Studying Brain-Behavior Correlations with fMRI

Catherine Walsh, Ph.D.

Section on Functional Imaging Methods, NIMH

FMRIF Summer Neuroimaging Course

July 30, 2024

Outline

- **What** are brain-behavior correlations?
- **Why** study brain-behavior correlations?
- **How** to study brain-behavior correlations?

What are brain-behavior correlations?

Brain – Behavior Correlations

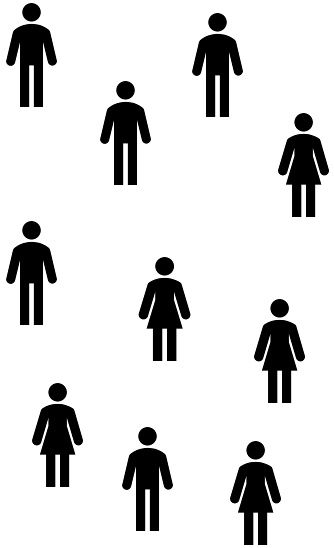
- Activation (in an ROI, a network)
- Functional connectivity (at rest, during task)
- Structural measures (cortical thickness, DWI)
- Other modalities (EEG, MEG, etc)

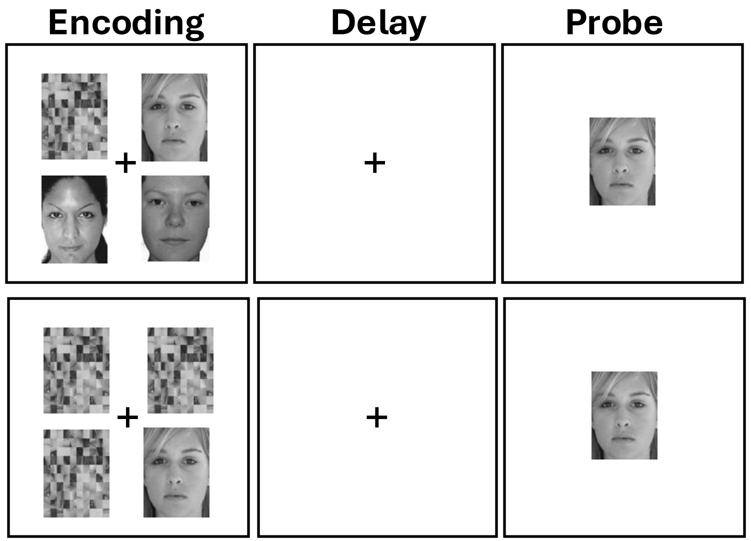
- Performance on a task (accuracy, RT)
  - Could be in or outside the scanner
- Self-report measures (personality traits, psychiatric symptoms)

- Some sort of statistical connection/association between the two

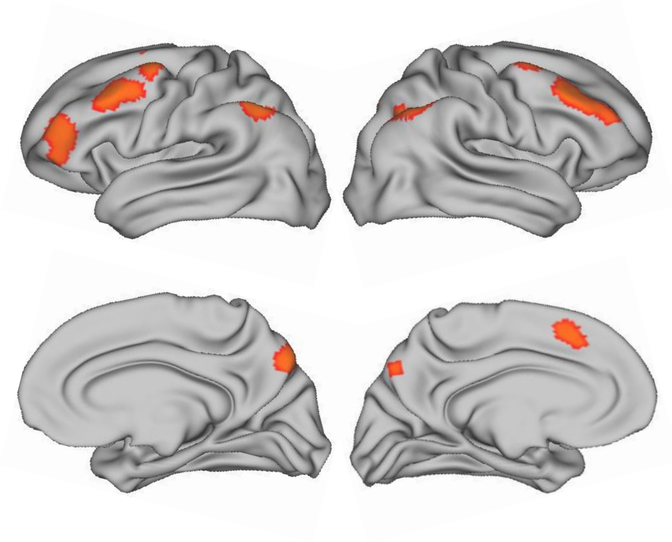
Why study brain-behavior correlations?

Traditional fMRI (group-level) analyses



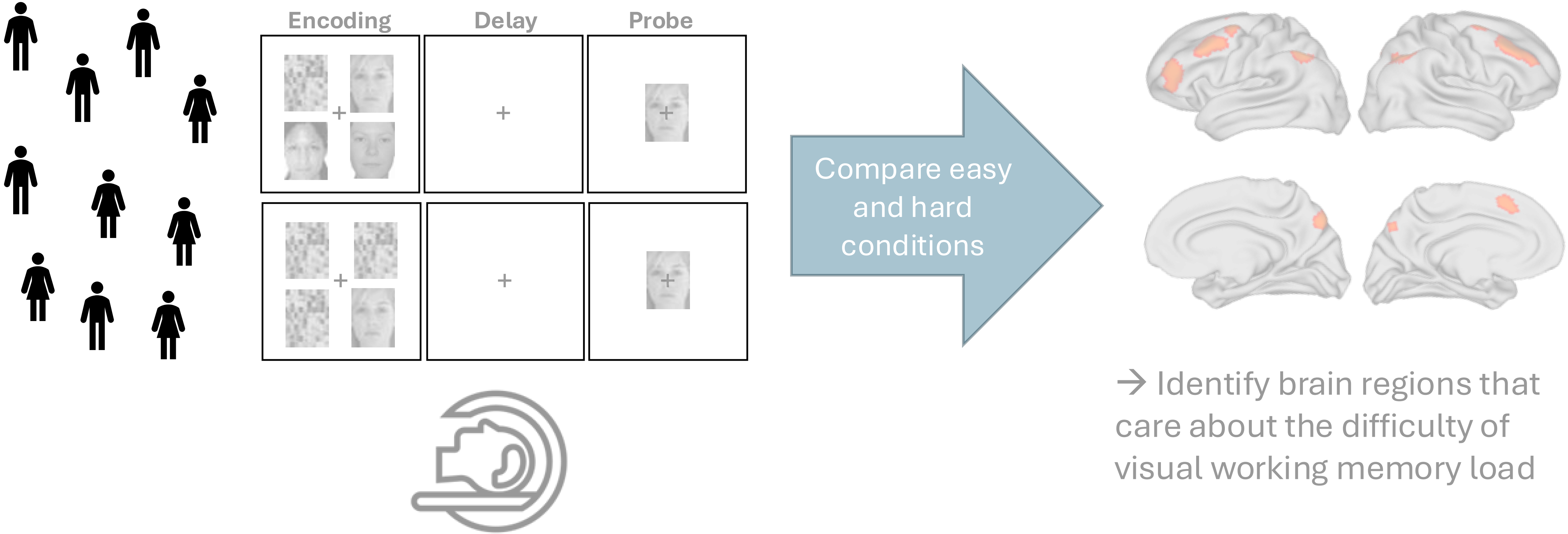


Compare easy and hard conditions



→ Identify brain regions that care about the difficulty of visual working memory task

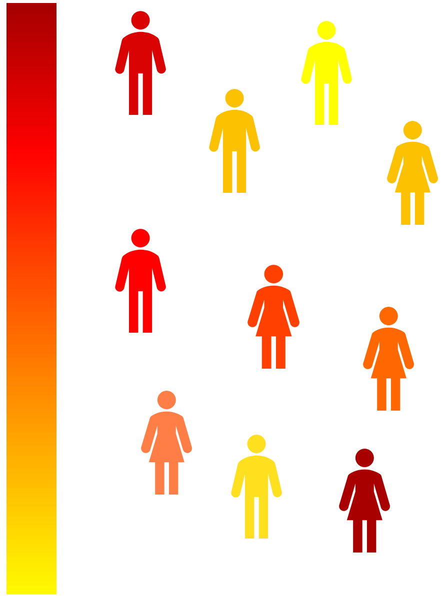
Traditional fMRI (group-level) analyses



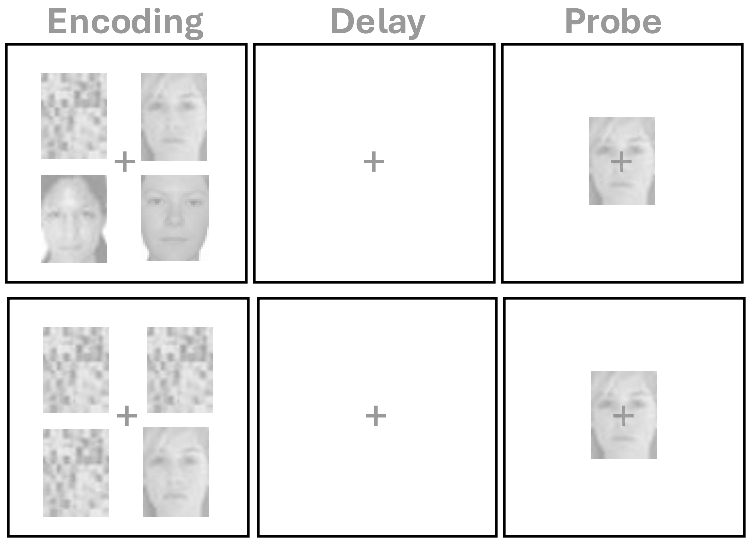
→ Identify brain regions that care about the difficulty of visual working memory load

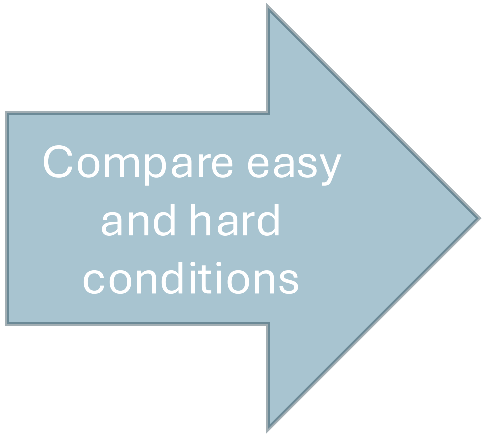
But wait, there's more (information!)

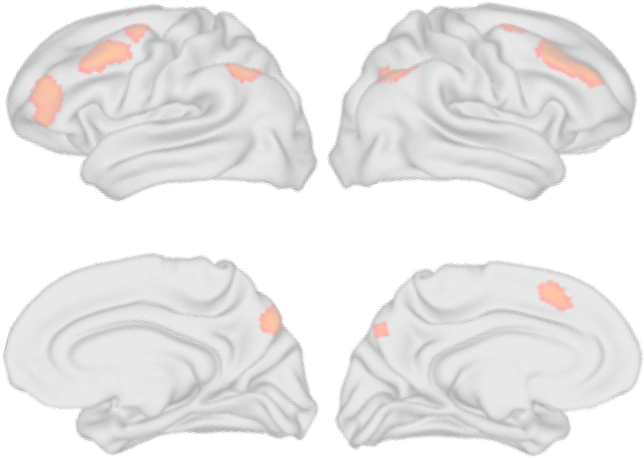
Better
score



Worse
score



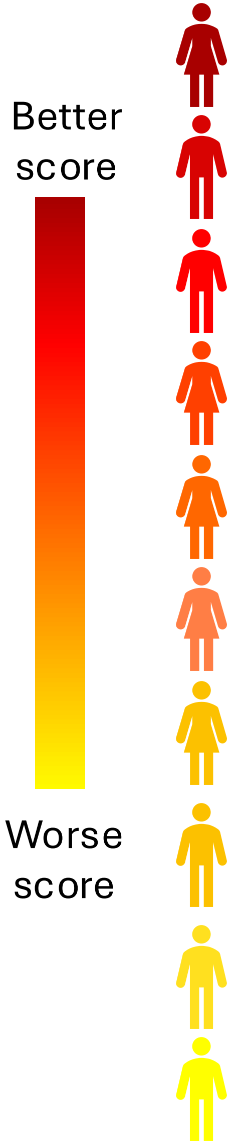




→ Identify brain regions that care about the difficulty of visual working memory load

But wait, there's more (information!)

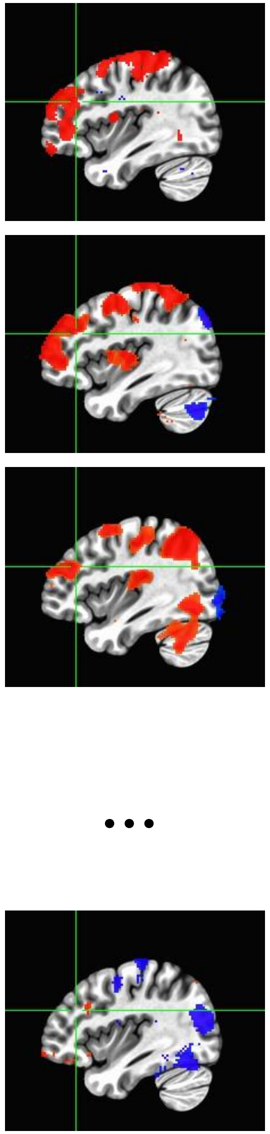
Better score



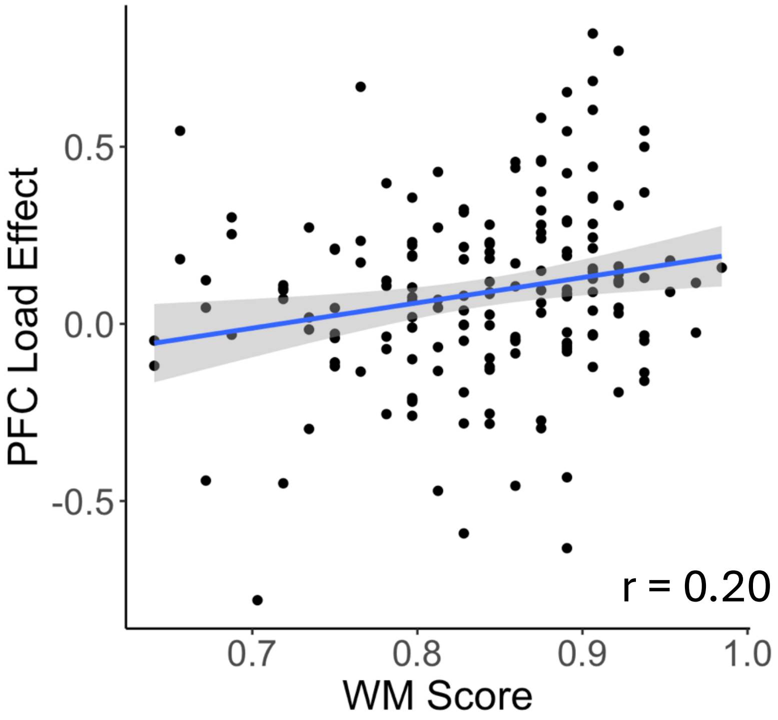
Worse score

Compare conditions

for each participant



Correlation

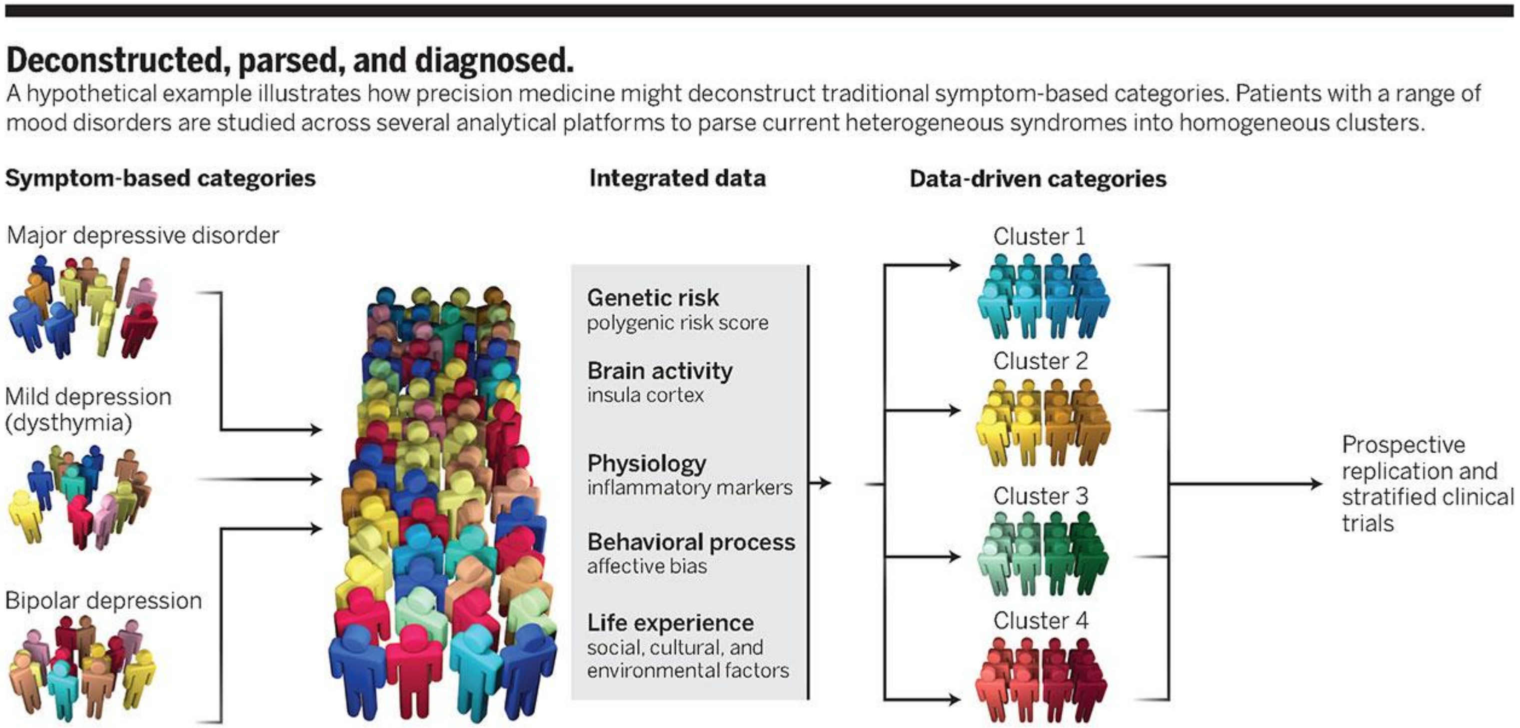


→ More activation in lateral PFC associated with better performance on task!

Psychiatric symptoms as dimensional, not categorical

Research Domains Criteria (RDoC)

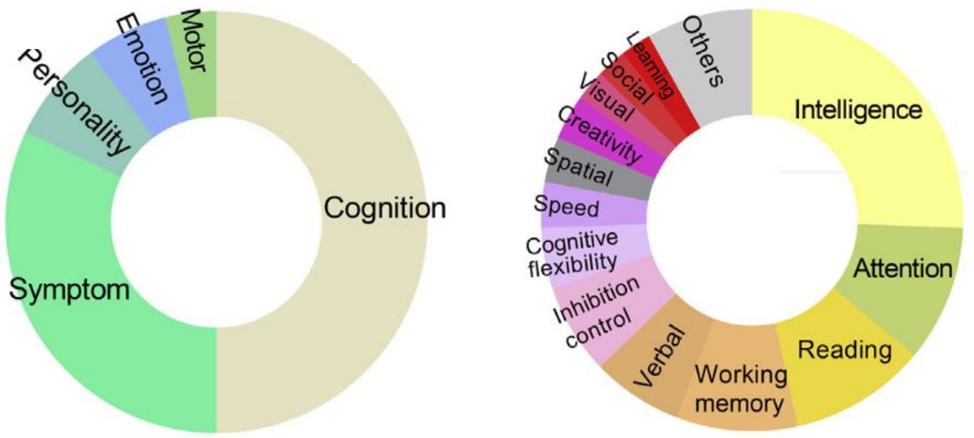
Goal: "to develop, for research purposes, new ways of classifying mental disorders based on dimensions of observable behavior and neurobiological measures"



So what?

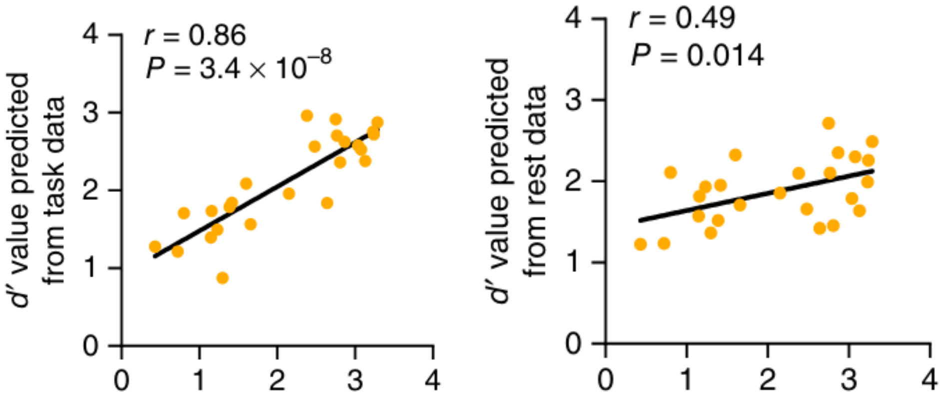
- Better understanding of disease pathophysiology
- Track disease progression
- Develop new (more effective) treatments
- Understand who will respond to which treatments

Predicting cognition



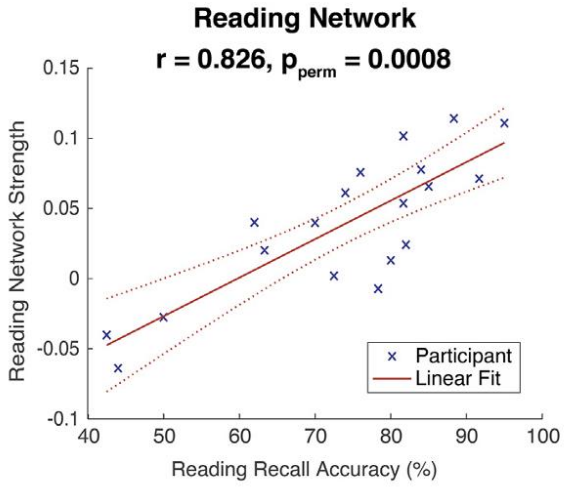
Sui et al., Biol Psych 2022

Sustained Attention



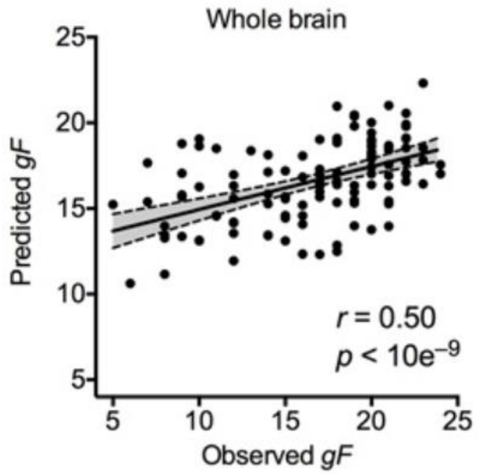
Rosenberg et al., *Nat Neuro* 2015

Reading Ability



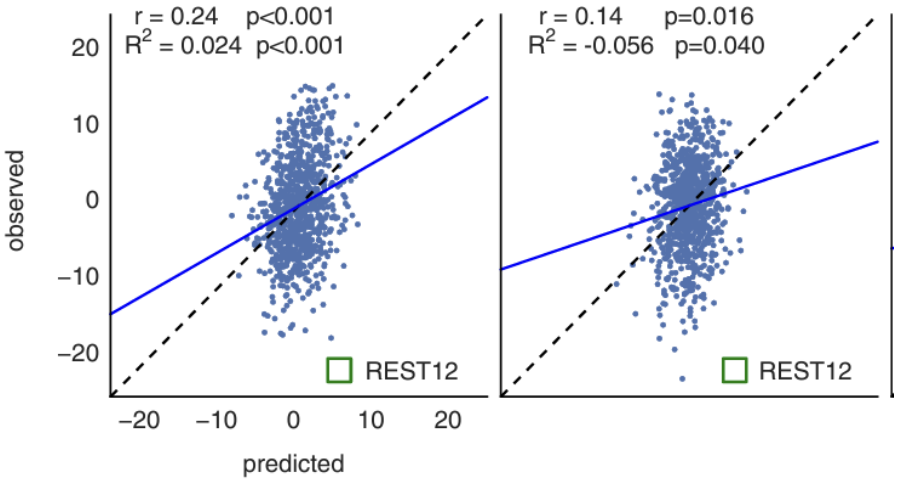
Jangraw et al., *NeuroImage* 2018

Fluid Intelligence



Finn et al., *Nat Neuro* 2015

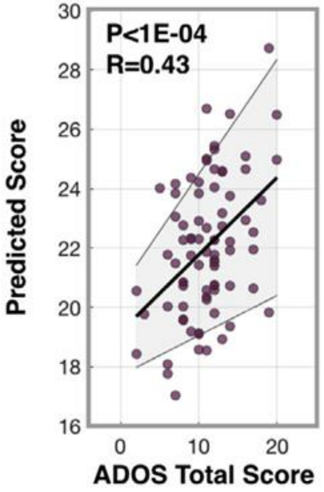
Openness and Conscientiousness



Dubois and Galdi et al., *Personality Neuro* 2018

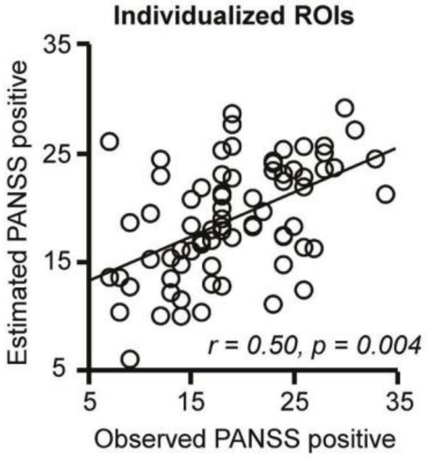
Predicting psychiatric symptoms

Autism



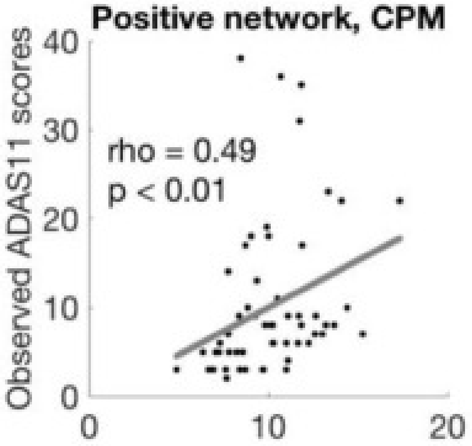
Lake et al., Biol Psych 2019

Schizophrenia



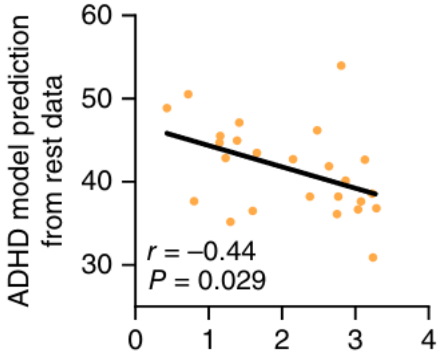
Wang et al., Mol Psych 2020

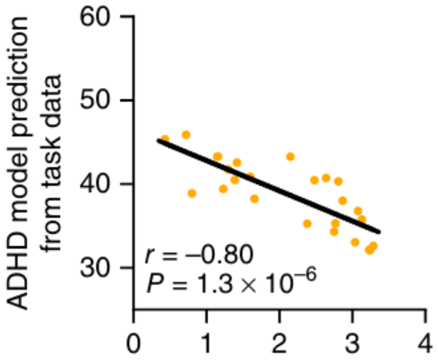
Alzheimer's Disease



Lin et al., Front Aging Neurosci 2018

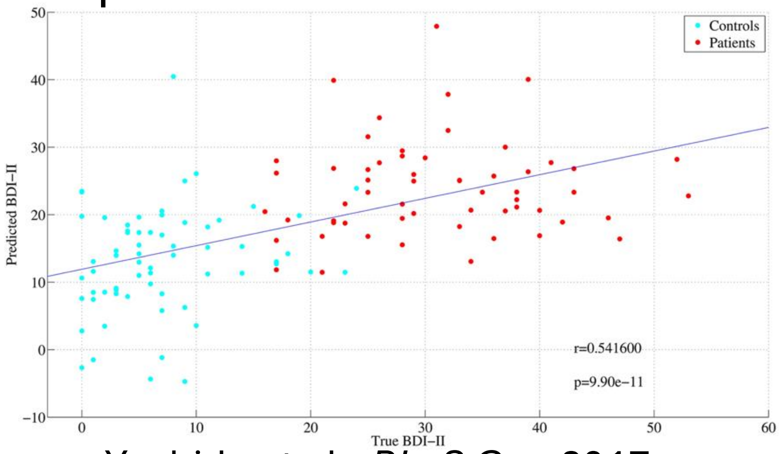
ADHD





Rosenberg et al., *Nat Neuro* 2015

Depression



Yoshida et al., *PLoS One* 2017

Why study brain-behavior correlations?

- Basic science!
  - New insights into neural processes
- Move towards brain-based psychiatry
  - Better understanding of how diseases work and how they progress
  - More personalized, targeted interventions: drugs, therapies, etc

But wait!

Are most brain-behavior correlations even meaningful?!?

nature

Explore content About the journal Publish with us

nature > articles > article

Article | Open access | Published: 16 March 2022

Reproducible brain-wide association studies requ thousands of individuals

Scott Marek, Brenden Tervo-Clemmens, Finnegan J. Calabro, David F. Montez, Benjamin P. Kay, Alexander S. Hatoum, Meghan Rose Donohue, William Foran, Ryland L. Miller, Timothy J. Hendrickson, Stephen M. Malone, Sridhar Kandala, Eric Feczko, Oscar Miranda-Dominguez, Alice M. Graham, Eric A. Earl, Anders J. Perrone, Michaela Cordova, Olivia Doyle, Lucille A. Moore, Uriarte, Kathy Snider, Benjamin J. Lynch, ... Nico U. F. Dosenbach +

Nature **603**, 654–660 (2022) | Cite this article

NEWS | 17 March 2022

Can brain scans reveal behaviour? Bombshell study says n

Most studies linking features in brain imagi small to be reliable, argues a controversi

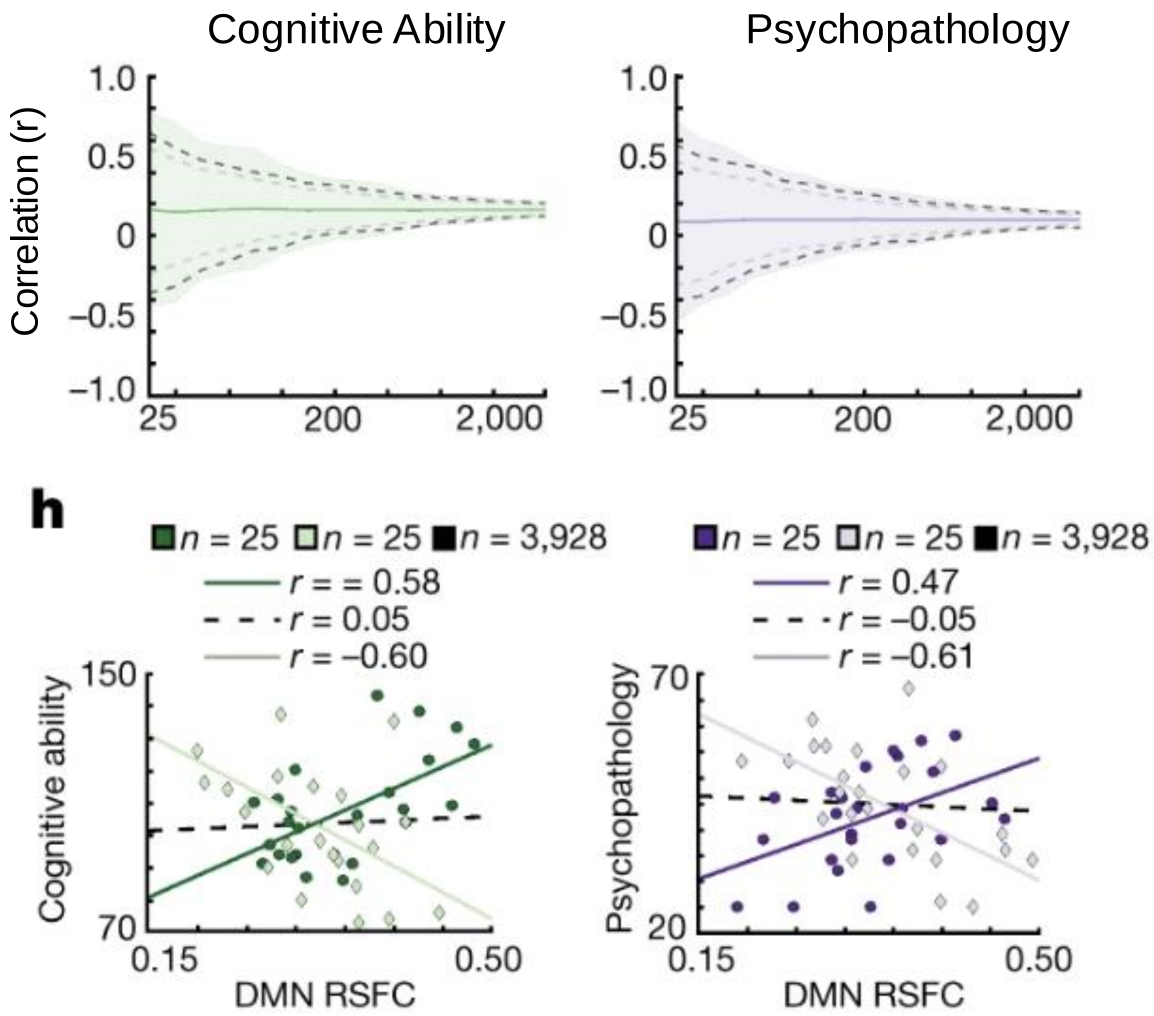
By Ewen Callaway

Scanning the Brain to Pr 'Task' for MRI

June 3, 2020

TAGS: BRAIN IMAGING | FMRI | NEWS | TASK PERFORMANCE

We need thousands of individuals?!?



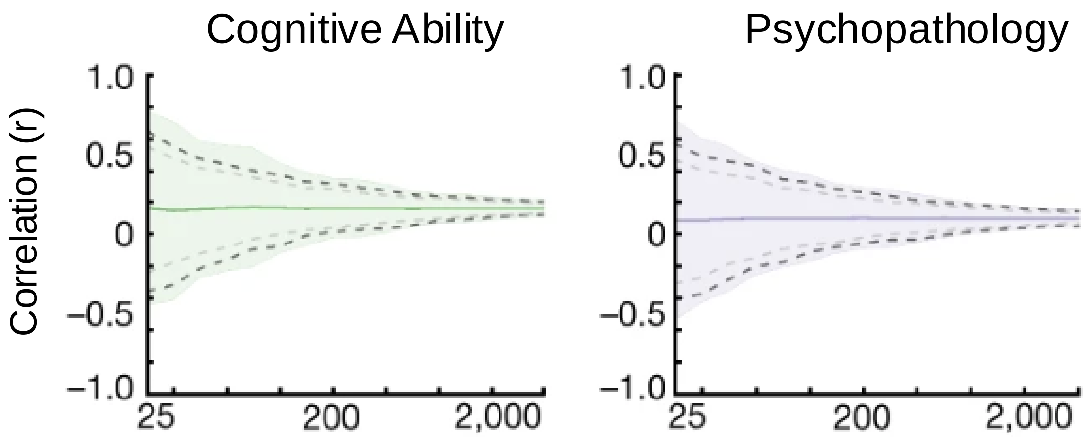
Marek et al., *Nature* 2022

→ Widely varying effect sizes at small sample sizes – even producing completely opposite results!

How to study* brain-behavior correlations?

*robustly

Sample size



Marek et al., *Nature* 2022

Benefits of a larger sample size:

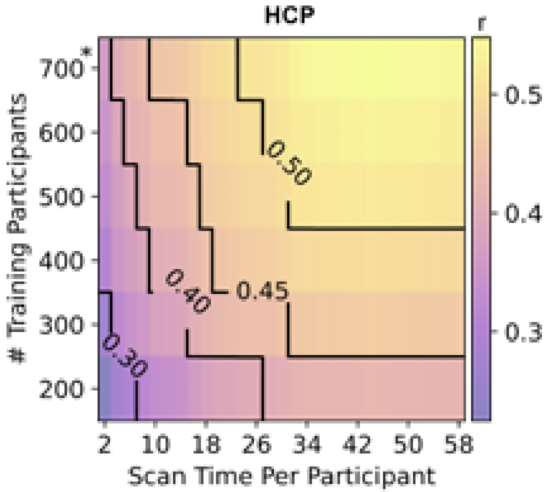
- Correlations get more reliable!
- Wider distribution of phenotypes
- Allows us to use more robust statistical/machine learning methods like cross validation

Problems with using a larger sample size:

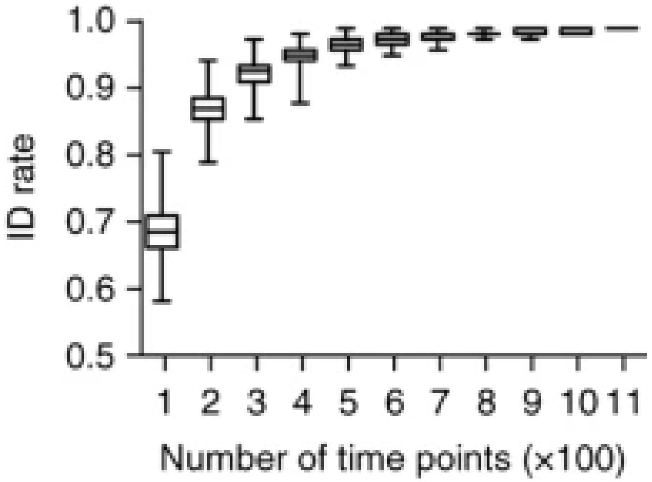
- Expensive
- Time consuming
- Hard to recruit patients

So what if I can't recruit thousands of subjects?

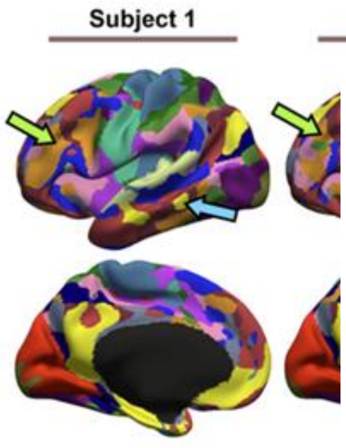
→ Get more scan time with fewer subjects



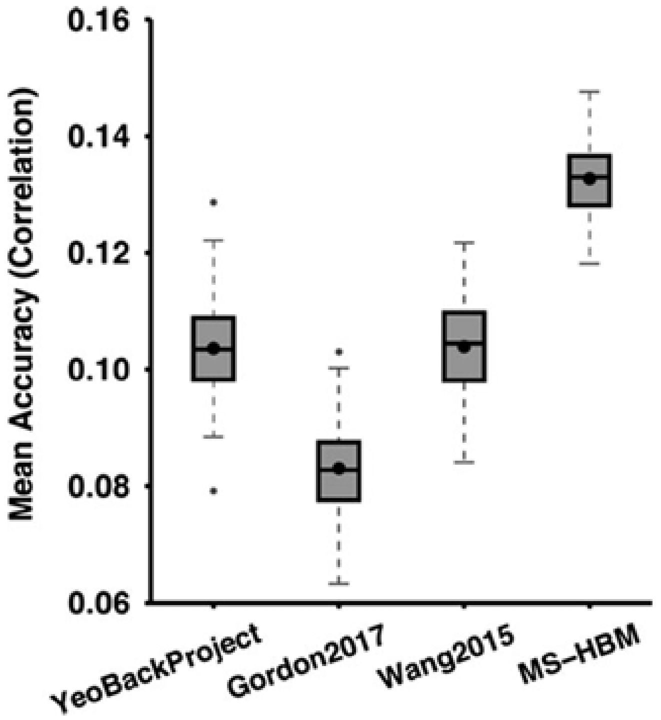
Ooi et al, *bioRxiv* 2024



Finn et al., *Nat Neurosci* 2015



Kong et al., *Ce*



So what if I can't recruit thousands of subjects?

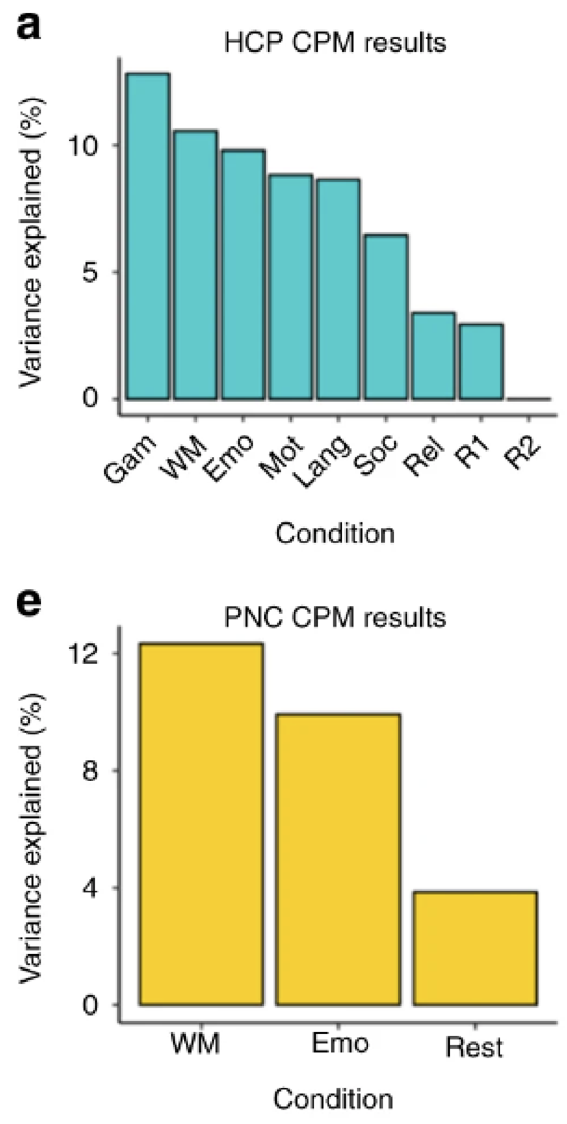
→ Consider the measures you use: task vs rest

Many studies use resting state fMRI to predict behavior

- Relatively reliable across session ("trait-like")
- Reflects functional networks that are a "backdrop" to anything that happens during task
- Easy to measure – no task to learn, requires relatively little scan time (~15 minutes)
- Many big, open datasets include it!

BUT: is rest the best for prediction?

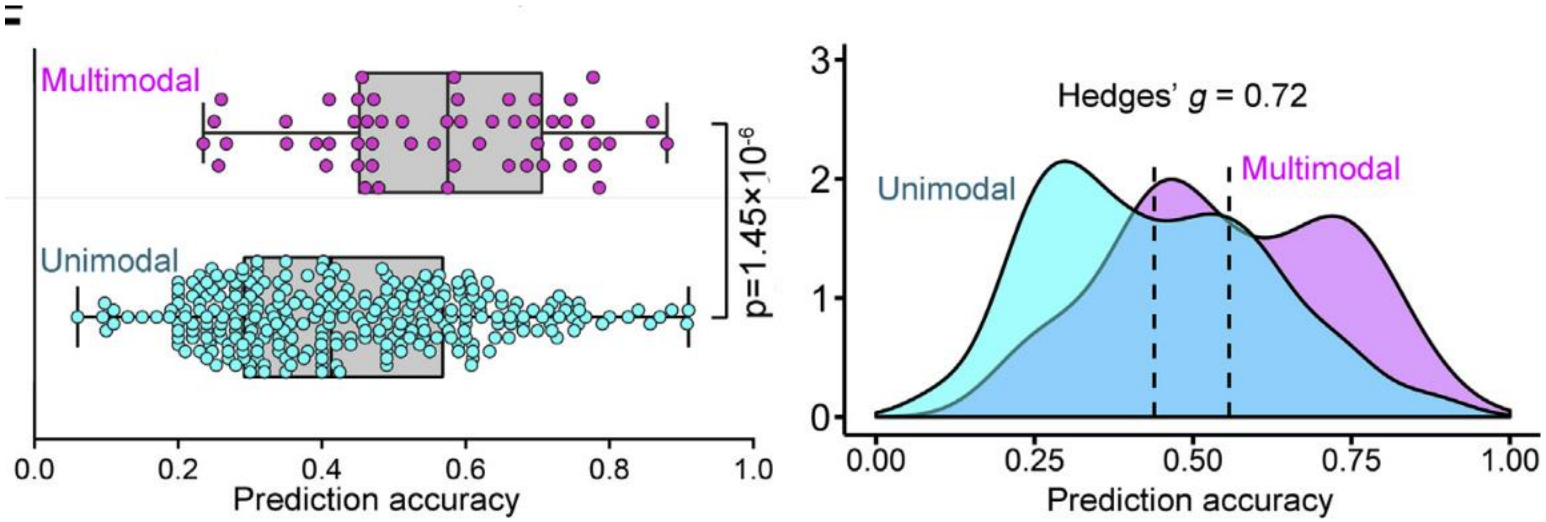
Maybe not.



Greene et al., *Nat Comms* 2018

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: task vs rest (or both!?)



Sui et al., *Biol Psych*. 2020

→ Integrating across neuroimaging features can improve prediction performance and leverage unique aspects of brain structure and function to better characterize behavioral traits

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: self-report vs cognitive task

Take for example: face blindness (prosopagnosia)

Option 1: Self-report measures

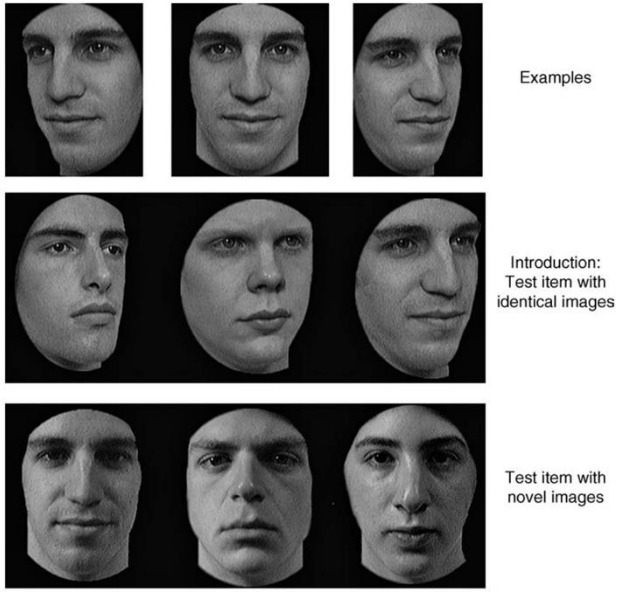
20 item prosopagnosia index (PI20)

1	My face recognition ability is worse than most people
2	I have always had a bad memory for faces
3	I find it notably easier to recognize people who have distinctive facial features
4	I often mistake people I have met before for strangers
5	When I was at school I struggled to recognize my classmates

Shah et al., *R. Soc. Open sci*, 2015

Option 2: Data from cognitive tasks

Cambridge Face Memory Test



Duchaine et al., Neuropsychologia 2006

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: self-report vs cognitive task

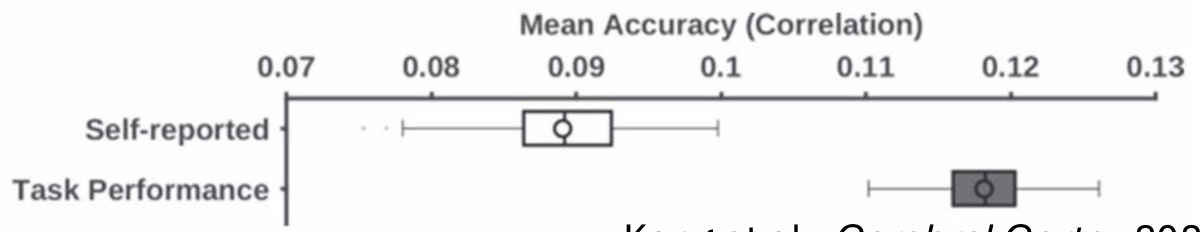
Take for example: face blindness (prosopagnosia)

Option 1: Self-report measures

- Pros:
    - Easy to administer
    - Might be more relevant for psychiatric conditions
    - Potentially more stable within an individual
- Cons:
    - Potential for bias
    - Relies on participant's ability to introspect

Option 2: Data from cognitive tasks

- Pros:
    - Less potential for bias
    - Able to do in scanner
    - Easier to predict?
- Cons:
    - Might have less relevance for "biomarkers"
    - Potentially less generalizable



Kong et al., *Cerebral Cortex* 2021

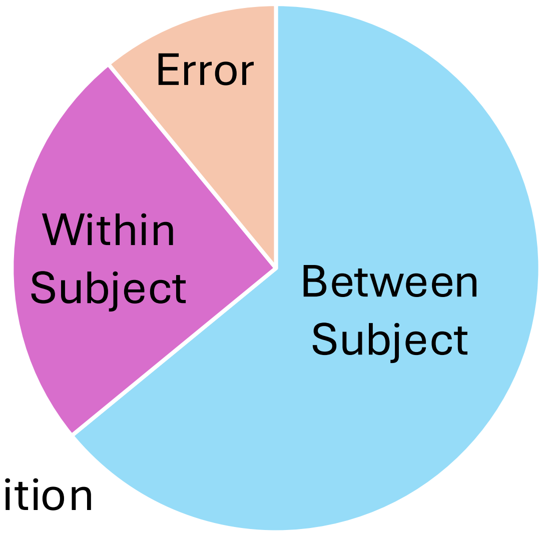
So what if I can’t recruit thousands of subjects?

→ Consider the measures you use: optimize sources of variance

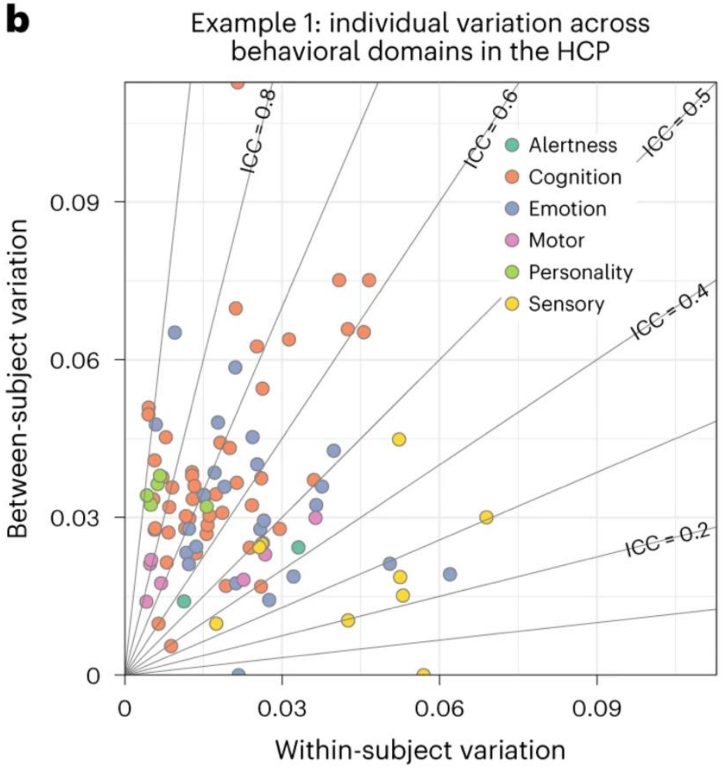
% of variance explained by a given measure

Example:
Measurement noise

Examples:
Metabolic changes
Physiological state
Arousal
Attention
Experimental Condition



True differences across individuals



Xu et al., *Nature Methods* 2023

Between and within subject variance can both be interesting targets of prediction, they just are asking different questions!

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: optimize sources of variance

Between and within subject variance can both be interesting targets of prediction, they just are asking different questions!

→ Different questions necessitate different tasks

Within Subject Effects
(Group/Condition Differences)

$$t = \frac{Difference\ in\ conditions}{Error}$$

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: optimize sources of variance

Between and within subject variance can both be interesting targets of prediction, they just are asking different questions!

→ Different questions necessitate different tasks

Within Subject Effects (Group/Condition Differences)

$$t = \frac{Difference\ in\ conditions}{Variance\ between\ individuals + Error}$$

Between Subject Effects (Brain-behavior Correlations)

$$ICC = \frac{Variance\ between\ individuals}{Error}$$

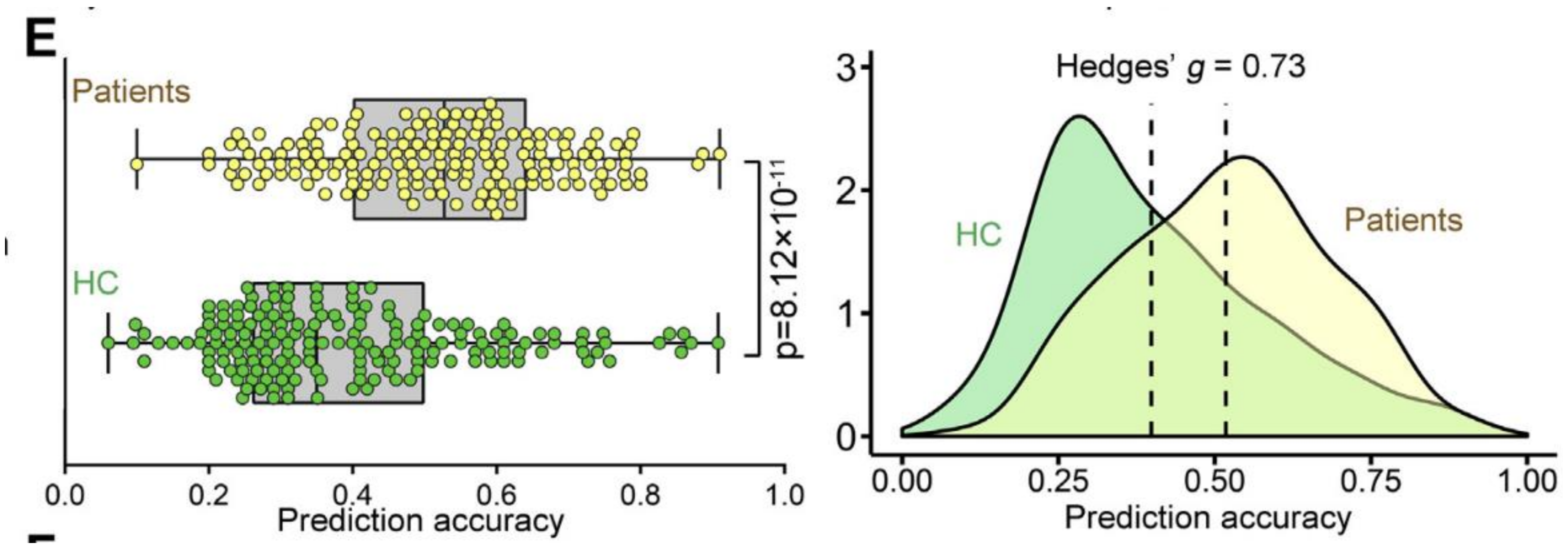
→ Tasks that are optimized for within-subjects effects may be poorly optimized for between-subjects effects

See Hedge et al., *Beh. Res. Methods* 2017 for more info/math

So what if I can't recruit thousands of subjects?

→ Consider the measures you use: optimize sources of variance

Also consider what sample you're using!

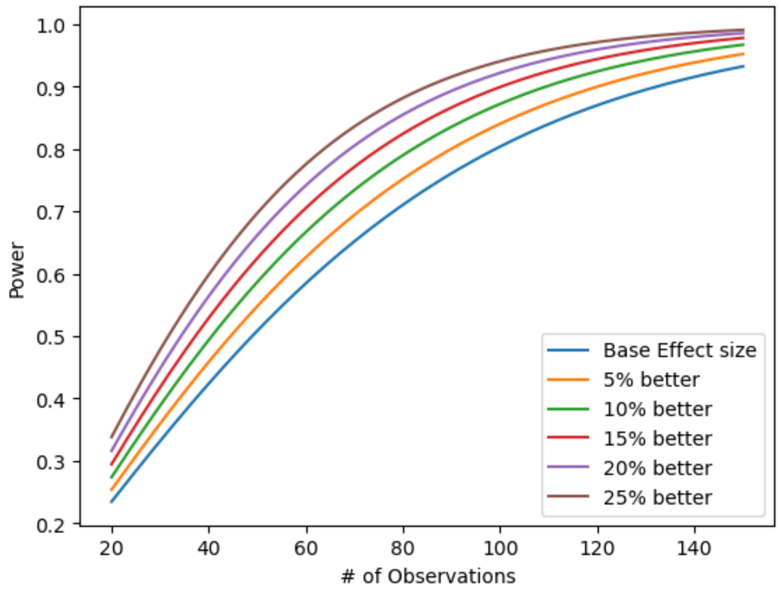


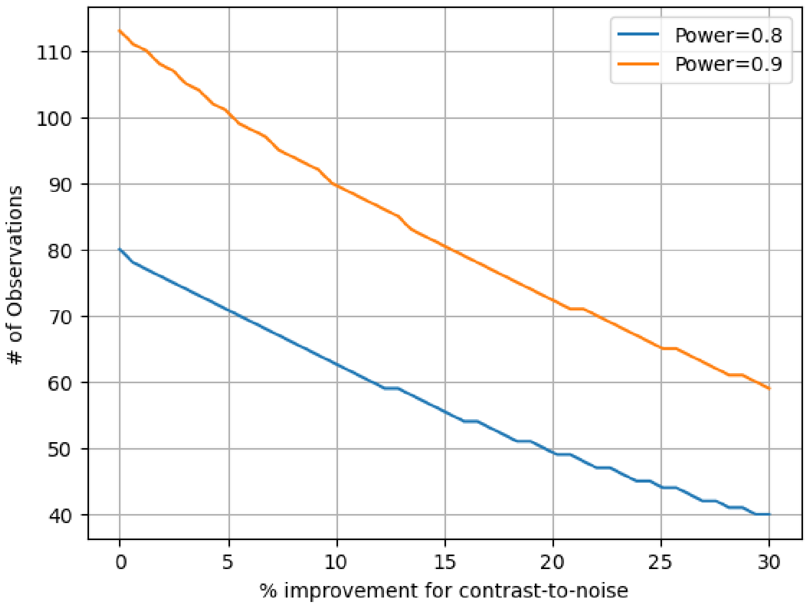
Sui et al., *Biol Psych*. 2020

Consider: if you're trying to find a biomarker for a specific psychiatric phenotype in a sample of healthy volunteers, there might not be enough between-subject psychiatric variance for a model to pick up on

So what if I can’t recruit thousands of subjects?

→ Improve your data quality





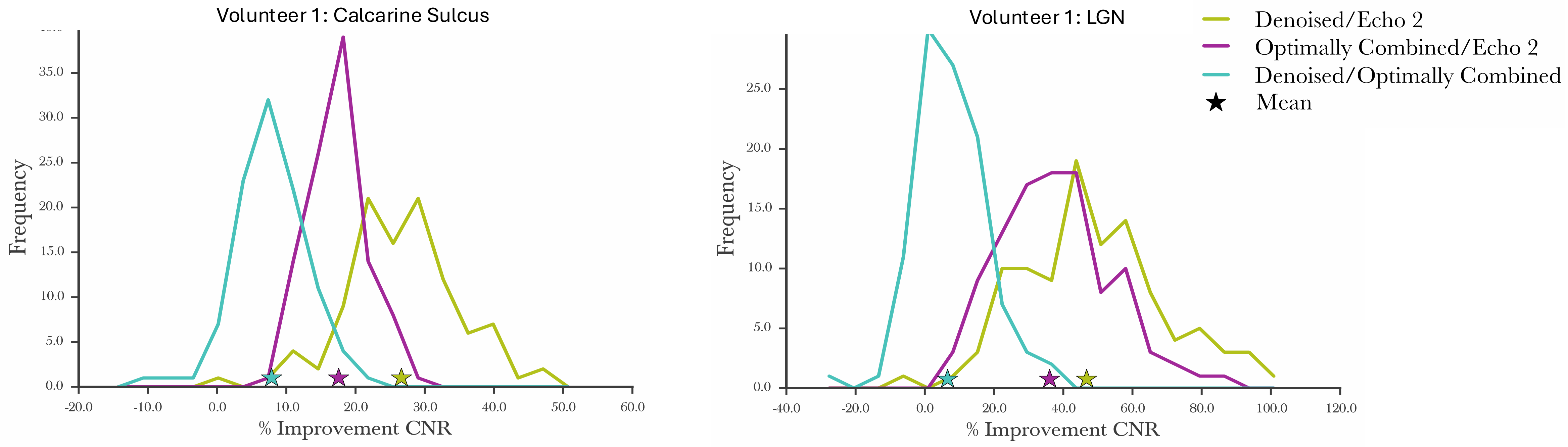
A 10% improvement in contrast-to-noise could mean a statistical power of 0.8 is possible with 63 vs 80 subjects

Slide from Dan Handwerker

So what if I can't recruit thousands of subjects?

→ Improve your data quality: acquisition parameters (use multi-echo!)

CNR % Improvement Between Preprocessing Methods



Slide from Dan Handwerker

So what if I can't recruit thousands of subjects?

→ Improve your data quality: decrease head motion

Subject measures	Pearson r
ReadEng (AgeAdj)	−0.23
ReadEng (Unadj)	−0.23
Vocabulary (AgeAdj)	−0.19
Dexterity (Unadj)	−0.18
CardSort (Unadj)	−0.18
Dexterity (AgeAdj)	−0.18
CardSort (AgeAdj)	−0.18
Education	−0.17
Fluid intelligence	−0.17
Spatial orientation	−0.17
Vocabulary (unadjj)	−0.17
Emotion recognition	−0.16
DSM somatic problems (pct)	0.16
DSM antisocial (raw)	0.16
ASR externalizing (raw)	0.16
DSM somatic problems (raw)	0.16
Tobacco use 7 day	0.18
Diastolic blood pressure	0.18
ASR externalizing	0.18
Tobacco use today	0.2
Systolic blood pressure	0.23
Weight	0.52
Body mass index (BMI)	0.66

Siegel et al., *Cerebral Cortex* 2017

→Head motion significantly correlated with subject measures from HCP

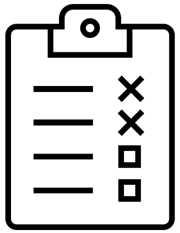
→Also see greater head motion in certain populations: kids, older adults, psychiatric patient populations

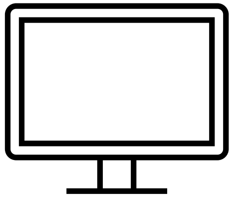
So what if I can’t recruit thousands of subjects?

→ Improve your data quality: maximize reliability

Key assumption of brain-behavior correlations: we are measuring stable, trait-like things

Put another way: the things we are measuring are **RELIABLE**





Trial 1 ✓

Trial 2 ✓

Trial 3 ✗

Trial 4 ✓

Trial 5 ✓

Trial 6 ✗

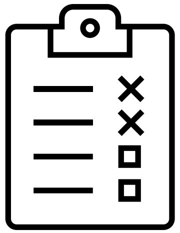
Overall accuracy: 0.66

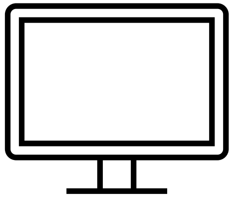
So what if I can’t recruit thousands of subjects?

→ Improve your data quality: maximize reliability

Key assumption of brain-behavior correlations: we are measuring stable, trait-like things

Put another way: the things we are measuring are **RELIABLE**



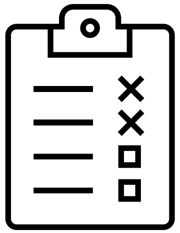


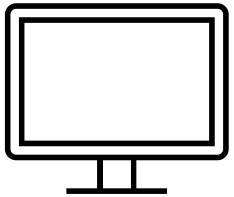
Trial 1 ✓	Trial 2 ✓
Trial 3 ✗	Trial 4 ✓
Trial 5 ✓	Trial 6 ✗
Accuracy 1: 0.66	Accuracy 2: 0.66

So what if I can’t recruit thousands of subjects?

→ Improve your data quality: maximize reliability

Key assumption of brain-behavior correlations: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**





Trial 1	✓	Trial 2	✓
Trial 3	✗	Trial 5	✓
Trial 4	✓	Trial 6	✗

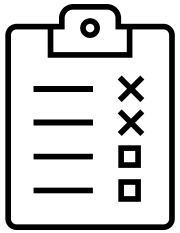
Accuracy 3: 0.66 Accuracy 4: 0.66

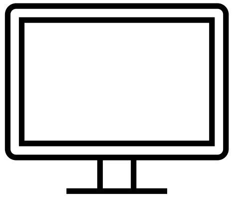
So what if I can't recruit thousands of subjects?

→ Improve your data quality: maximize reliability

Key assumption of brain-behavior correlations: we are measuring stable, trait-like things

Put another way: the things we are measuring are **RELIABLE**





Trial 5	✓	Trial 2	✓
Trial 6	✗	Trial 1	✓
Trial 4	✓	Trial 3	✗

Accuracy 5: 0.66

Accuracy 6: 0.66

Split-halves reliability: the task you're using is internally consistent (i.e. task is measuring the same construct all the way through)

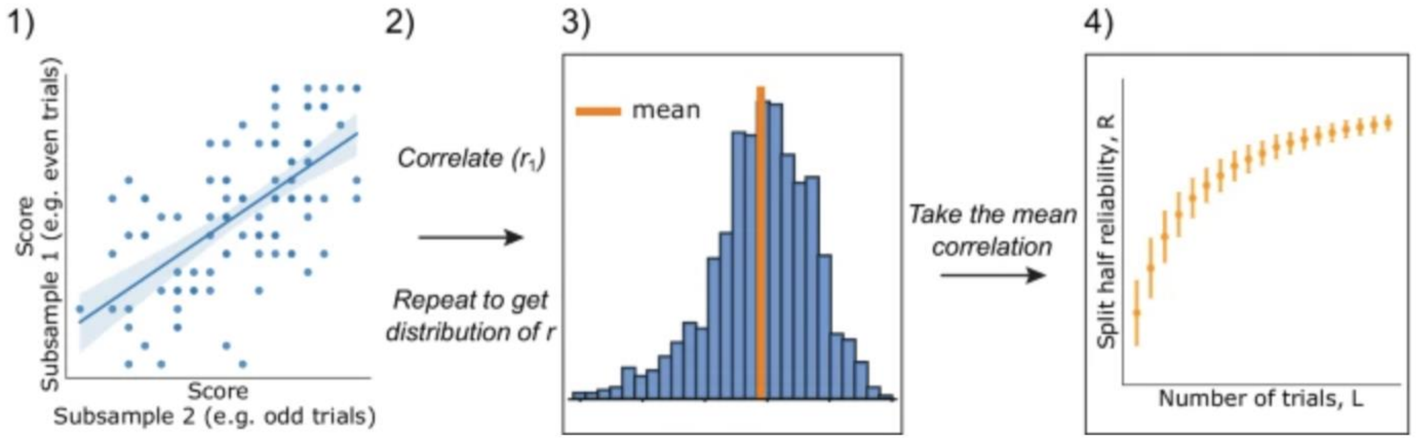
→ *Do all tasks show acceptable levels of split-halves reliability??*

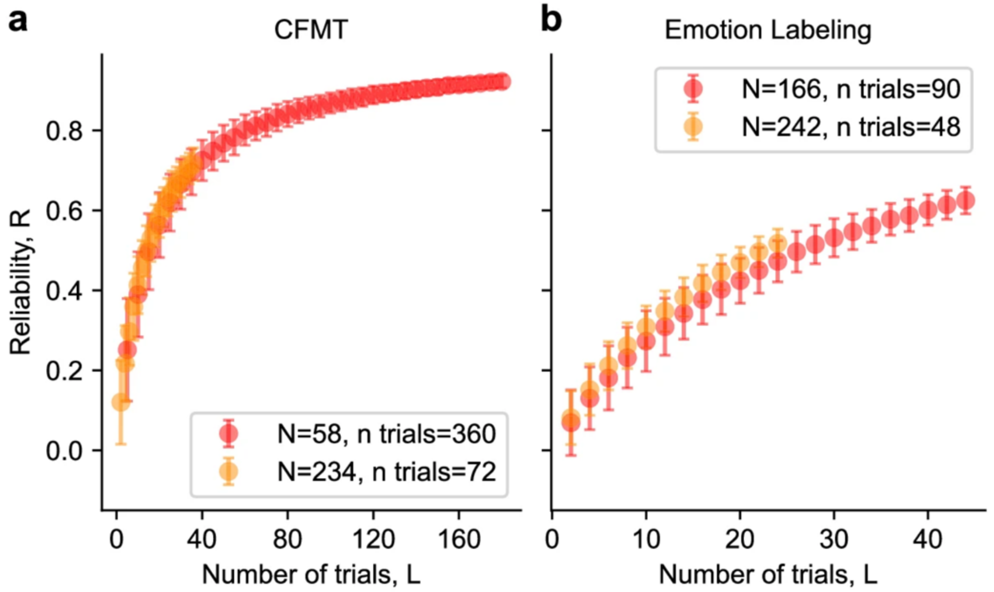
So what if I can't recruit thousands of subjects?

→ Improve your data quality: maximize reliability

Key assumption of brain-behavior correlations: we are measuring stable, trait-like things

Put another way: the things we are measuring are **RELIABLE**





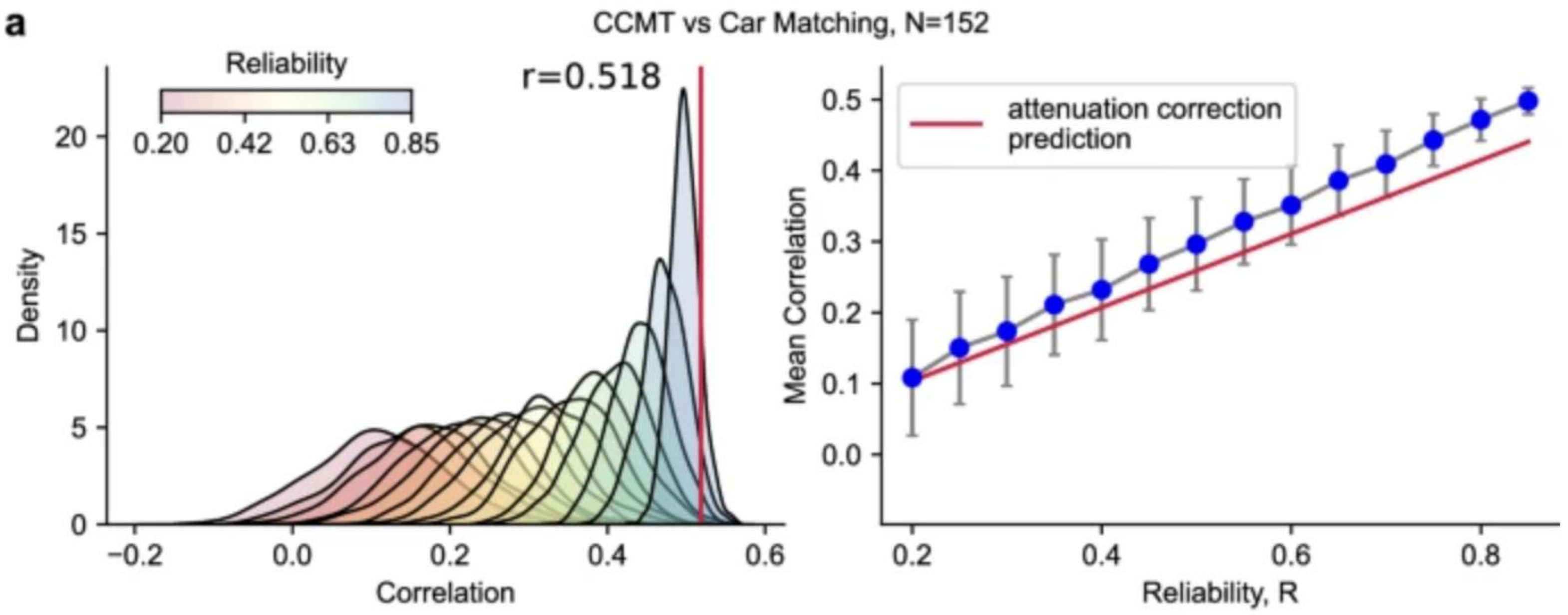
→ Number of trials it takes to reach an "acceptable" level of reliability varies across task

Kadlec, *Walsh* et al., *Comms Psych* 2024

So what if I can't recruit thousands of subjects?

→ Improve your data quality: maximize reliability

Key assumption of brain-behavior correlations: we are measuring stable, trait-like things
Put another way: the things we are measuring are **RELIABLE**



→Less reliable tasks show attenuated correlations, even when they are truly related

Kadlec, *Walsh* et al., *Comms Psych* 2024

*** Mathematically the same for brain-behavior, behavior-behavior, brain-brain, etc ***

So what if I can’t recruit thousands of subjects?

→ Improve your data quality: maximize reliability

Key assumption of brain-behavior correlations: we are measuring stable, trait-like things

Put another way: the things we are measuring are **RELIABLE**

→ *Test-retest reliability*: If we measure you at two different time points, you will score similarly

So what if I can't recruit thousands of subjects?

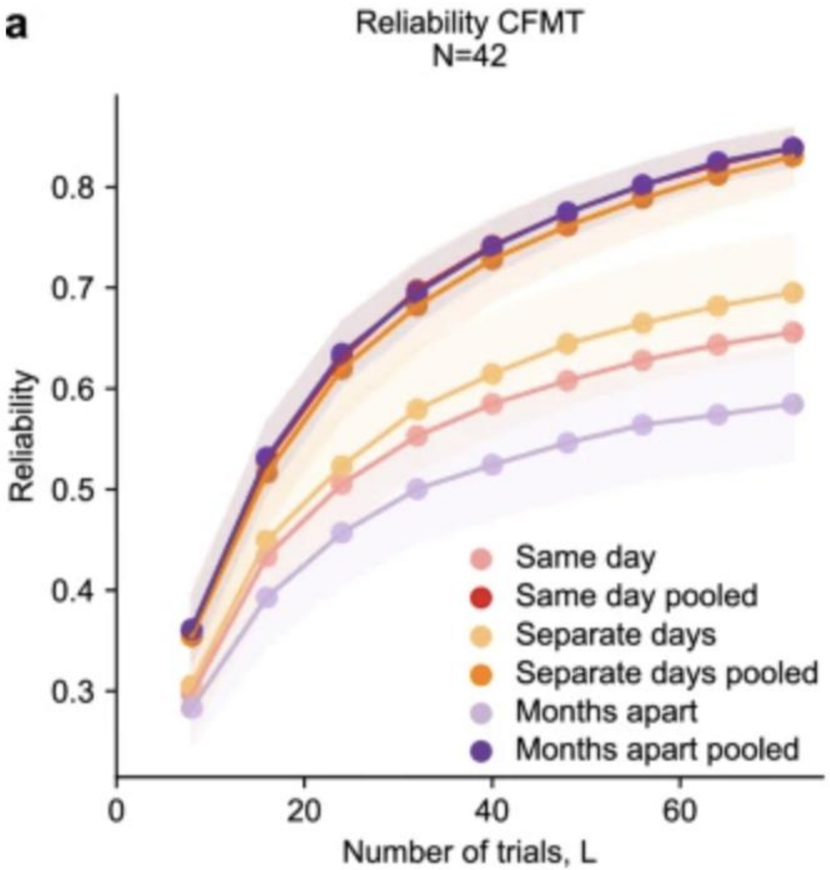
→ Improve your data quality: maximize reliability

3 batches of (non-overlapping) participants completed 2 versions of CFMT:
- Same day
- A few days apart
- 7-8 months apart

If sessions are equivalent (i.e. time doesn't matter), test-retest = split halves

Light colors: test-retest reliability (forms are kept separate)
Dark colors: split-halves reliability (forms are pooled)

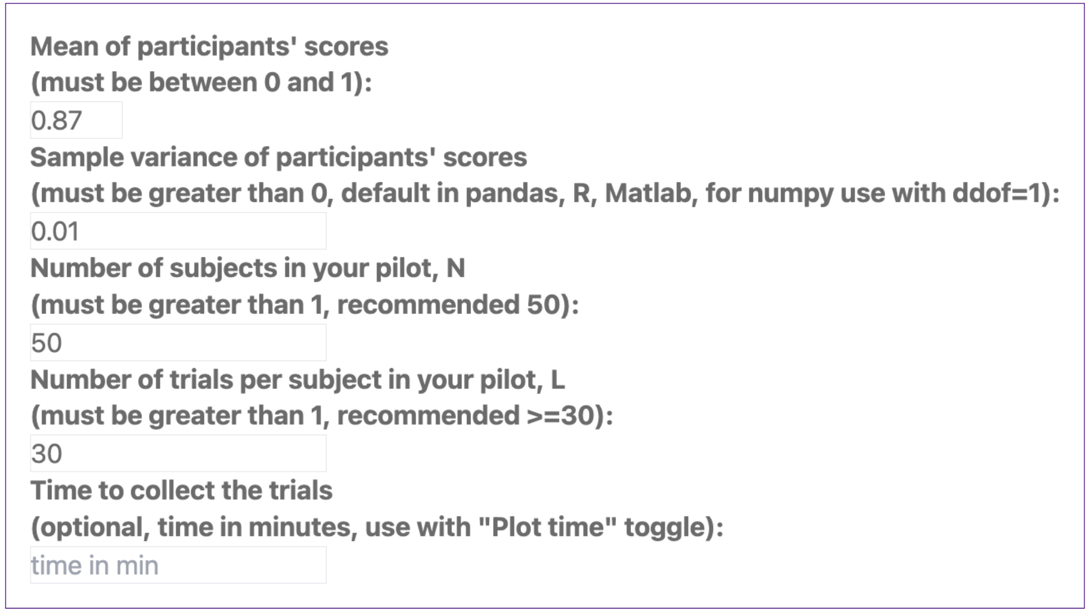
→ **Pooling data across sessions increases reliability of measures**

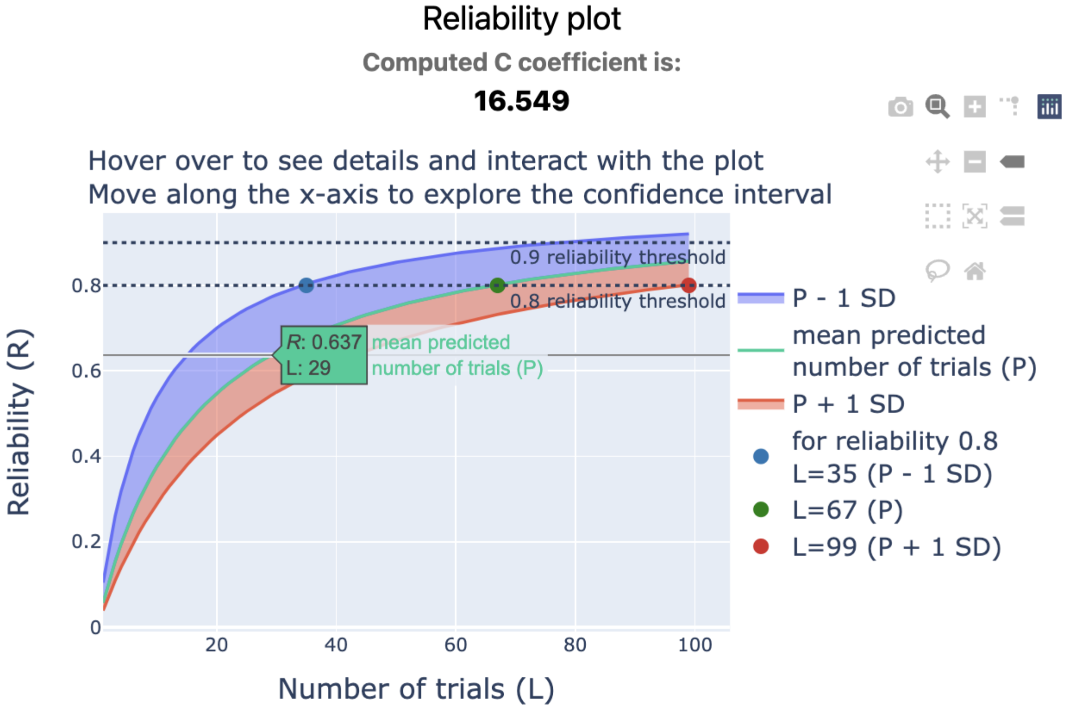


Kadlec, *Walsh* et al., *Comms Psych* 2024

So what if I can't recruit thousands of subjects?

→ Improve your data quality: maximize reliability





→ Web app to help explore reliability of tasks from a pilot sample (N ~ 50)

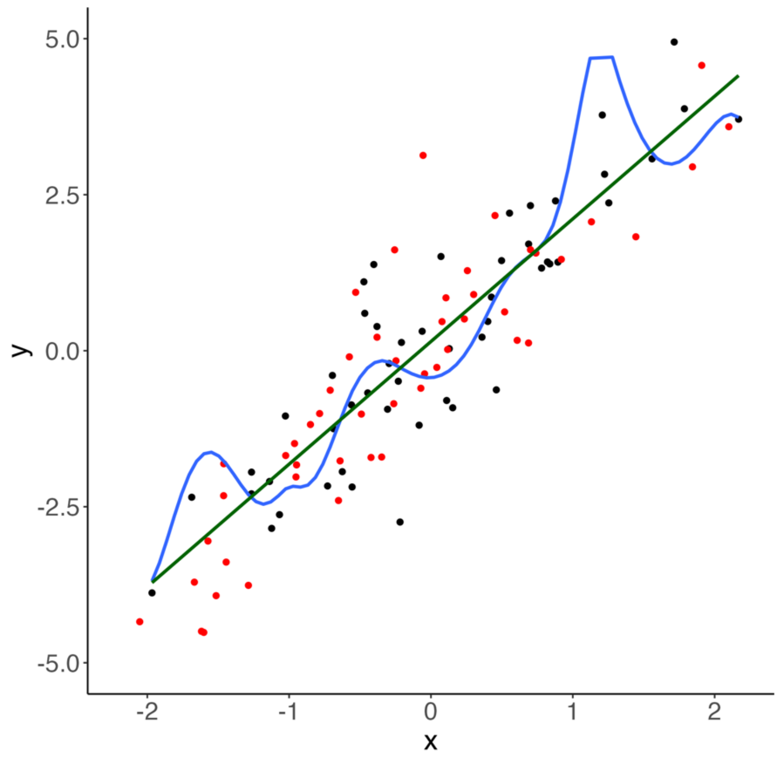
Kadlec, *Walsh* et al., *Comms Psych* 2024

https://jankawis.github.io/reliability-web-app/

So what if I can't recruit thousands of subjects?

→ Use better statistical methods: Cross-Validation

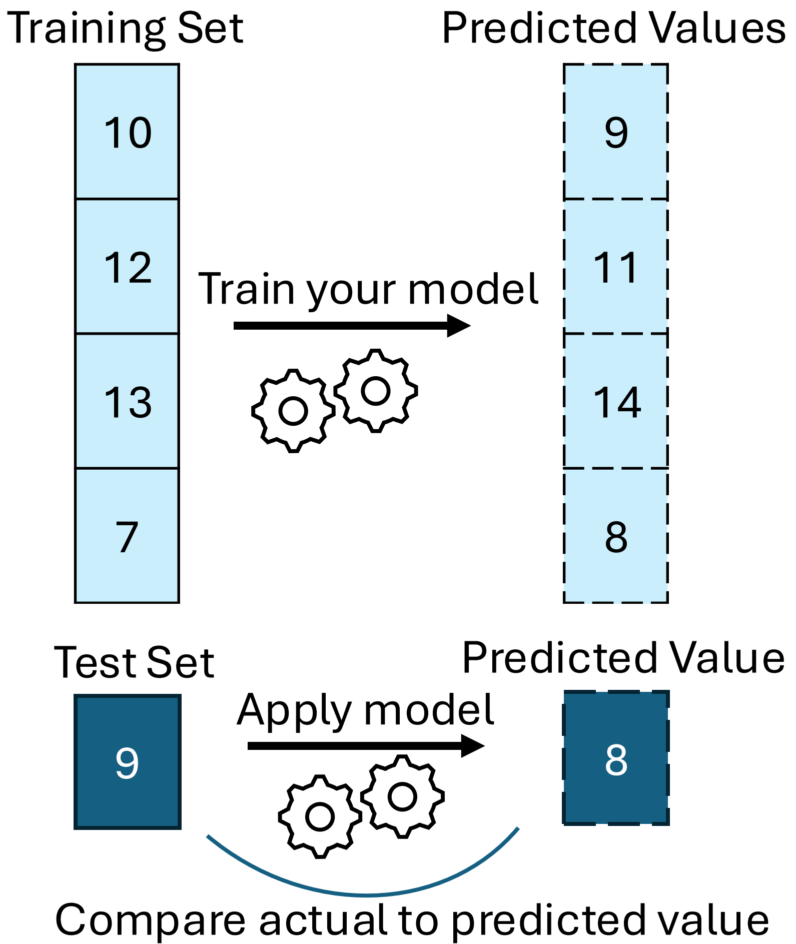
x = randomly drawn from normal distribution
y = 2*x + noise



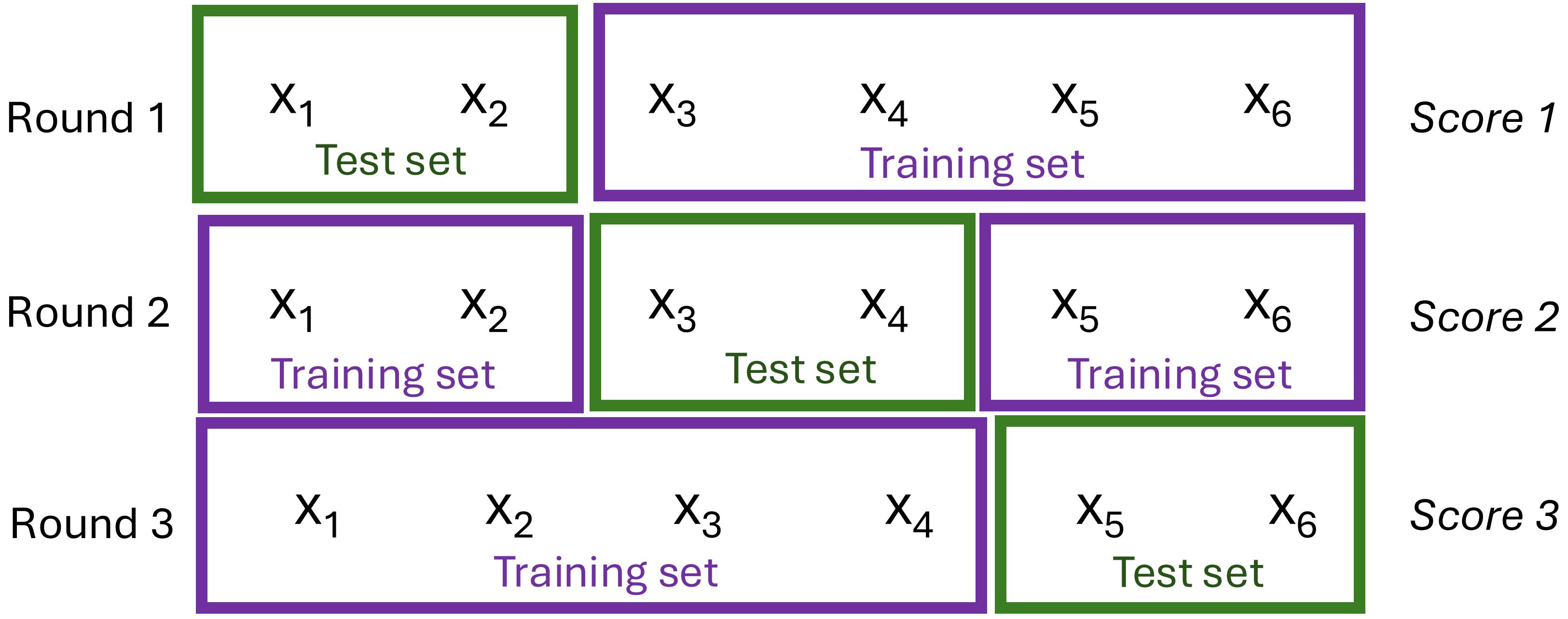
→ Creating your model in your entire sample can lead to fitting to sample specific characteristics (i.e. noise) and lead you to think you're doing better than you actually are

So what if I can't recruit thousands of subjects?

→ Use better statistical methods: Cross-Validation



K-fold cross validation

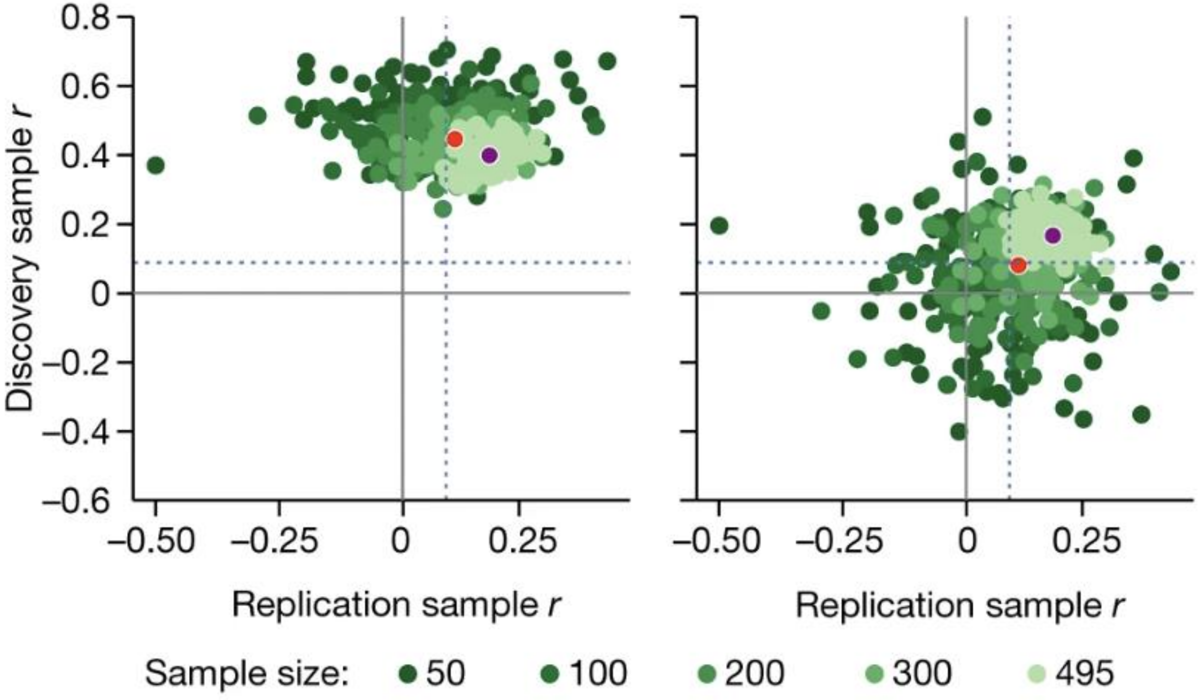


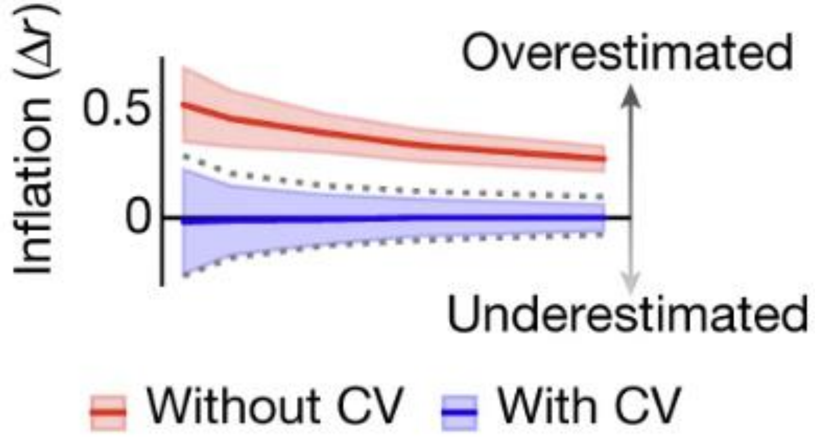
→ Average score from test sets across rounds

So what if I can’t recruit thousands of subjects?

→ Use better statistical methods: Cross-Validation

Without CV　　　　With CV



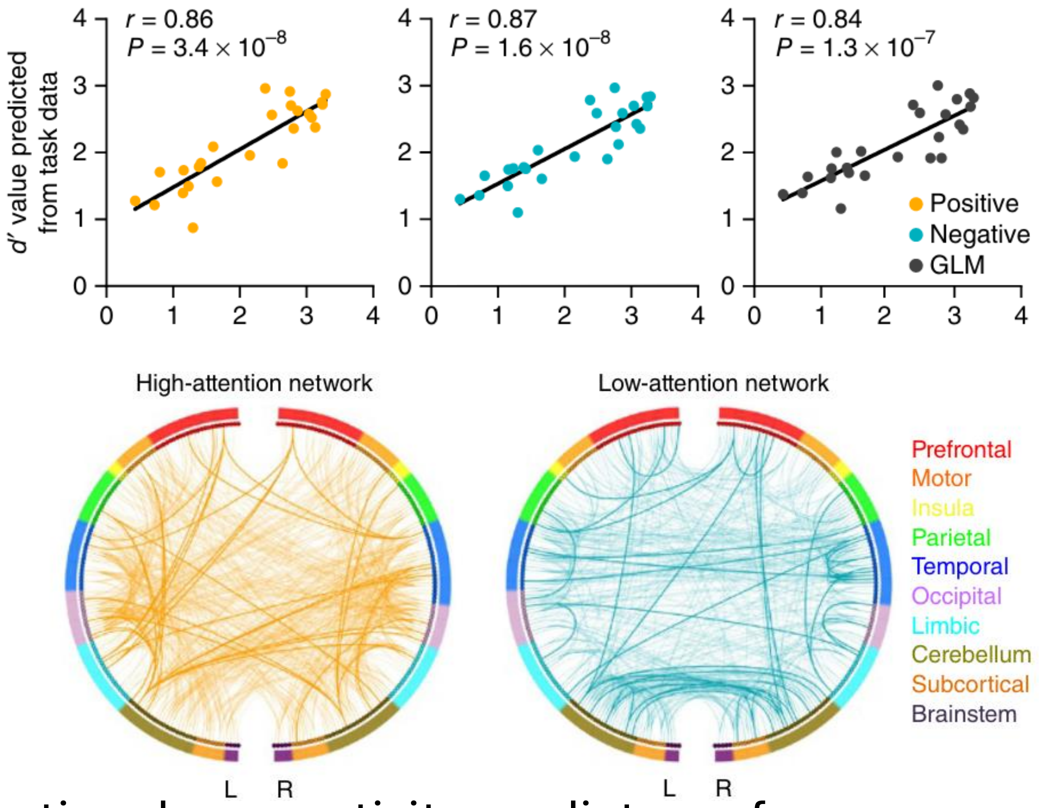


→ Without CV, effects are overestimated in discovery sample
→ Cross validation un-biases effect estimates

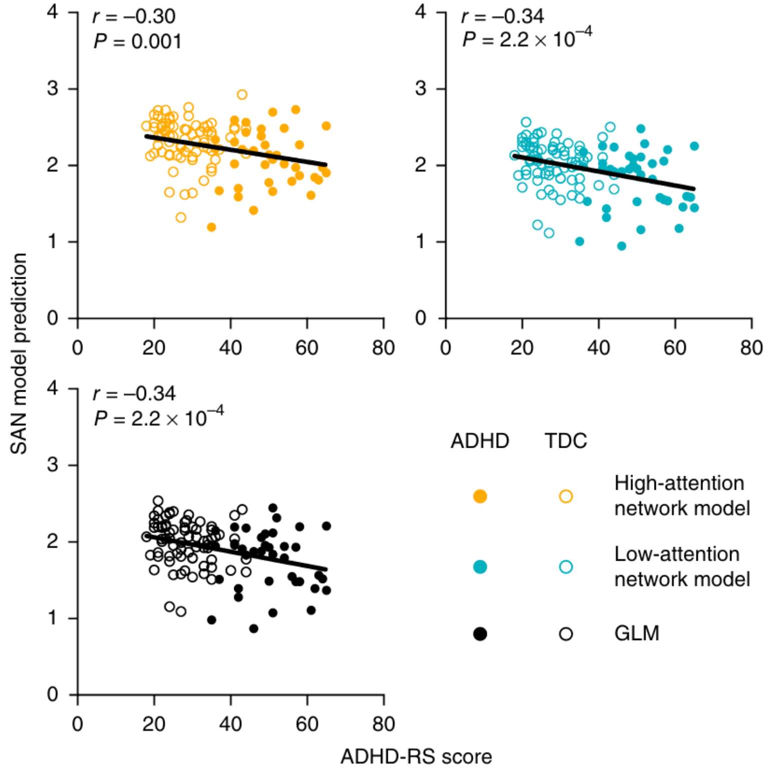
Spisak et al., *Nature* 2023

So what if I can’t recruit thousands of subjects?

→ Use better statistical methods: Generalize to independent sample



Functional connectivity predicts performance on sustained attention task in healthy volunteers



Model can also successfully predict ADHD scores in an independent sample

Rosenberg et al., *Nat Neuro* 2015

Wrapping it all up

- What are brain-behavior correlations?
  - Statistical association between some sort of brain measure and some sort of behavioral measure
- Why should we care about brain-behavior correlations?
  - They give us new insights into nuances behind neural processes and support the move towards brain-based psychiatry
- How can we robustly study brain-behavior correlations?
  - Increase your sample size and/or scan duration
  - Think critically about which measures you're going to collect
  - Improve data quality of both the brain and behavior
  - Use appropriate statistical/machine learning methods like cross validation and generalization in an independent sample

Acknowledgements

Section on Functional Imaging Methods

- Peter Bandettini
- Dan Handwerker**
- Javier Gonzalez Castillo
- Pete Molfese
- Burak Akin
- Josh Faskowitz
- Sharif Kronemer
- Fernando Ramirez
- Isabel Gepart

- Tori Gobo
- Josh Dean
- Marly Rubin
- Plyfaa Suwanamalik-Murphy
- Stephanie Swegle

- Jan Kadlec
- Michal Ramot
- Jesse Rissman

- Robert Bilder
- Jean-Baptiste Pochon
- Agatha Lenartowicz
- Kristen Enriquez
- Holly Truong
- Sandra Loo
- Catherine Sugar
- Carrie Bearden

Questions? catherine.walsh@nih.gov