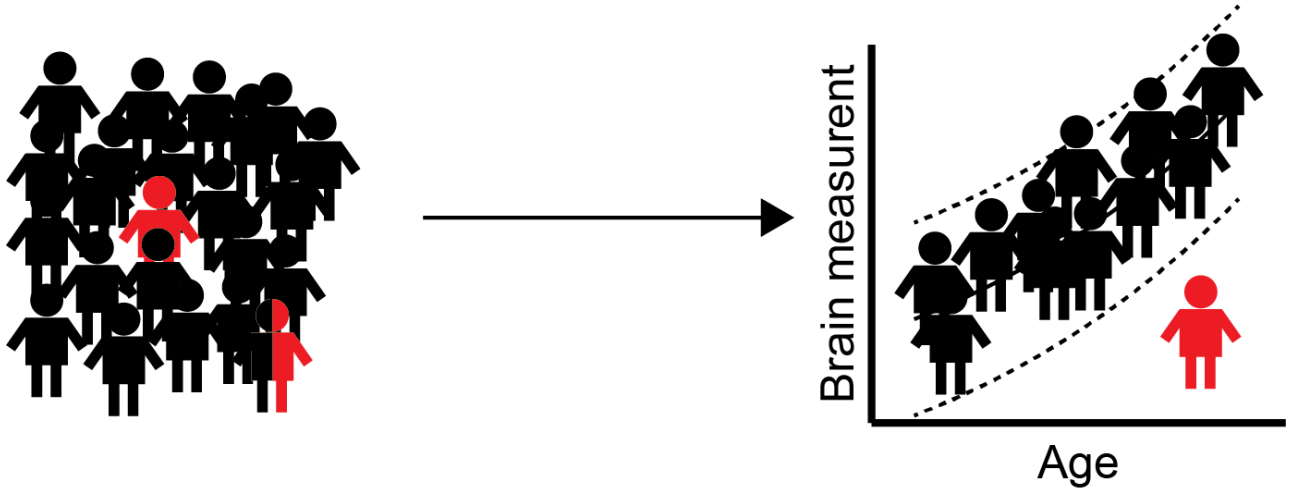
FUNCOIN: A whole-brain functional connectivity regression method for normative modelling

Janus Rønn Lind Kobbersmed, MD, BaSc, Phd-student, Center of Functionally Integrative Neuroscience, Aarhus University
Professor Diego Vidaurre, Center of Functionally Integrative Neuroscience, Aarhus University

Analysis workshop, May 26 2025, Aarhus University

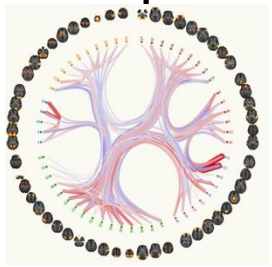


INDEPENDENT RESEARCH FUND DENMARK
AARHUS UNIVERSITY

Motivation:

Brain function biomarkers of disease/disorder

- Challenges:
  - Complex nature



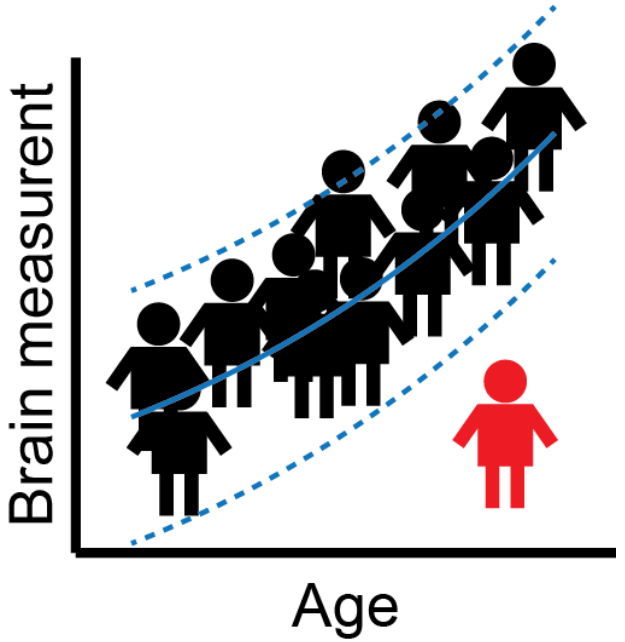
  - Heterogenous symptoms

  - Described on a spectrum

  - Hard to recruit cases

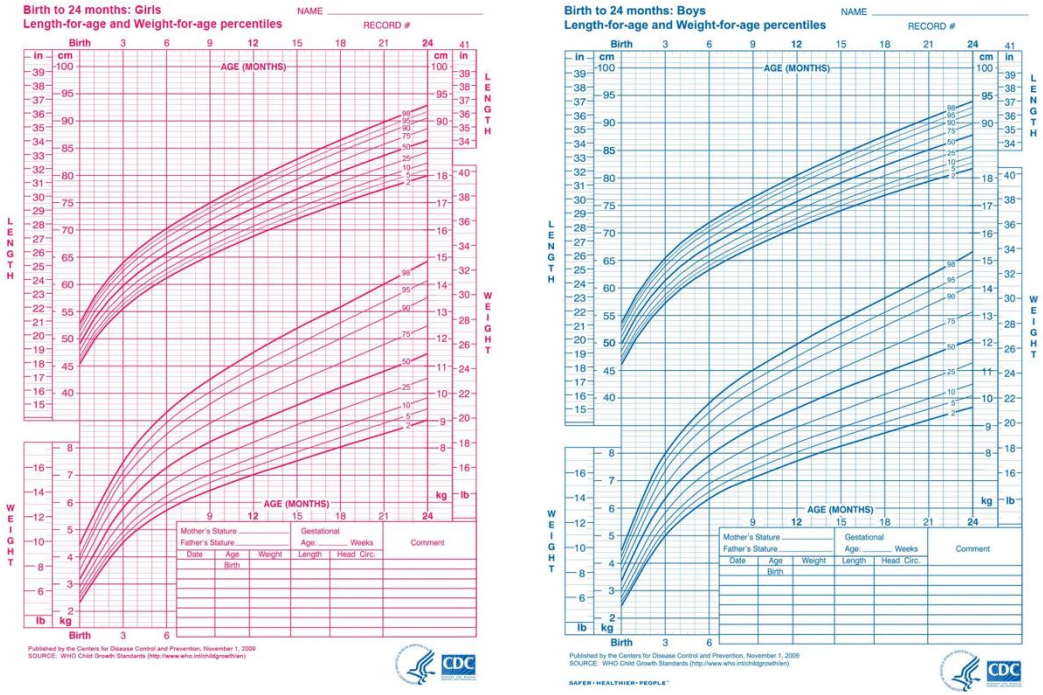
- Suggestion:

Normative modelling



Normative modelling

- Like growth charting of children
- Describes the normal variation in height and weight given sex and age
- Individual measures are evaluated as percentiles or Z-scores
- Allows identifying outliers or spotting longitudinal effects

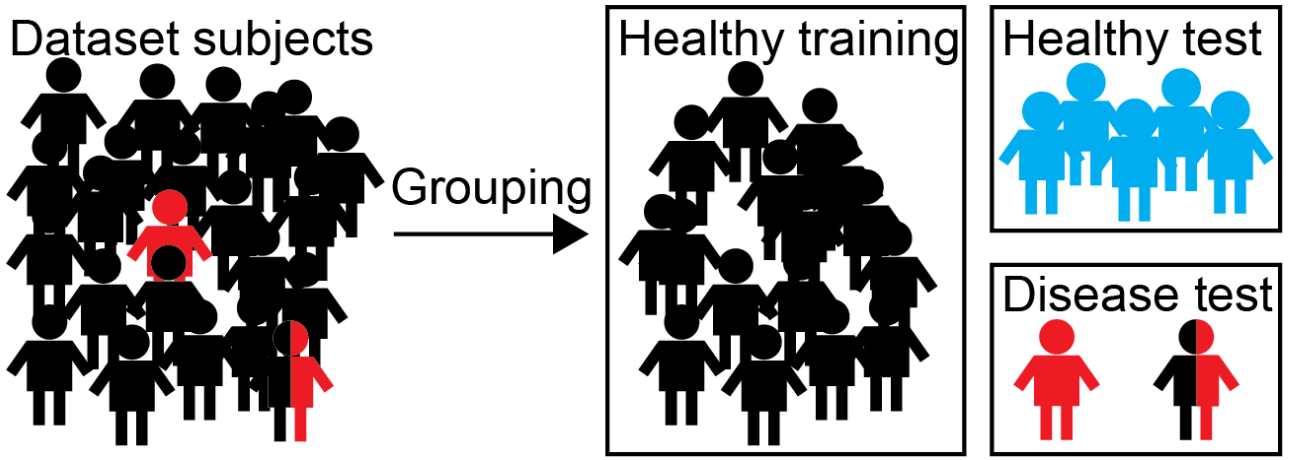


The W.H.O. growth charts for girls (left) and boys (right) from birth to age 2, published by the *C.D.C. (Centers for Disease Control and Prevention), 2009*

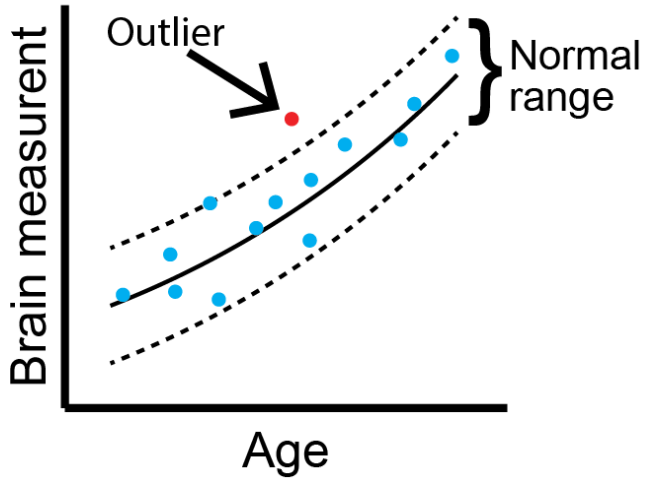
Elements in normative modelling

1. A large dataset of healthy subjects and (possibly fewer) subjects with diagnoses
2. A method to predict the brain meassure from sex and age

1.

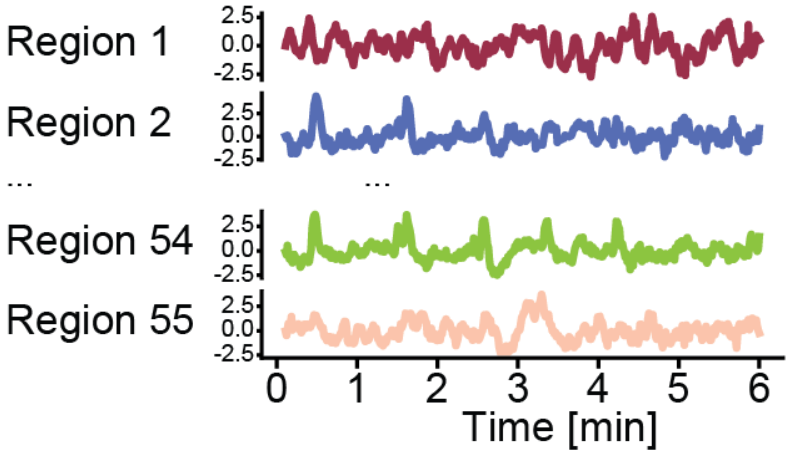


2.



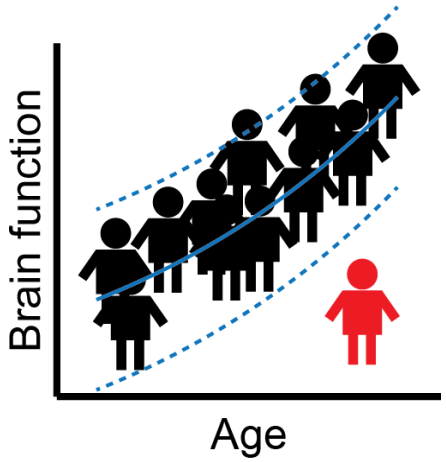
UK Biobank, fMRI data

- rs-fMRI from >60k subjects
  - 10684 with diagnoses in ICD10 F or G
  - 49721 with no diagnoses in ICD10 F or G
- Scan time 6 minutes (TR = 0.735 s)
- Network parcellations from ICA (p=55)

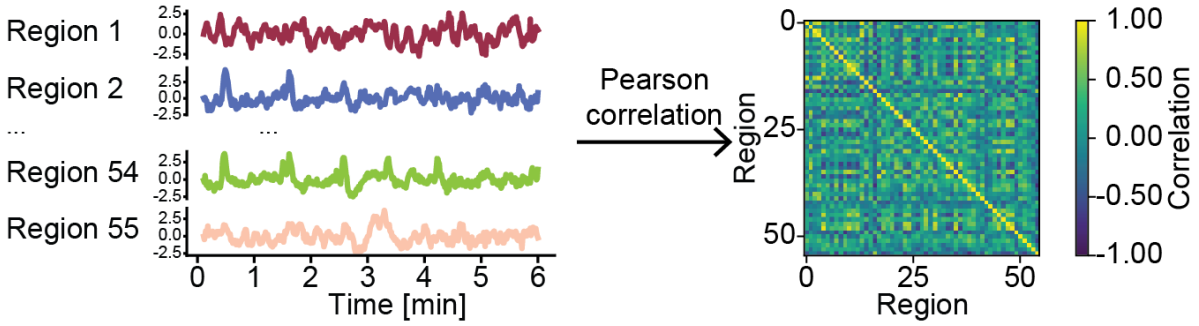


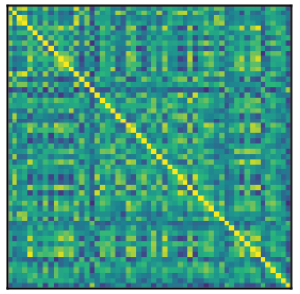
Normative modelling of brain function

Goal:

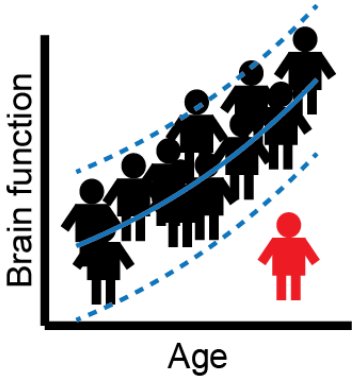


Assessing brain function with functional connectivity (FC)





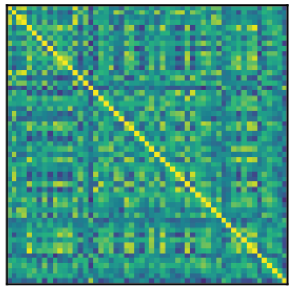
?



that is:

(age, sex)

?



FUNCOIN: Functional Connectivity Integrative Normative Modelling[1]

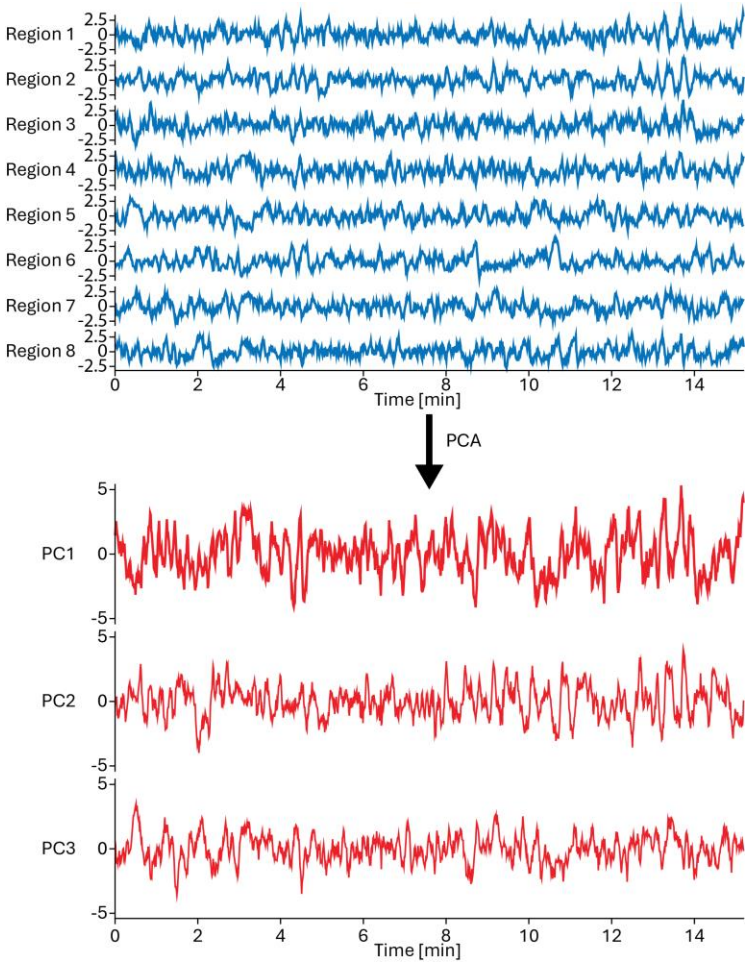
- A novel framework for network-level normative modelling
- An adaptation of a recently developed covariance regression method: Covariate Assisted Principal Regression[2]
  - Modified for normative modelling
  - Dimensionality reduction and regression in one go
- Released as a ready-to-use Python package

[1]: Kobbersmed, J.R.L. et al. (2025). 'One-shot normative modelling of whole-brain functional connectivity', *BioRxiv*.

[2]: Zhao, Y. et al. (2021). 'Covariate Assisted Principal regression for covariance matrix outcomes', *Biostatistics*, 22(3), pp. 629-45.

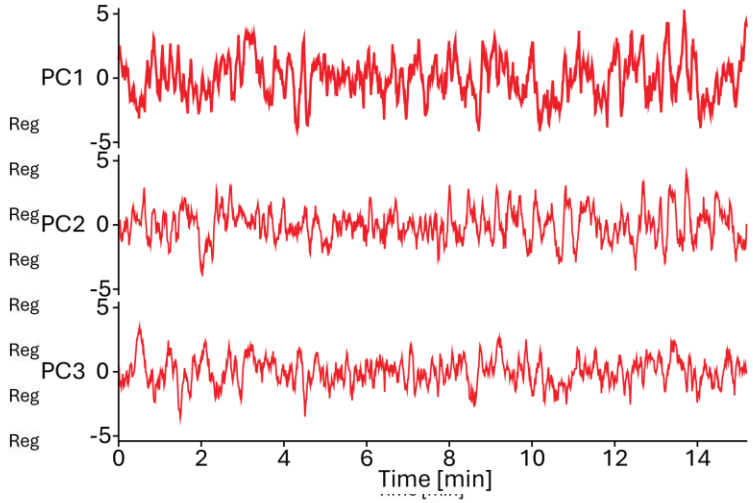
Background: Principal components analysis

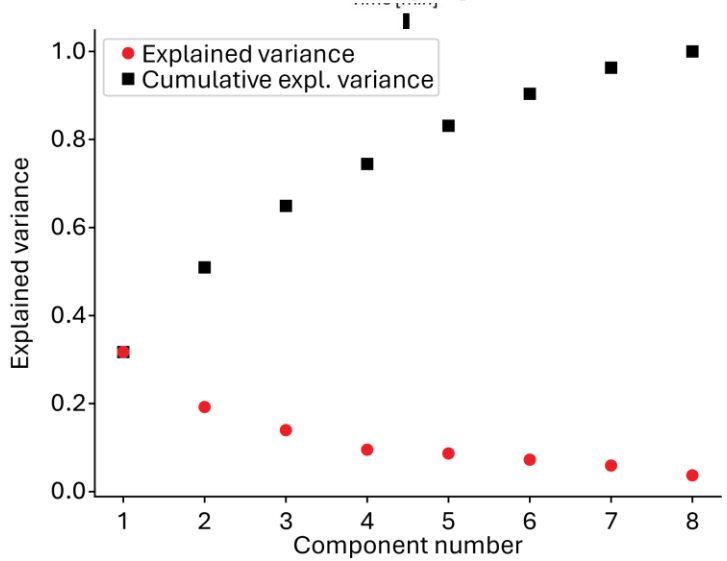
- We identify the components that **capture most variation** in the data
- => Benefits:
  - We **reduce data to fewer dimensions**
  - The component loadings reveal **which brain regions contribute most** to each component



Background: Principal components analysis

- We identify the components that **capture most variation** in the data
- => Benefits:
  - We **reduce data to fewer dimensions**
  - The component loadings reveal **which brain regions contribute most** to each component
  - From the **strength** of a component we can compute how much variance the component explains



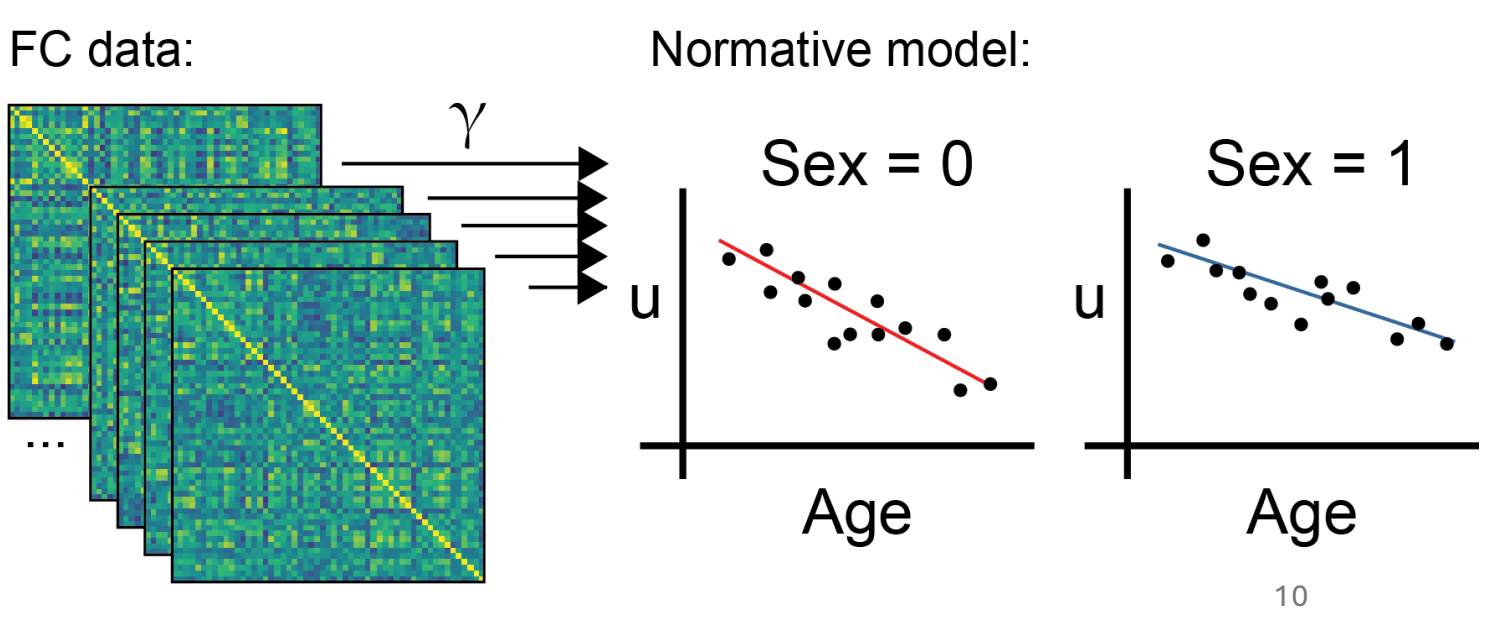


Functional Connectivity Integrative Normative Modelling (FUNCOIN)

- Identifies ***components*** in a way similar to PCA
- The identified components are ***shared*** among the subjects
- The ***strength*** of the components ***depend on covariates*** (e.g. sex and age)

Model:

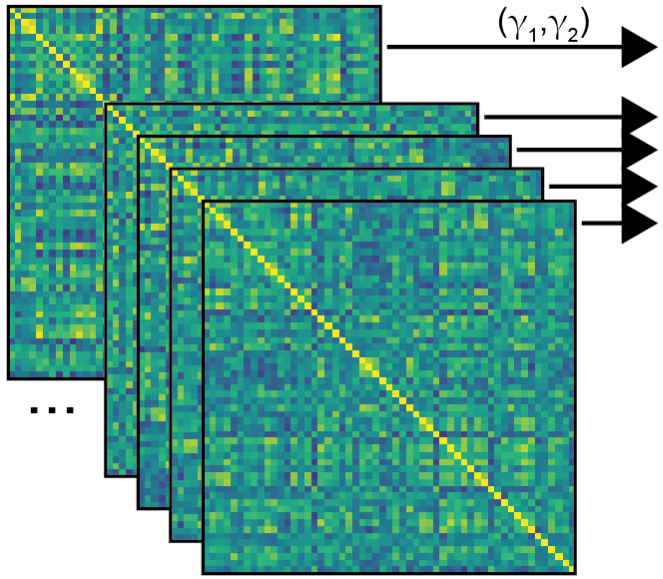
$$u_i = \log(\boldsymbol{\gamma}^T \boldsymbol{\Sigma}_i \boldsymbol{\gamma}) = \beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta}$$



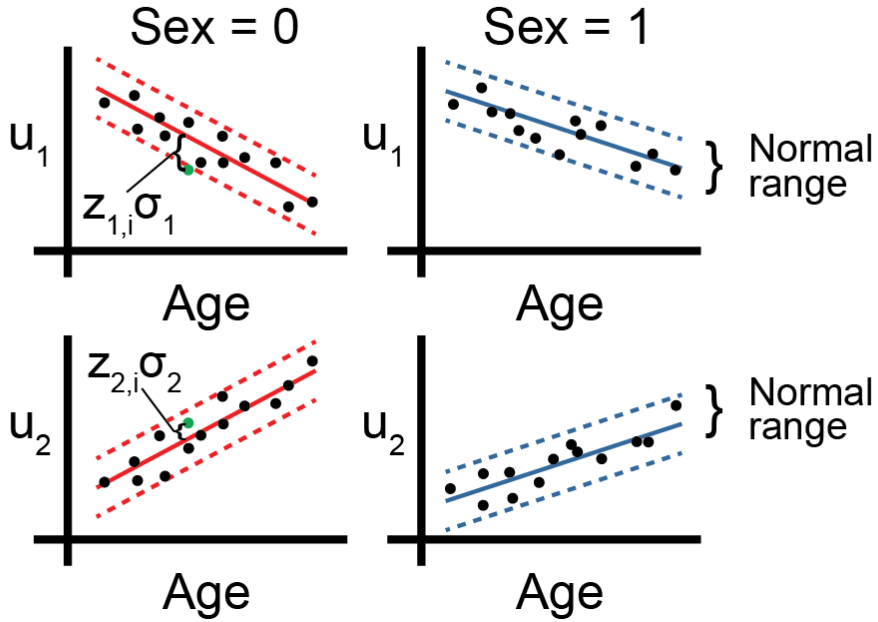
Normative modelling of FC

- Healthy training: n=32000; healthy test: n=14000
- Deviation is assesed as two Z-scores

Healthy training FC:

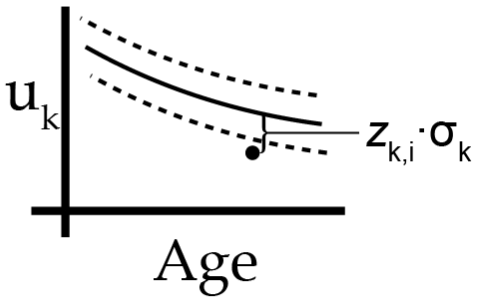


Normative model:



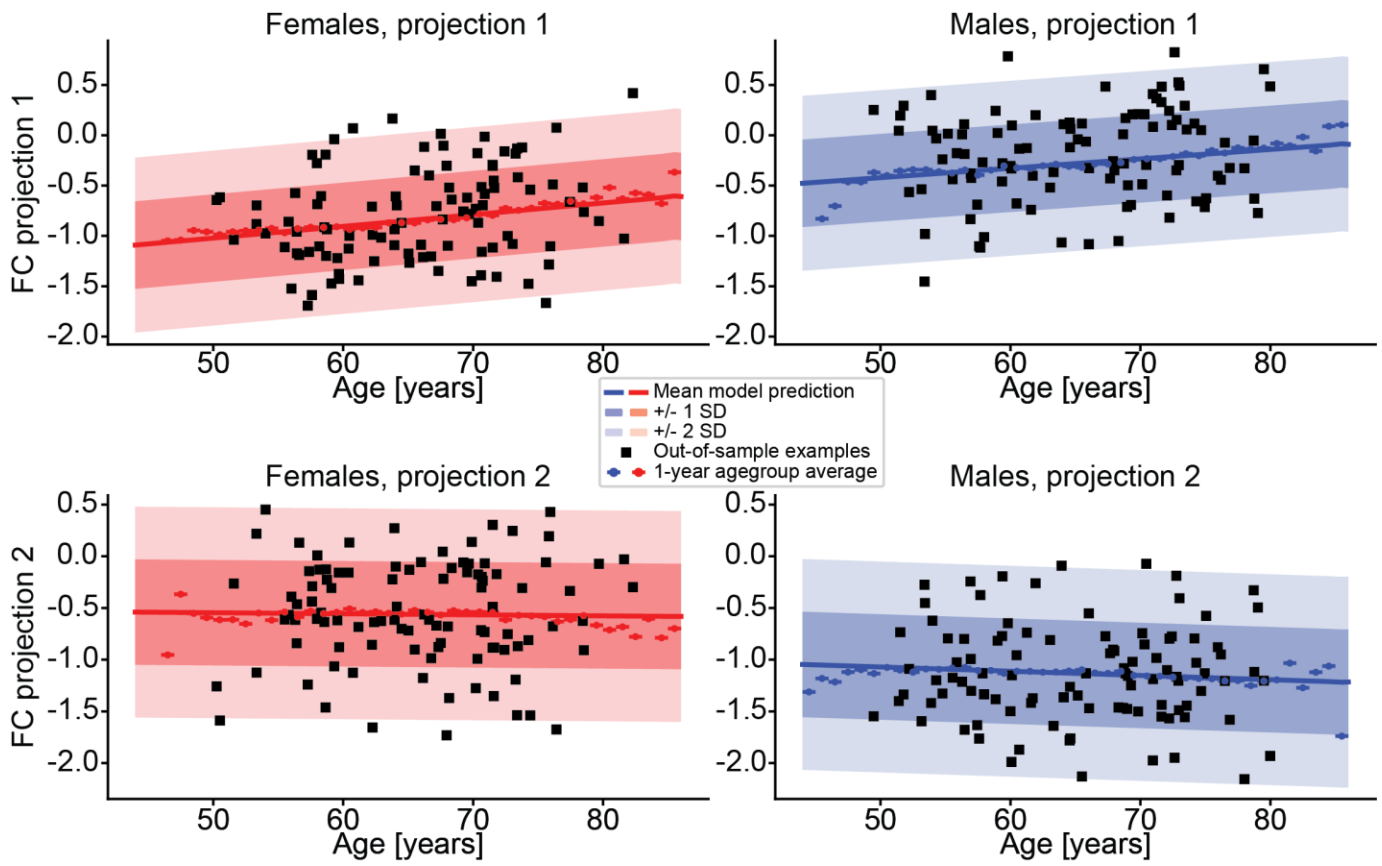
Individual deviation:

$$z_{k,i} = \frac{u_{k,i} - \overline{u_k}}{\sigma_k}$$



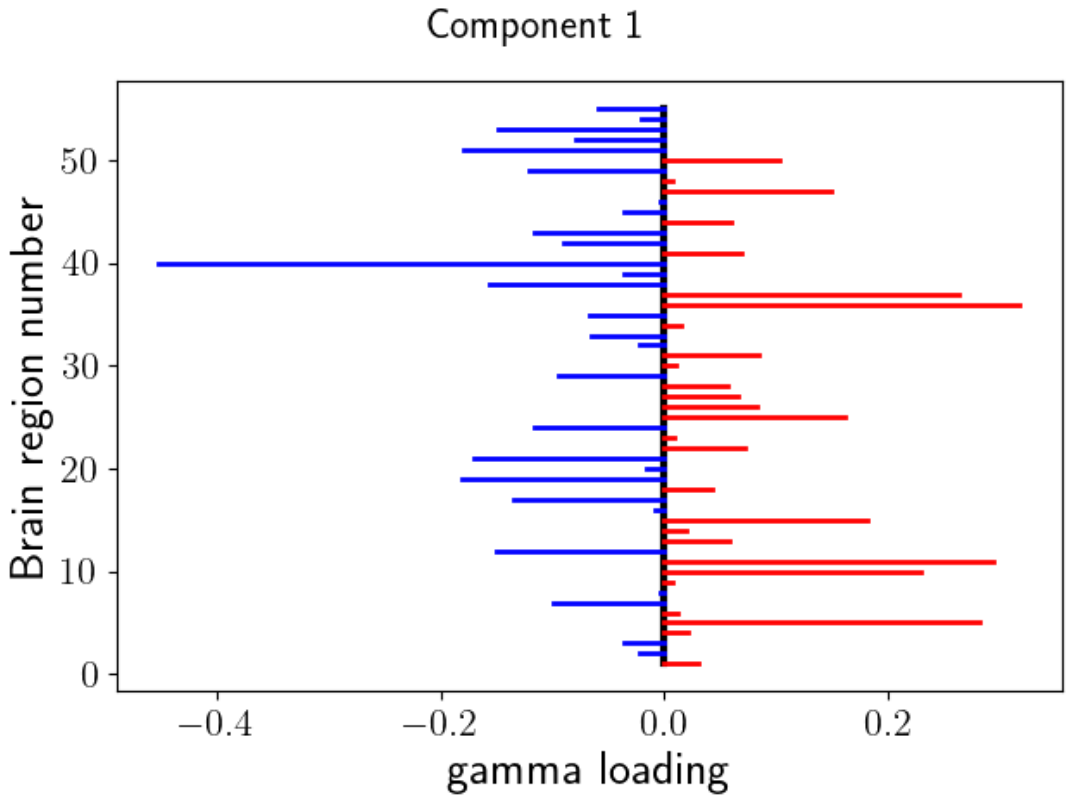
Results: Normative models with FUNCOIN

- We identify 2 components:
  - 1: Depends on sex and age
  - 2: Depends (mainly) on sex



Results: Inspecting components and coefficients

- The weighting of the brain regions are the values of $\boldsymbol{\gamma}$s
- The association with covariates is seen from the $\boldsymbol{\beta}$ coefficients

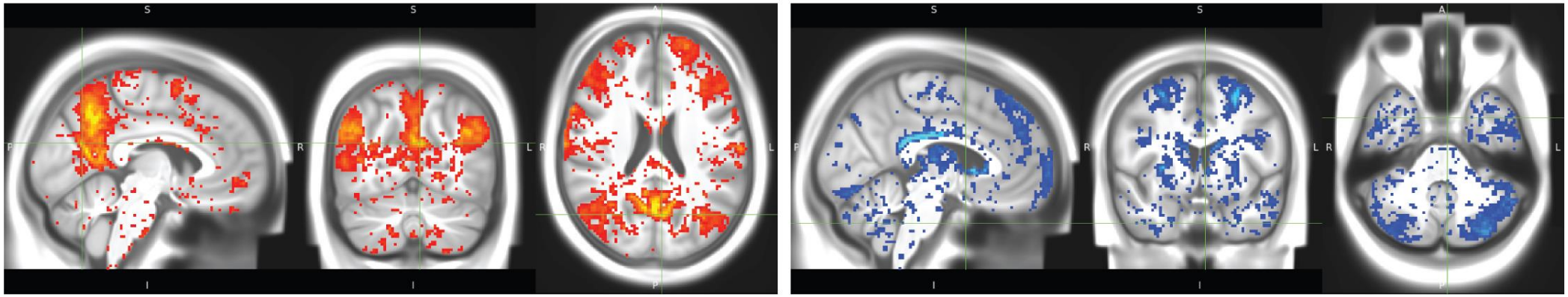


	Coeff.	Variable	Estimate (SE)	95%-CI
First projection	$\beta_{1,0}$	Intercept	-1.085 (0.010)	[-1.101; -1.066]
	$\beta_{1,1}$	Sex	0.6129 (0.014)	[0.585; 0.641]
	$\beta_{1,2}$	Age	0.4755 (0.018)	[0.440; 0.511]
	$\beta_{1,3}$	Sex-age	-0.0935 (0.026)	[-0.145; -0.043]

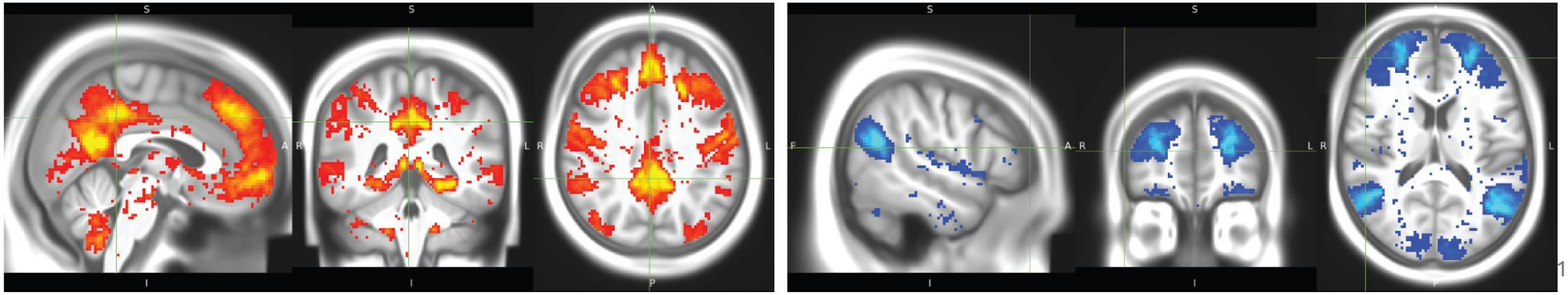
Results: Sex and age components

- Components illustrated with brain maps:

Projection 1: Dependents on sex and age

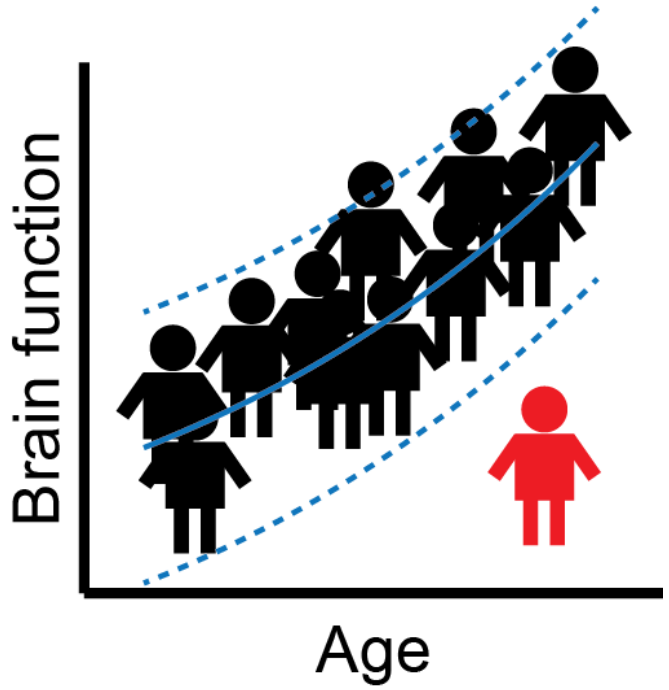


Projection 2: Depends on sex



Diagnosis prediction

- We consider |Z|>2 as an outlier
  - Expect 2,3 % of healthy subjects to be outliers in each direction
- Reasons for being an outlier
  - Coincidence
  - Non-identified pathology
  - Brain disease/disorder



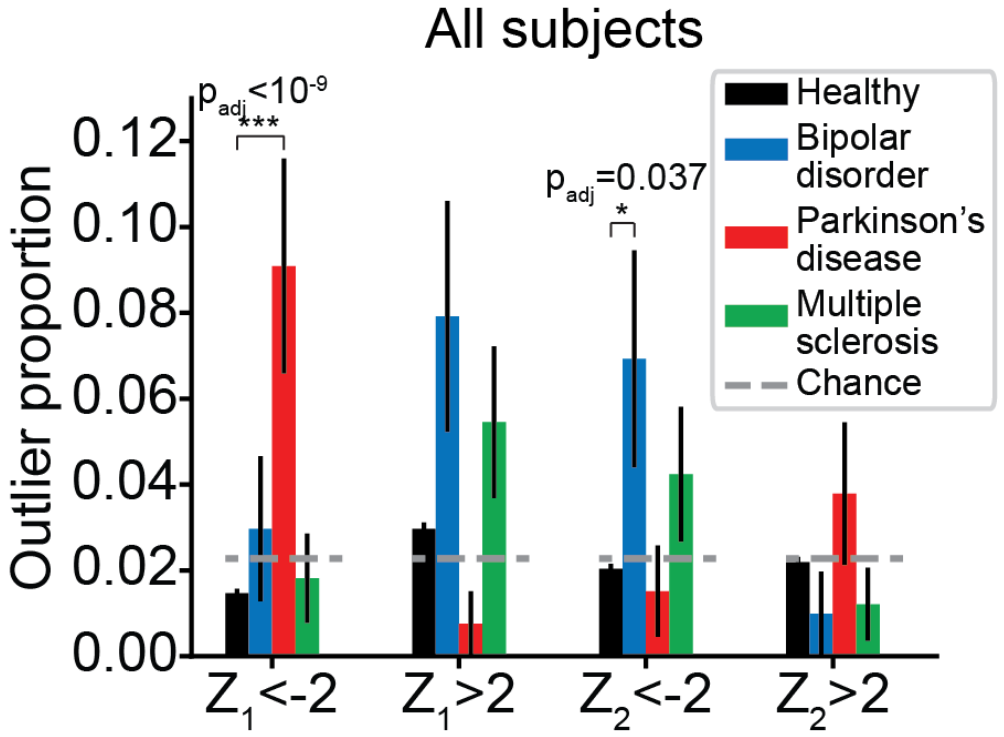
Diagnosis data

- Focus on chronic conditions
- Focus on N>100

Diagnosis	N
Alzheimer's dementia (F00)	34
Vascular dementia (F01)	25
Unspecified dementia (F03)	55
Mild cognitive impairment (F06)	28
Schizophrenia (F20)	33
Bipolar disorder (F31)	**101**
Autism/Asperger's (F840 and F845)	24
Parkinson's disease (G20)	**132**
Alzheimer's disease (G30)	48
Multiple sclerosis (G35)	**164**

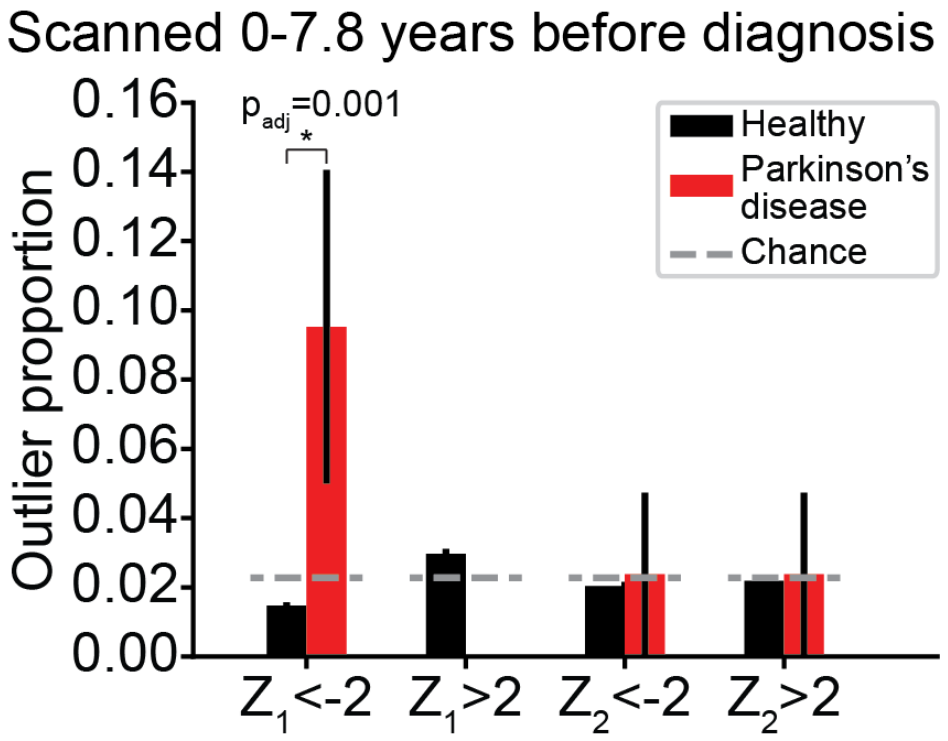
Diagnosis prediction

- FUNCOIN reveals significant FC alterations in Parkinson's disease and bipolar disorder
- Subjects with bipolar disorder divided into two groups



Early diagnosis prediction

- Some subjects were diagnosed with PD years after their scan
  - 54 subjects were scanned 0-8 years before PD diagnosis (average ~3.7 years before)



Future work

- Currently finishing a version of FUNCOIN for MEG/EEG
  - Spectral properties
  - High dimensionality (large p)
  - Large amount of time points

See more:

- Paper: https://www.biorxiv.org/content/10.1101/2025.01.13.632752v1
- The FUNCOIN Python package: https://github.com/kobbersmed/funcoin
- FUNCOIN tutorial for this workshop: https://github.com/CFIN-analysis/analysis_workshop_26May/blob/main/Notebooks/4_Tutorial_funcoin.ipynb

Extra slides (leftout but feel free to ask)

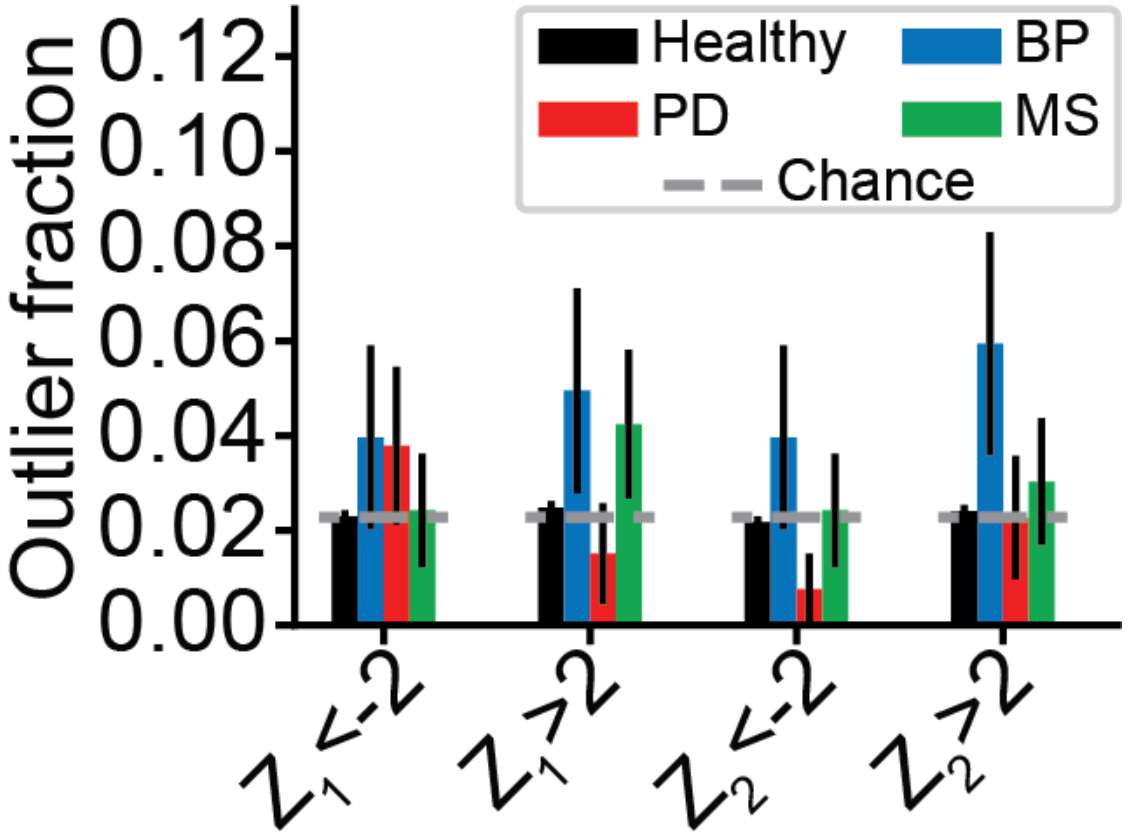
Covariate-assisted principal regression

- For each subject, *i*, let $y_{it} \in \mathbb{R}^p$, t $= 1, \ldots \mathrm{T_i}$ be independent, identically distributed multivariate, normal random variables with mean zero and variance Σ_i
- Assume that there exists a vector, $\gamma \in \mathbb{R}^p$, such that $z_{it} = \gamma^T y_{it}$ satisfies the following multiplicative heteroscedasticity model:

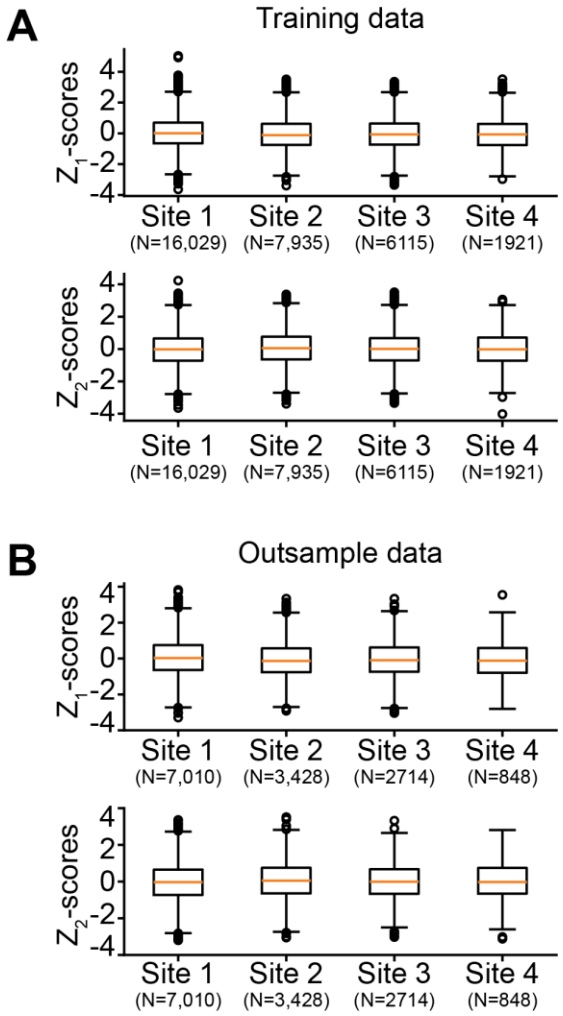
$$\log\big(Var(z_{it})\big) = \log(\boldsymbol{\gamma}^T \boldsymbol{\Sigma}_i \boldsymbol{\gamma}) = \beta_0 + \boldsymbol{X}_i^T \boldsymbol{\beta}$$

where $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^{q-1}$ model coefficients.

Disease prediction – elementwise model



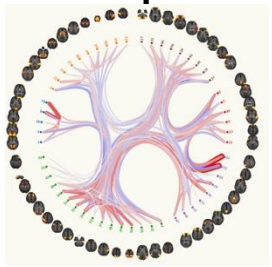
Taking scanning site into account



Benefits of normative modelling

Brain function biomarkers of disease/disorder

- Challenges:
  - Complex nature

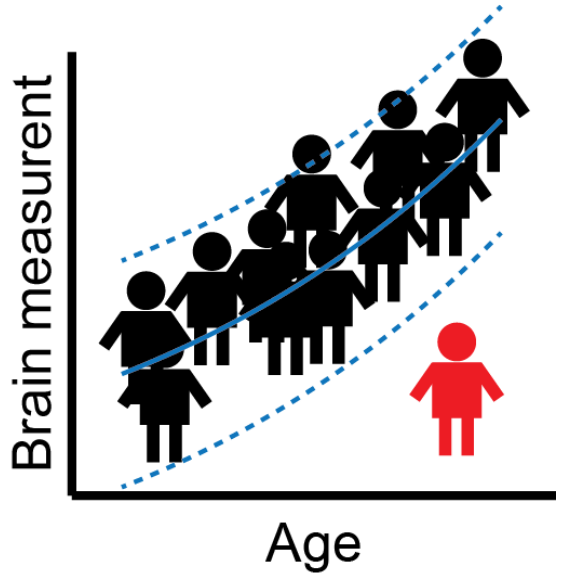


  - Heterogenous symptoms

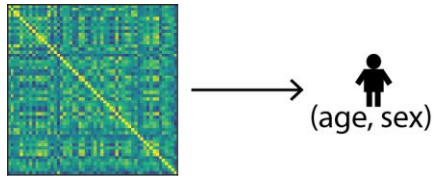
  - Described on a spectrum

  - Hard to recruit cases

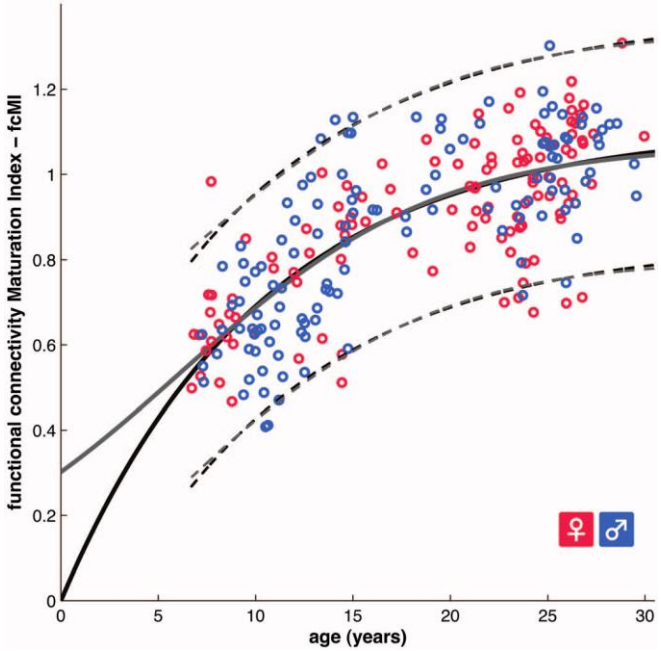
- Benefits of normative models:
  - Subject-level inference
  - Inference is per definition on a spectrum
  - Does not require a large number of cases



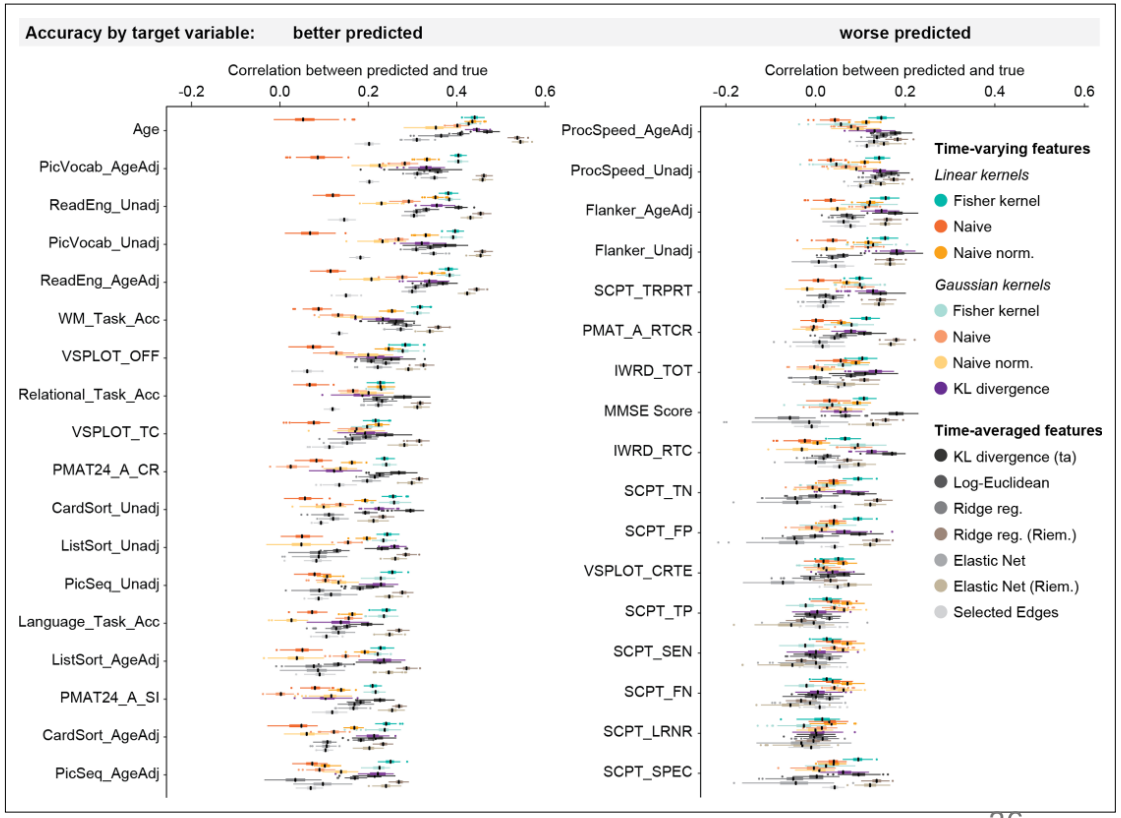
Prediction from FC



- Age and other traits can be predicted from rsfMRI FC, e.g. with
  - Linear models
  - SVMs (and other ML)
  - HMM



From: Dosenbach, N.U. et al. (2010). Prediction of individual brain maturity using fMRI. Science.

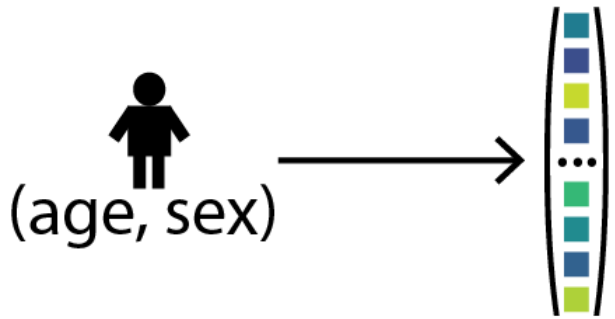


From: Ahrends, C. et al. (2025). Predicting individual traits from models of brain dynamics accurately and reliably using the Fisher kernel. eLife.

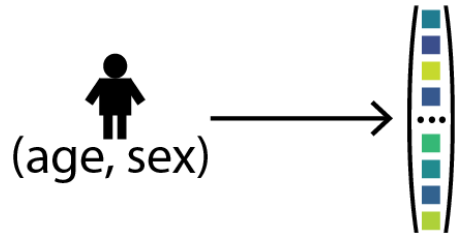
Previous normative models in rsfMRI FC

- Normative models of FC have predicted **individual connections**
  - Gaussian-Gamma mixed models (Looden et al, *Molecular Autism*, 2022*)*
  - Warped Bayesian Linear Regression (e.g. Fraza et al, *Neuroimage*, 2021)
  - Gaussian Processes (e.g. Marquand et al, *Biological Psychiatry*, 2016)

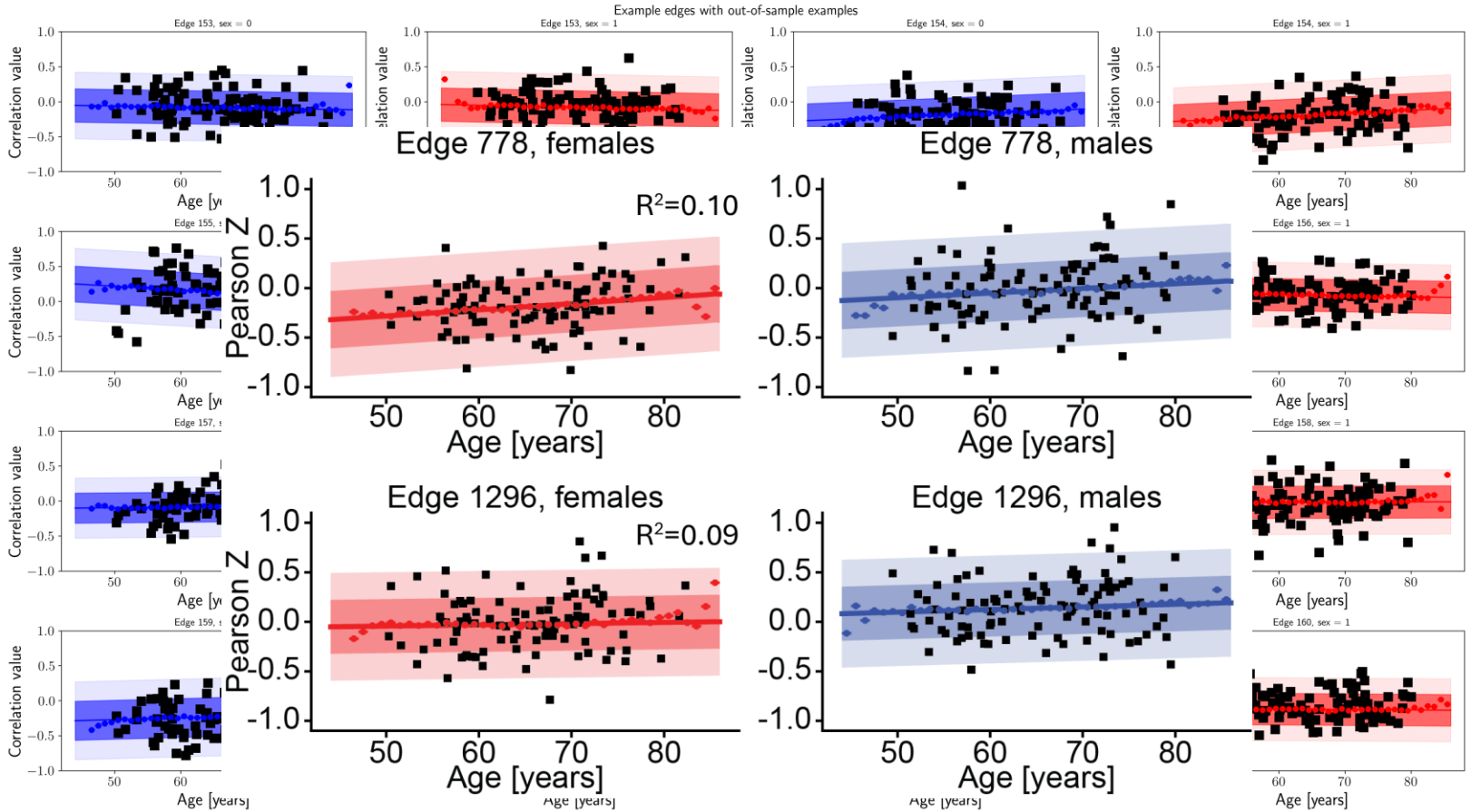
  - Good overview in: Marquand et al, *Molecular Psychiatry*, 2019



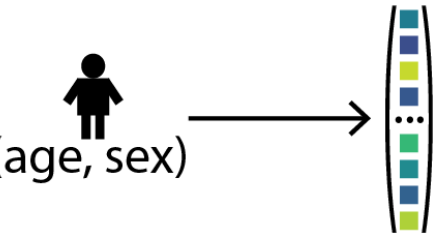
Elementwise FC prediction



- Predicting each correlation value (Fisher z) from sex and age
- Best ec

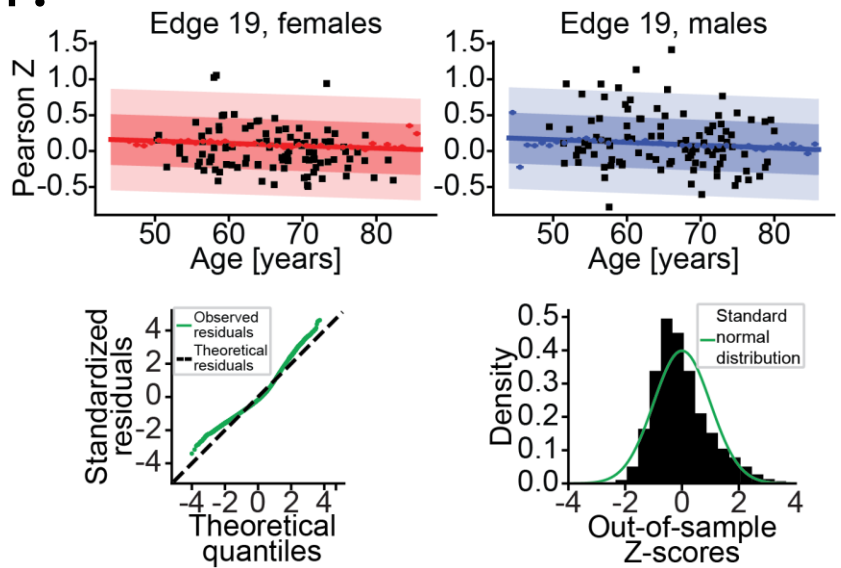


Problems in elementwise FC prediction

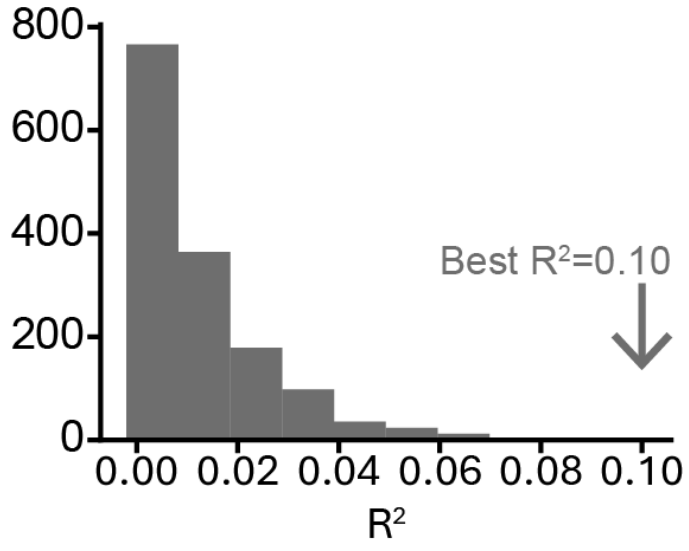


1. Massive amount of models to validate
2. Each element holds little information about sex and age
3. Massive hypothesis testing
4. Not predicting valid FC matrices

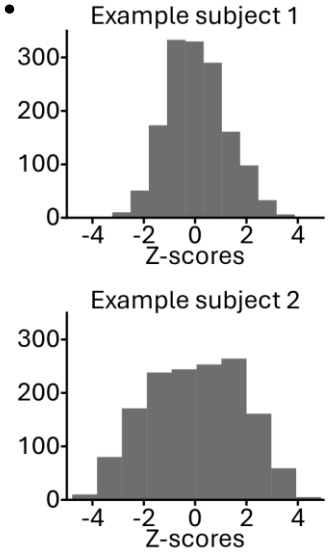
1:



2:



3:



Z-score distribution

- The out-of-sample Z-scores from healthy subjects follow a standard normal distribution