One-shot normative modelling of whole-brain functional connectivity

Janus R. L. Kobbersmed[a]*, Chetan Gohil[b], Andre Marquand[c,d], Diego Vidaurre[a]

[a]Department of Clinical Medicine, Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark

[b]Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[c]Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands

[d]Donders Institute for Cognition, Brain and Behavior, Radboud University, Nijmegen, The Netherlands

*Corresponding author

Abstract

Many brain diseases and disorders lack objective measures of brain function as indicators of pathology, which has recently spurred the use of normative modelling in neuroimaging. Normative models characterize the normal variation of brain measurements given sex and age, thereby allowing identification of abnormalities as deviations from normal. Normative modelling of brain function is typically based on predicting functional connectivity (FC) between each pair of brain regions. But the human brain is an extremely integrated organ, and brain disease often has widespread effects that are not well captured by piecemeal analyses, i.e. connection by connection. We propose Functional Connectivity Integrative Normative Modelling (FUNCOIN), developing a whole-brain normative model of FC from a large resting-state fMRI data set that captures whole-network-level changes associated with sex and age. This model can significantly, and substantially, uncover abnormal FC patterns in Parkinson's disease patients even on scans up to 5.5 years before diagnosis.

Introduction

While structural brain scans are widely used in clinical settings for identifying e.g. stroke and tumors, functional brain imaging modalities like fMRI and MEG are still primarily used in research (Zhang et al. 2021; Calhoun, Pearlson, and Sui 2021; Nour, Liu, and Dolan 2022). The search for brain function biomarkers of brain diseases and disorders is challenged by the complex nature of the brain as well as the lack of precision in diagnostic categories (Calhoun, Pearlson, and Sui 2021). Traditionally, studies have identified focal abnormalities associated with brain diseases, like loss of dopaminergic cells in substantia nigra resulting in lower striatal dopamine levels in Parkinson's disease (Lotharius and Brundin 2002). However, recent studies suggest that brain disease should be better explored by the consequences that known focal abnormalities have on wider brain function (Zhang et al. 2021; Fornito, Zalesky, and Breakspear 2015). The growing knowledge of the complex mechanisms leading from focal abnormality to network-level, functional dysfunction and, ultimately, observable symptoms on which the clinician bases a diagnosis, motivates the search for ways of incorporating brain function in clinical diagnostic procedures and decision making e.g. in treatment.

On these grounds, normative modelling has become increasingly popular particularly in structural, but also functional, brain imaging, as a way of identifying individuals that deviate from the healthy population (Rutherford et al. 2023; Marquand et al. 2019). Normative models of a given brain measurement aim to describe the normal variation of the measurement based on a large dataset from a (considered healthy) background population. These models predict the expected values of a set of measurements given covariates like sex and age, and quantify the variation in the background population by determining the distribution of the brain measurement around the predicted mean (Rutherford, Kia, et al. 2022). Individual deviance from the normative model, as a possible marker of disease, is quantified with a Z-score or a percentile of the expected distribution around the predicted mean. This way, normative models differ from the traditional case-control approach by relying on determining subject-level deviance from the population norm rather than group-level differences. For this reason, normative models have potential in rare diseases, where recruiting large groups of cases is impossible, or in diseases with very heterogenous patient groups, as is the case for many psychiatric and

neurological diseases (Marquand et al. 2016; Cuthbert and Insel 2013; Insel et al. 2010; Hyman 2010; Cai, Choi, and Fried 2020; Moriarity et al. 2022; Kapur, Phillips, and Insel 2012).

Based on the aforementioned idea that many brain diseases are better described at an integrative level, we here focus on the search for brain function biomarkers of disease by doing network-level normative modelling of functional connectivity (FC) from resting-state fMRI (rsfMRI). As a matter of fact, rsfMRI FC has proven a useful tool for understanding brain function and has led to insights on brain organization and the identification of several networks and subnetworks, e.g. (Smith et al. 2009; Margulies et al. 2016; Power et al. 2011; Smith et al. 2013; Rasero et al. 2018). Out of the normative modelling approach, previous studies have revealed sex differences in rsfMRI FC as well as changes in FC with age over the entire human lifespan (Edde et al. 2021; Deery et al. 2023; Scheinost et al. 2015; Lopez-Larson et al. 2011; Biswal et al. 2010). Machine learning methods predicting age (Dosenbach et al. 2010; Bi, Guan, and Tian 2024; Lund et al. 2022; Smith et al. 2019) as well as a range of other variables (Schulz et al. 2024; Griffin et al. 2024; Rasero et al. 2021) have revealed that the rsfMRI functional connectome contains information on non-imaging traits, and constitutes a subject-specific “fingerprint” that allows identification of individual subjects from their FC (Abbas et al. 2023; Finn et al. 2015). In patients, rsfMRI studies have revealed complex, network-level, disease-related changes in FC over a range of neurological and psychiatric diagnoses (Zhang et al. 2021), e.g. changes in thalamic, sensorimotor and cerebellar networks in subjects with Parkinson’s disease (Tinaz et al. 2016; Gratton et al. 2019) and the lack of anticorrelation between medial and dorsolateral prefrontal cortex in bipolar disorder (Chai et al. 2011; Zhang et al. 2021).

FC has been the basis in all previous normative modelling studies with rsfMRI. To formulate a normative model, one must predict FC from covariates — typically sex and age, and draw inference from the residual variance caused by other sources including individual traits and brain diseases (Rutherford, Kia, et al. 2022). Most previous normative models have been based on element-wise regression methods (Looden et al. 2022; Rutherford et al. 2023; Fraza et al. 2021; Jalbrzikowski et al. 2019; Jiang et al. 2024), while a few have predicted the region-wise FC strength after aggregating the FC elements of each brain region (Sun et al. 2023; Liu et al. 2024), or applied dimensionality reduction on the FC matrix and then predicted

the independent components of the matrix (Kessler, Angstadt, and Sripada 2016). Prediction of individual connections from covariates like sex and age is challenging because of the high dimensionality of the FC matrix, with *p* brain regions resulting in $p*(p-1)/2$ unique elements; this leads to a multiple comparisons issue and a lack of capacity to generalize to network properties of the matrix. Additionally, all the above approaches ignore the special mathematical structure of the FC matrix, i.e. the fact that it is semi-positive definite.

In this paper, we propose Functional Connectivity Normative Modelling (FUNCOIN), a flexible, consistent, and scalable framework for normative modelling of full-matrix FC. FUNCOIN is an adaptation of a recently developed method from the statistics literature, covariate-assisted principal regression (Zhao, Wang, et al. 2021), which performs dimensionality reduction and regression of covariance matrices. In essence, FUNCOIN, implemented in a Python package, finds FC components (i.e. linear projections) that are maximally related to sex and age (or other specified covariates) while respecting the mathematical (positive-semi definite) structure of the FC matrix, in a way that is appropriate for normative modelling. This provides a parsimonious representation of the FC matrix using a few relevant components, thus avoiding mass hypothesis testing, which is a major limitation of current normative modelling approaches on FC.

Using data from UK Biobank (UKB), we use FUNCOIN to create a normative model of FC from 32000 individuals and test it on 14000 healthy subjects. We show that our method is scalable, generalizes well, and is well suited for describing a healthy background population as required for normative modelling. Then, we demonstrate our model's capability of identifying brain diseases: bipolar disorder (BD), Parkinson's disease (PD), and multiple sclerosis (MS). Remarkably, PD subjects were identifiable in our model years before being diagnosed. Finally, we compare to an element-wise approach and identify several challenges of element-wise methods that our method, modelling the full FC matrix, can mitigate.

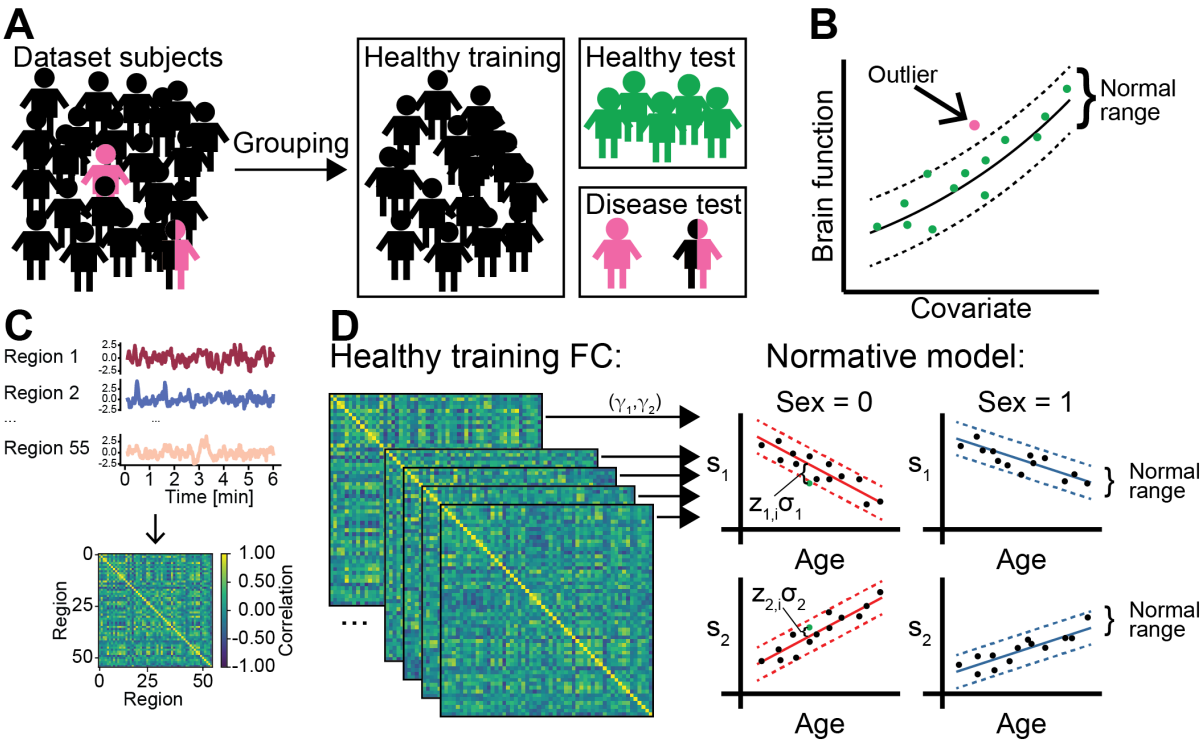

Figure 1: Schematic illustration of normative modelling with FUNCOIN. A. The input dataset has measurements from a large group of healthy and non-healthy subjects from the UK Biobank. We train our model on a large subset of the healthy subjects (training data, black, N = 32k). The generalizability of the model is evaluated on a smaller, out-of-sample subset of the healthy subjects (test data, green, N =14k). The ability of the model to identify diagnosed subjects is evaluated on the available data on subjects with brain diseases (pink). **B.** The training data produce an estimation of a normative distribution of the brain measurements (mean +/- 2 SD), which determines the normative range of the brain measurements and allows identifying outliers. **C.** Quantification of brain function. We aim at estimating a normative model of rsfMRI FC, calculated as the Pearson correlation of the time series from each pair of ICA components/brain regions. **D.** FUNCOIN identifies two projections (γ_1,γ_2) in the training group such that their magnitude follow a (2-dimensional) linear model (after taking the logarithm, see Methods). Individual (out-of-sample) subject deviation is determined as Z-scores (green points and annotation).

Results

FUNCOIN: A framework for normative modelling of FC

The aim of this study is to perform normative modelling of rsfMRI FC. For this, we need: 1) A large rsfMRI dataset of subjects representing a healthy background population as well as subjects that are expected to deviate from the healthy cohort in specific dimensions due to brain disease (Rutherford, Kia, et al. 2022); and 2) a method to predict FC as a function of sex and age.

We use rsfMRI data from UKB, which consists of rsfMRI time series from more than 60,000 people (illustrated in Figure 1A, left). We use the UKB-parcellation with 55 components from independent component analysis (ICA) (see Methods, (Beckmann and Smith 2004). In this setting, we consider a subject healthy if they have no psychiatric or neurological diagnoses (group F or G diagnoses in

International Classification of Disease 10 (ICD10)), which is the case for approximately 50000 of the subjects. The dataset is therefore well-suited for normative modelling. We divide the subjects into three groups (Figure 1A, right)(Rutherford, Kia, et al. 2022): A large group of healthy subjects to train the model on (training data, N=32,000), a set of healthy subjects on which we test the out-of-sample performance of the model (test data, N=14,000), and a group of subjects who have psychiatric or neurological diagnoses. We focus on diseases with more than 100 subjects in the dataset: BD (N=101), PD (N=132) and MS (N=165).

In general, a normative model of a brain measure predicts the mean (Figure 1B, solid black curves) and standard deviation (Figure 1B, dashed black curves illustrate +/- 2 standard deviations (SD)) in the healthy population as a function of covariates like sex and age. The model is evaluated on out-of-sample test data (Figure 1B, green points) to assess the ability of the model to generalize to healthy subjects unknown to the model thereby making sure that the normative model is representative of the background population. Subject-level deviation from the average is calculated as Z-scores (deviation from prediction in the unit of number of standard deviations). A Z-score of $|z|>2$ is considered an outlier, i.e. a deviation more extreme than expected from the natural variation in the healthy population (Figure 1B, pink point, see Methods).

Our measure of brain function is FC, determined in this case as the Pearson correlation matrix from all pairs of the UKB ICA parcellation's 55 components (akin to ROIs) (Figure 1C)(Beckmann and Smith 2004). The FC matrix is high-dimensional and positive semidefinite. To handle this, we propose FUNCOIN, an adaptation of a recently developed covariance regression method, covariate-assisted principal regression (Zhao, Wang, et al. 2021), for normative modelling. This method identifies components of the FC matrices across subjects, i.e. projections whose magnitude maximally depend on sex and age.

By fitting the model to the training data (Figure 1D left), we chose to identify two projections (γ_1 and γ_2, Figure 1D middle) of the correlation matrices as well as the regression coefficients that relate these projections to age and sex (through a log-linear model; see Methods). This gives rise to prediction of the average component strength given sex and age as well as the normal range in the training data (Figure 1D right). With the out-of-sample healthy test subjects, we assess

explanatory power and generalizability of the model, while the Z-scores, (z_1,z_2), of the subjects with brain diagnoses are used to probe the model's ability to identify subjects with brain disease as outliers.

We release a python package for using FUNCOIN, including training the model (identifying projections and coefficients), calculating a variety of model evaluations metrics, applying identified projections to out-of-sample subjects, calculating Z-scores, determining non-parametric and parametric confidence intervals, performing statistical testing, and simulating data. The package is available at https://github.com/kobbersmed/funcoin and is installable via pip.

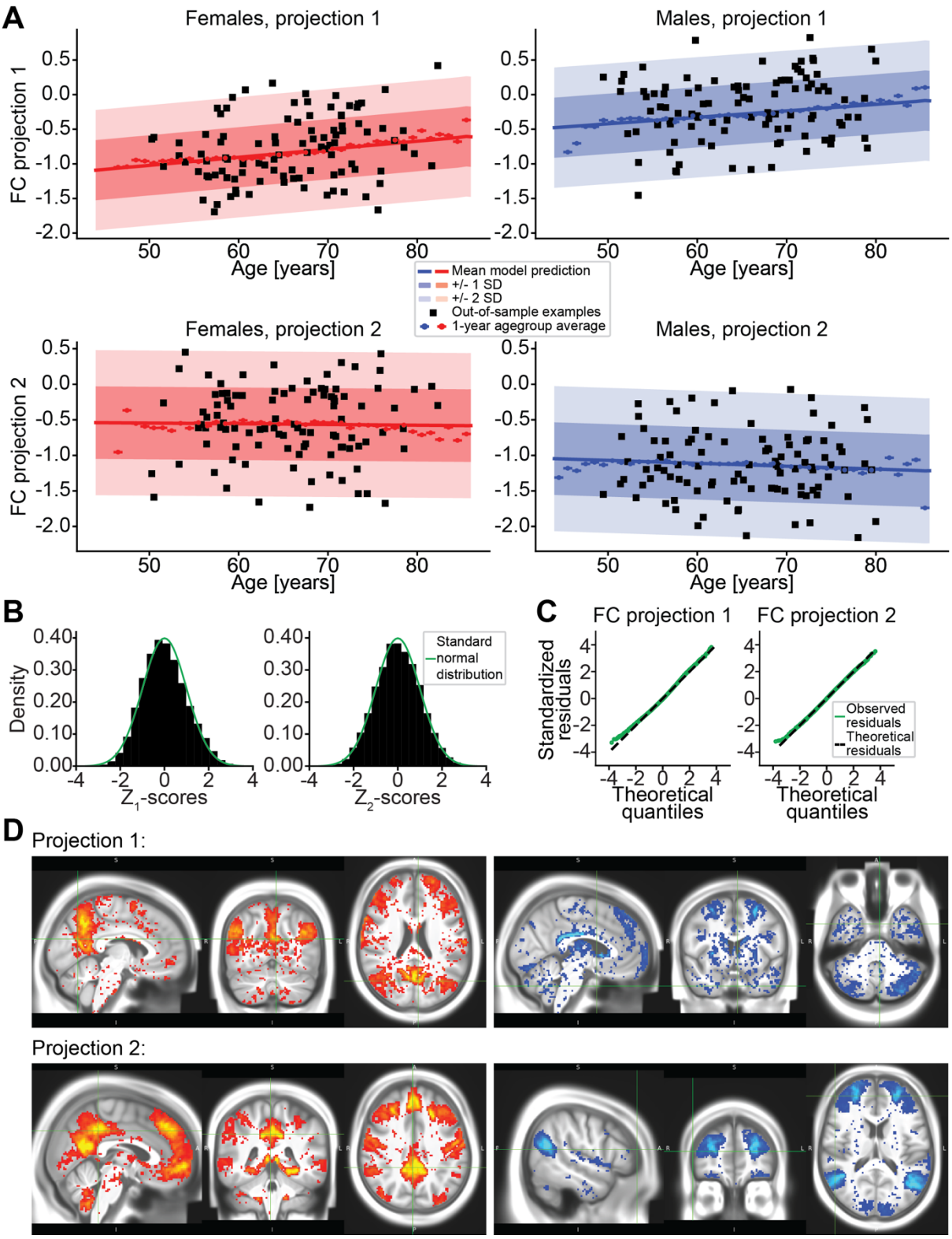

Figure 2: Normative modelling of a healthy population from the UK Biobank using FUNCOIN. A. Normative graphs from the identified projections of the training data. For each sex (females red, males blue), the normative graphs of the two projections (top and bottom) are plotted as a function of age. 100 random out-of-sample subjects for each sex are plotted (black) to show the spread in the background population. As observed, component 1 is associated with both sex and age, while component 2 is mainly associated with sex. **B.** Histograms of the out-of-sample Z-scores for the two projections. The bars are scaled to an area of 1 and show good agreement with the probability density function of a standard normal distribution, which visualizes the good generalizability of the model to out-of-sample subjects. **C.** QQ-plots of out-of-sample Z-scores (standardized residuals) show that the residuals are normally distributed, ensuring the statistical validity of the model. **D.** Brain maps showing the positive (warm

colors) and negative (cool colors) parts of each component. Component 1 primarily consists of default mode network areas, dorsolateral prefrontal cortex, thalamus, retrosplenial cortex, cerebellum, and striatal structures. Component 2 mainly consists of default mode network areas, temporoparietal junction, parahippocampal gyrus, and medial and dorsolateral prefrontal cortex.

Characterizing dependence of FC on sex and age

Normative modelling of FC requires consistent and accurate prediction of FC from sex and age. Simulation results and careful assessment of the model performance on real data showed that the covariate-assisted principal regression model is consistent but biased (see Methods, Figure S1A-B, Table S1-S2). This leads us to introduce FUNCOIN: an unbiased version of the previous method that we can use for normative modelling of FC (see Methods).

Using FUNCOIN, we identify two projections from the training data (n=32,000) with sex, age, and sex-age interaction as covariates. The log-transformed projected FC matrices give rise to a linear model (Figure 2A), where the first projection is associated with both sex and age (Figure 2A, top, R^2=0.32), and the second projection is associated mostly with sex (Figure 2A, bottom, R^2=0.24). The shaded areas show ±2 SD. Contrary to many other normative models (Bethlehem et al. 2020), the residuals in FUNCOIN were homogenous and normally distributed (see Methods, Figure S1C-D). This yields a more precise SD estimation and therefore allows better identification of outliers. See Figure S2 for γ loadings and Table S3 for β_0 and $\boldsymbol{\beta}$ coefficients and statistics.

For the model to be useful for normative modelling, it is crucial that it generalizes to out-of-sample healthy subjects, i.e. data which has not been utilized in the training of the model (black example points in Figure 2A). The Z-scores quantify the individual deviation as the difference between predicted and observed value in the unit of number of standard deviations. Thus, with normally distributed residuals, the Z-scores of the healthy population are expected to follow a standard normal distribution. Histograms of 14000 out-of-sample Z-scores normalized to an area of 1 for the two projections (Figure 2B, bars) show great agreement with the probability density function of a standard normal distribution (Figure 2B, green graph), which indicates excellent generalizability of the model. Quantile-quantile (QQ) plots and residual vs. fitted plots of out-of-sample test data (Figure 2C and S1E) were similar to the plots of training data (Figure S1C-D), showing that the model is well-calibrated and generalizable. In terms of prediction accuracy, the model generalizes well with

out-of-sample R^2 values (0.32 and 0.24 respectively) approximately equal to the training data R^2 values.

Further, inspecting the positive and negative parts of the components identified (Figure 2D), we find that the sex-age associated component 1 consists of parts of the default mode network (most pronounced precuneus, angular gyrus, and posterior cingulate cortex), retrosplenial cortex, and parts of dorsolateral prefrontal cortex (warm colors) as well as parts of the basal ganglia (striatal structures like nucleus caudatus and putamen), thalamus, cerebellum, and, to some extend, parts of dorsal premotor cortex (cool colors). Component 2, which was mainly associated with sex differences, captures parts of the default mode network (medial prefrontal cortex, precuneus, posterior cingulate cortex), parahippocampal gyrus and retrosplenial cortex (warm colors), as well as temporo-parietal junction, part of fusiform gyrus, and dorsolateral prefrontal cortex (cool colors). This mapping was verified with NeuroSynth (Yarkoni et al. 2011). We emphasize that the projections found are sign invariant, meaning that we cannot interpret the positive and negative parts of the components (warm and cool colors respectively) as activation or deactivation. For two regions of opposite sign, as e.g. medial and dorsolateral prefrontal cortex in component 2, observing negative correlation between the regions will tend to increase the value of the projection of the FC matrix, while positive correlation will tend to lower the value.

Since the data come from 4 scanning sites, we asked if this could impact our results. In our model, the Z-scores showed similar distributions regardless of the acquisition site (Figure S3). Fitting a model with scanning site as a dummy coded categorical covariate in addition to sex, age, and sex-age interaction yielded the same γ_1, γ_2, and β coefficients as our model in Figure 2 (see Discussion, Methods, and Tables S3-S4 for a more thorough discussion).

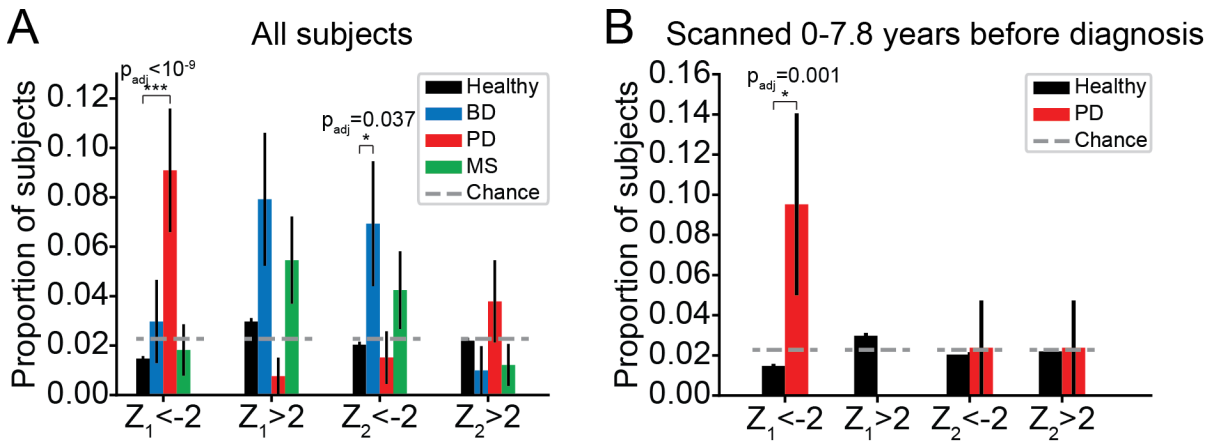

Figure 3: Disease prediction on selected brain diseases using FUNCOIN. A. Proportions of out-of-sample healthy and diagnosed subjects with extreme Z-scores in the two projections. PD and BD subjects show significantly more outliers after FDR correction: For PD: 9.1% have Z_1<-2; for BD 7.9 % had Z_1>2 and 6.9 % had Z_2<-2. **B.** Focusing on subjects with PD scanned before diagnosis: 9.3 % of the subjects with PD are outliers with Z_1<-2.

FUNCOIN can identify brain pathology even years before diagnosis

One of the motivations for normative modelling in psychiatry and neurology is the search for new biomarkers of brain diseases and disorders (Marquand et al. 2016). We therefore evaluated the proportion of outliers among individual out-of-sample subjects with and without brain diagnoses. Table 1 reports the number of subjects with different brain diagnoses in the UKB dataset. We focused our analysis on diseases with n>100. The diagnosis data is based on diagnosis occurrence in health records, and we therefore did not consider diseases that may have undergone remission at scan time (e.g. depression). Based on these criteria we focused on BD (N=101), PD (N=132), and MS (N=165). Some of the UKB subjects with brain diseases were scanned before the time of diagnosis (first occurrence in health record). For PD subjects, this was the case for 54 subjects, and the time of scan spanned from approx. 7.8 years before diagnosis to 23.2 years after the scan.

In Figure 3A, we included all subjects with BD, PD, or MS and computed the proportion of outliers (Z<-2 or Z>2) for the two projections. Given that the healthy population Z-scores follow a standard normal distribution, we expect p(Z>2)=p(Z<-2)≈2.3 % of healthy individuals to be outliers by chance for each direction (>2 or <-2) in each projection (Z_1 and Z_2)(Figure 3A-B, grey, dashed lines). For each diagnosis, we compared the proportion of outliers to the group of healthy subjects with a χ^2 test and FDR-corrected the *p*-values with Benjamini-Yekutieli correction (Benjamini and Yekutieli 2001).

For the first FUNCOIN projection, 9.1 % of the subjects with PD had Z_1-scores below -2, which was significantly more than in the group of healthy outliers ($p_{adj}<10^{-9}$). This indicates that the disease tends to decrease this component of FC, which is also indicated by the lower amount of PD subjects with $Z_1>2$. This component is composed of structures that were previously shown to be affected in PD at a focal level (nucleus caudatus)(Lotharius and Brundin 2002) and at a network level (thalamic and cerebellar networks)(Gratton et al. 2019; Tinaz et al. 2016). Subjects with BD showed a significantly higher amount of outliers in the second component with 6.9 % having $Z_2<-2$ ($p_{adj}=0.037$). The medial and dorsolateral prefrontal cortex are part of this component with opposite signs. Lower Z-scores in this component is compatible with a lack of anticorrelation between these two areas, which has been observed with rsfMRI in BD (Chai et al. 2011). In the first component, 7.9% had $Z_1>2$, which did not survive FDR-correction ($p_{adj}=0.112$). Interestingly, none of the subjects with BD had both $Z_1>2$ and $Z_2<-2$. This may indicate that the model is able to detect subgroups in the group of BD subjects (e.g. different subtypes or different stages of BD), and the effective prediction accuracy for BD is thus 14.8 %. In subjects with MS, 5.5 % of the subjects had Z_1-scores above 2, which was not significant.

Many brain diseases (e.g. PD) cannot be cured, but diagnosis at an early stage is important to slow down disease progression and diminish future symptom burden. We therefore tested the model's ability to identify subjects who were scanned before diagnosis with PD (n=54, scanned up to 7.8 years before diagnosis, on average 3.7 years). Remarkably, the proportion of outlier subjects with PD was approximately the same (9.3 %) as when we considered the whole group of subjects with PD (9.1 %). Despite this analysis being restricted to fewer subjects (n=54) this was still significant ($p_{adj}=0.001$). The outliers in this group were scanned up to 5.5 years before diagnosis (on average 3.2 years). These results show a potential of our method as a clinical tool for diagnosing early-stage brain disease or in screening of at-risk individuals, e.g. people with a family history of PD.

Table 1

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

Table 1: Number of subjects for selected psychiatric and neurological diseases in the UKB fMRI dataset. Diagnosis codes in parenthesis are from ICD10. Green rows indicate the diagnoses with n>100, which were included in our study.

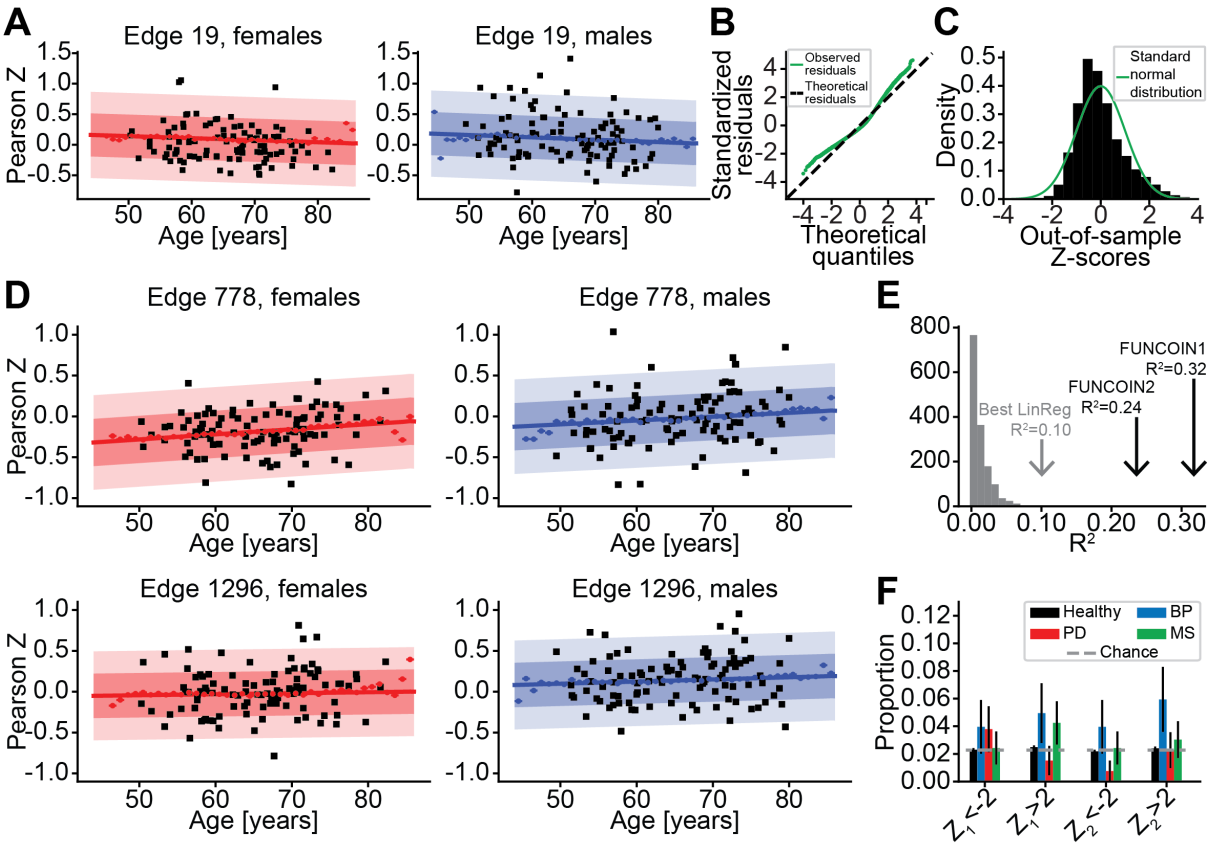

Figure 4: Comparing FUNCOIN to an element-wise normative model. A. Normative graphs for an example edge from an element-wise linear regression model highlighting some of the challenges of element-wise normative models (see text). **B.** QQ-plot from training data shows that the assumption of normally distributed residuals is violated (compare to our model in Figure S1C). **C.** The non-normality of the out-of-sample Z-scores is clear when comparing the histogram to a standard normal probability density function (compare to our model in Figure 2B). **D.** Normative graphs of the two edges with highest explained variance (R^2). For each sex (females red, males blue), the normative graphs of the two projections (top and bottom) are plotted as a function of age. 100 random out-of-sample subjects for each sex are plotted (black) to show the spread in the population. Comparing to Figure 2A, we see much less explained variance captured with sex and age as covariates. **E.** Histogram of edgewise R^2 values in the linear regression model. Arrows point to the best R^2 in the element-wise model as well as the R^2 values of the two components in our model (depicted in Figure 2A). **F.** In the element-wise linear regression model, the proportion of outliers in each diagnosis was not significantly different from the healthy group.

Full-matrix normative modelling outperforms element-wise methods

Previous normative models of FC act at the level of the individual elements of the FC matrix, using versions of linear or polynomial regression (Rutherford et al. 2023; Jalbrzikowski et al. 2019; Fraza et al. 2021; Jiang et al. 2024) or mixture models (Looden et al. 2022). Having established the feasibility of FUNCOIN for whole-brain normative modelling of FC, we next compared it to an element-wise linear regression model of Fisher z-transformed correlation values. The Fisher transformation is a standard way to make the correlation values approximately normal and is widely used in modelling of FC (including normative modelling) e.g. (Looden et al. 2022; Smith et al. 2023; Jalbrzikowski et al. 2019; Rutherford et al. 2023). This analysis

highlights some of the limitations of element-wise methods, which FUNCOIN can help bypass.

Common to all element-wise FC prediction methods is the fact that they test a massive amount of hypotheses (in our case 1485*2: above or below normal for each edge). After correcting for multiple comparisons, we would need very strong effects to draw any significant conclusion. For normative models in particular, this would effectively result in having to threshold Z-scores at a level way above |Z|>2 (in this case |Z|>4.3 with Bonferroni correction). This massively increases the number of false negatives, i.e. subjects that should be considered out of the normative distribution but are not. Our model, on the other hand, handles this problem by reducing dimensionality at the same time as performing the regression on the relevant covariates (sex and age). While we could in theory circumvent the multiple comparisons issue by integrating the Z values like e.g. (Rutherford et al. 2023), the following example edge (edge 19, where the residuals are most skew) shows three potential problems of this approach.

First, we see from the normative graphs of the example edge (Figure 4A) that the relation to sex and age is very weak. This is the case for most of the 1485 FC edges. Thus, when drawing conclusions from a set of 1485 regression models, one must either weigh the contributions of each model or dismiss most of them. On the other hand, FUNCOIN avoids this problem by simultaneously extracting and fitting to the aspects of FC that best relate to the covariates.

Second, element-wise linear regression, like FUNCOIN, requires the assumption of normally distributed residuals for the model to be valid and applicable. For the example edge, the violation of this assumption is visualized in the QQ plot of residuals from the training data, where we would expect the points to lie on a straight line with unit slope (Figure 4B, compare to Figure S1C). This shows that the model is not appropriate. Applying element-wise regression upstream of Z-score calculation requires dealing with the non-gaussian residuals, e.g. by likelihood-warping (Fraza et al. 2021). Without taking this into account, like e.g. (Jalbrzikowski et al. 2019), a large number of careful model sanity checks are required (Fraza et al. 2021). Our model deals with this issue by reducing dimensionality to a few relevant components thereby making it straightforward to verify the model assumptions (as in Figure 2 and Figure S1).

Third, related to the second issue, in normative modelling, we are interested in generalizability to out-of-sample healthy subjects as well as identifying subjects with brain pathology as outliers from the model. Bad or invalid model fits may result in a skewed distribution of Z-scores, as in the example edge (Figure 4C). The skewness of the Z-scores increases the false positive rate in one end of the distribution and the false negative rate in the other end. This arbitrariness makes this Z-score hard to compare to or combine with Z-scores from other edges and underlines the importance of validating all models used for Z-score calculations.

Instead of combining all the Z-scores, we now analyze the most relevant edges. We use the R^2 value as a goodness-of-fit measure as suggested by (Rutherford, Kia, et al. 2022). Figure 4D shows the normative graphs of the two edges with the highest out-of-sample R^2 values (R^2 = 0.10 and 0.09). In these edges, the model fit is much better than in the example edge (Figure S4). For the best edge, FC depends on age and both edges show a small difference between sexes. Examining out-of-sample R^2 values from all the edges in the element-wise model (Figure 4E), we see that the R^2 values of the selected edges (0.10 and 0.09) is well below the out-of-sample R^2 values obtained with FUNCOIN (0.32 and 0.24), and most other FC edges have much lower R^2 values. Thus, FC holds information about sex and age scattered out across FC edges, which can only be modelled integratively. By doing so, we utilize relevant information from the whole brain rather than picking out the specific connectivity between two brain regions. This has major implications for the model's ability to identify brain disease. We conducted the same analysis as in Figure 3 for the element-wise linear regression model, using the two edges selected based on R^2 values (Figure 4F). The model had less predictive accuracy than FUNCOIN, and all comparisons to healthy individuals showed non-significant differences after Benjamini-Yekutieli FDR-correction.

Discussion

In this paper, we propose a framework for normative modelling of rsfMRI FC with FUNCOIN and illustrate its usability in detecting brain disease. The model showed promising results in several ways. First, our model is flexible, consistent, scalable, and generalizable, which is essential for normative modelling. Second, it shows a quantifiable characterization of disease, with ~9 % of PD subjects and ~15% of the

BD subjects falling as outliers, which was significantly more than in the healthy population. Remarkably, for PD subjects the same accuracy was achieved when restricting the analysis to scans before the time of diagnosis. Third, it circumvents several challenges from element-wise models by simultaneously performing covariate-relevant dimensionality reduction and regression thereby extracting the relevant information for FC regression and normative modelling.

Because it only requires basic information and a resting-state brain scan, FUNCOIN is not disease-specific, but can be applied to any condition where we expect FC to deviate from normal, which is the case for a wide range of psychological and neurological diseases (Zhang et al. 2021). Given the ability of our model to identify early stage brain diseases, FUNCOIN could be part of screening/diagnosis, where an fMRI scan of the individual with suspected brain disease is evaluated against a pretrained normative model.

Normative modelling is tailored for subject-level inference and is well-suited to differentiate between subtypes of disease in heterogenous patient groups (Marquand et al. 2016). For example, earlier normative modelling studies in structural MRI have revealed large heterogeneity in patients with BD (Wolfers et al. 2018; Wolfers et al. 2021). In our study, the different Z-scores come from orthogonal components of the FC and therefore convey information about different parts of the FC. By differentiating between subjects being outliers in different components, our method has the potential to shed light on subtypes of brain diseases with different underlying pathologies despite being labelled with the same diagnosis. In our results, this may have been the case for the subjects with BD, since no subjects had both $Z_1>2$ and $Z_2<-2$. This may indicate two different subgroups of BD. In our study, disease heterogeneity can also pose a challenge, because we are applying a very general model on heterogenous disease categories with few subjects. This can be the reason for not seeing a significantly higher proportion of outliers among subjects with MS, which can be divided into 4 subtypes identifiable with structural MRI and machine learning (Eshaghi et al. 2021). Future studies applying our method to more patients with known disease subtype can explore this further.

Previous normative models in rsfMRI are all based on predicting FC. Most studies have utilized edgewise methods including versions of linear or polynomial regression (Fraza et al. 2021; Rutherford et al. 2023; Jalbrzikowski et al. 2019; Jiang et al. 2024) or Gaussian-Gamma mixture models (Looden et al. 2022). Two studies used Gaussian processes to model the ROI-wise FC strength (Sun et al. 2023; Liu et al. 2024), and an earlier study applied dimensionality reduction of the FC with group-ICA before predicting the components (Kessler, Angstadt, and Sripada 2016). Above, we identified several challenges to the edgewise approaches. Strategies for dealing with massive hypothesis testing include the use of extreme value distributions (Fraza et al. 2021; Kia and Marquand 2018; Marquand et al. 2016), correlation of Z-scores to cognitive trait scores (Fraza et al. 2021), normative probability maps (Marquand et al. 2016) or counting the above-threshold Z-scores (Rutherford et al. 2023). The inherent risk of very bad model fits in some edges, as identified in Figure 4, poses a problem for these strategies. Indeed, Savage et al., using voxel-wise Bayesian linear regression (BLR) in task-fMRI, noted that a subset of the Z-scores should be interpreted with caution due to bad model fits (Savage et al. 2024). Gaussian processes are extremely flexible and avoid the problem of bad model fits, but they do not scale well beyond a few thousand participants (Fraza et al. 2021). The use of likelihood warping in BLR is a way to (to some extend) deal with bad model fits in terms of non-Gaussian residuals while retaining good scalability, but this is not feasible with highly skewed data (Fraza et al. 2021). In contrast, FUNCOIN does not suffer from the above challenges. If shared, covariate-related components exist in the training data, the integrative approach of reducing dimensionality by focusing on the relevant components and doing regression at the same time ensures good model fits and few hypotheses to test. Furthermore, our method scales well and was trained on 32,000 subjects with 490 time points of 55 ICA components. On a standard office computer this took less than 25 minutes and computation time scaled linearly with number of subjects (Figure S5). Finally, the approach of Kessler et al, applying dimensionality reduction with group-ICA followed by prediction of ADHD diagnosis from the Z-scores, comes closest to our approach (Kessler, Angstadt, and Sripada 2016). The study provided insights on normal network-level maturation and deviations indicating reduced attention performance. However, for the diagnosis prediction, no components survived correction for multiple comparisons. We suggest

that the simultaneous integrative dimensionality reduction and regression in our model may offer an increased power.

For clinical application, it is crucial to be able to apply our normative model across different assessment sites. Our training data comes from 4 different scanning sites. In line with earlier normative modelling studies on structural MRI from UKB (Fraza et al. 2021; Kia et al. 2022), we found no scanning site effects on the Z-scores of the healthy subjects (Figure S3), which is probably due to good hardware and acquisition parameter alignment in UKB (de Boer et al. 2024). The flexibility of our model makes it easy to implement site effects in the model. We retrained the model with scanner site as an additional (categorical, dummy-coded) covariate (i.e. a model with sex, age, sex-age interaction, scanning site) and compared to the estimates of the basic model presented in Results (sex, age, sex-age interaction). This led to the same γ_1, γ_2, β_0, and $\boldsymbol{\beta}$. Identifying a third component in both models led to discrepancy because γ_3 in the model with site effects depended to some extend on scanning site (Table S4-S5). That is, FUNCOIN allows us to either account for site effects by including them in the model or focusing on components unaffected by the choice of site (as in this study). If accounted for in the model, the concluding regression step in FUNCOIN makes it straightforward to include subjects from different scanning sites after first estimating the site effect of the new site as has been suggested in a Bayesian linear regression setting (Rutherford, Fraza, et al. 2022). This will be left for future work.

As in other normative modelling studies of FC, interpretation of the Z-scores can be challenging. First, we do not know a priori whether a given brain pathology increases or decreases Z-scores. Second, given an outlier from the normative model it is important but difficult to determine whether the subject's deviation from normal is clinically relevant (e.g. brain disease) or not (e.g. artifact or chance)(Marquand et al. 2016). Thus, in a clinical setting the model must be combined with other clinical/paraclinical assessments. Third, some subjects with brain disease may be unidentifiable by a normative model, either because they used to be in one end of the normal spectrum and the brain pathology affects the FC in the other direction, or because some intervention has had a stabilizing effect. In that case, the subject is

not an outlier (yet), despite the impact of disease. A recent normative modelling study in structural MRI has proposed an analysis method for changes in Z-scores over time, which can be adapted to our model to take advantage of repeated scans (Bučková et al. 2024).

Further, we only consider linear effects on FC. Since FUNCOIN finds components related to the covariates provided, our good model fit does not provide evidence of the absence of non-linear effects identifiable with FUNCOIN. There may be diagnostic information in nonlinear effects, which are straightforward to include in further explorations of the model using nonlinear expansions.

Last, the FUNCOIN method is based on the assumption that the subjects' FC matrices share at least one eigencomponent that differs in strength between the subjects. It has been noted that this assumption is not necessarily fulfilled in the presence of brain diseases that change not only the connection strengths but also network architecture like e.g. Alzheimer's disease (Zhao, Wang, et al. 2021). However, we emphasize that this is not a problem in our adaptation of the method, since we are training our model on healthy subjects and identify subjects with disease as outliers.

Conclusion

In conclusion, we here propose FUNCOIN as a new framework for normative modelling of FC, which resolves challenges inherent to previous edgewise normative modelling approaches. We illustrate the ability of the model to detect brain diseases even years before clinical diagnosis, which shows the potential of this model in the search for brain disease biomarkers. Supported by a well-documented Python package, and given the flexibility, consistency, scalability, and generalizability of our model, we suggest that FUNCOIN can be used on a broad range of questions from explorative neuroscience to clinical practice.

Methods

Data

We use rsfMRI time series data from the UK-Biobank (Sudlow et al. 2015), which at the moment contains rsfMRI data from more than 60k subjects. 46,000 subjects without brain diagnoses (ICD-10 category F and G) were used for the normative

model in an approximately 70:30 train-test split (32,000 for training the model, 14,000 subjects for testing the model out-of-sample)(Rutherford, Kia, et al. 2022). For disease prediction, we used data from subjects with BD (ICD-10 F31, N=101), PD (ICD-10 G20, N=132), and MS (ICD-10 G35, N=165). The time of diagnosis used for calculating the time between diagnosis and scan (Figure 3, 4F) is the time of first occurrence in health records.

Regardless of which of the four scanning sites, rsfMRI scans were performed using a Siemens Skyra 3T with acquisition parameters harmonized across scanning sites (e.g. TR=735 ms, spatial resolution 2.4 mm). We used the UKB ICA parcellation with 55 ICA components (after removal of artifact components)(Beckmann and Smith 2004; Griffanti et al. 2014; Salimi-Khorshidi et al. 2014). Details on data acquisition and preprocessing can be found in (Alfaro-Almagro et al. 2018; Miller et al. 2016).

The data consist of time series data with 490 time points (6 minutes) per ICA component, except for 502 subjects, where there were 532 time points. In many subjects we noticed large fluctuations in the beginning of the measurement at some ICA components. We therefore right-truncated all time series to 490 points followed by removal of the first 8 time points.

Covariate-assisted principal regression

We here introduce FUNCOIN as a framework for normative modelling of FC. This is an adaptation of a newly proposed covariance regression method, covariate-assisted principal regression, which we briefly describe here. See supplementary materials of this paper and the paper originally proposing the method (Zhao, Caffo, et al. 2021) for more details. The method performs dimensionality reduction and regression in one go by finding components of a set of *p*-by-*p* FC matrices such that the components are most associated with covariates. The model is a log-linear, heteroscedastic model for covariance matrices for subject *i=1…N.* Assume that $\boldsymbol{y_{it}} \in \mathbb{R}^p$, the measured data for subject *i* at time point *t,* follows a *p*-dimensional multivariate normal distribution. It is assumed that there exists at least one projection, $\boldsymbol{\gamma} \in \mathbb{R}^p$, such that

$$\mathrm{u_i} = \log(Var(\boldsymbol{\gamma}^\top \boldsymbol{y_{it}})) = \log(\boldsymbol{\gamma}^\top \boldsymbol{\Sigma_i} \boldsymbol{\gamma}) = \beta_0 + \boldsymbol{x_i}^\top \boldsymbol{\beta}$$

where $\beta_o \in \mathbb{R}$ is the intercept, $\boldsymbol{\beta} \in \mathbb{R}^{q-1}$ are model coefficients, and $\boldsymbol{x_i} \in \mathbb{R}^{q-1}$ is the vector of *q* covariates of subject *i*. The identified $\boldsymbol{\gamma}$ is scaled to unit Euclidian norm,

and the logarithm ensures positive-definiteness of model predicted FC. Several $\boldsymbol{\gamma}$ (at most *p* components, but limited by the actual number of components shared among the subjects) can be identified sequentially, where each $\boldsymbol{\gamma_j}$ is identified after removing the first *j-1* components ($\Gamma^{j-1} = (\gamma_1, \dots, \gamma_{j-1})$ from all subjects' data.

In this study, the time series data was demeaned and scaled to variance 1 before identifying the components. The components identified are of the covariance matrix with *p* degrees of freedom, which for the standardized data (mean 0, SD 1) is equal to the Pearson correlation matrix.

Consistency of identified projections in or study

We assessed consistency of the $\boldsymbol{\gamma}$ projections obtained with FUNCOIN by repeatedly and randomly dividing subsets of healthy subjects of different sizes into 2 groups of equal size and fitting the model identifying up to 5 components (Figure S1A). For each choice of number of subjects, each of 100 repetitions gave rise to two sets of projections, $\boldsymbol{\gamma}_{1,1}, \dots, \boldsymbol{\gamma}_{1,5}$ and $\boldsymbol{\gamma}_{2,1}, \dots, \boldsymbol{\gamma}_{2,5}$. We assesed consistency of the projections estimated, by for each of the first 3 components in group 1 ($\boldsymbol{\gamma}_{1,1}, \boldsymbol{\gamma}_{1,2}, \boldsymbol{\gamma}_{1,3}$) computing the dot product with all five components from the second group ($\boldsymbol{\gamma}_{2,1}, \dots, \boldsymbol{\gamma}_{2,5}$) and finding the maximum absolute value. That is we evaluate

$$\max_{i=1,\dots,5} |\gamma_{1,j} \cdot \gamma_{2,i}|$$

for j=1,2,3.

The $\boldsymbol{\gamma}$s are sign invariant and norm 1, which makes the best possible absolute dot product value equal to 1. The projections identified in each group are orthogonal, so a good agreement between two $\boldsymbol{\gamma}_{1,i}$ and $\boldsymbol{\gamma}_{2,j}$ ensures, that the match is unique among the identified projections. Our estimations showed consistency of the two first projections identified for N larger than approximately 2000-4000 subjects. The third projection was not consistently identified for N≤24000, which was the highest possible number of healthy subjects in the half-split test.

Resolving bias

In the original paper, Zhao et al assessed bias, standard error, and coverage probabilities of confidence intervals of the estimated beta in a simulation study. However, they did not report on the intercept (β_0). Accurate numerical predictions of FC are required for normative modelling, and bias on the intercept would cause a

skew distribution of residuals (and thereby Z-scores). This is a problem because it makes the definition of outliers as |Z|>2 biased. We therefore reproduced the simulation study from (Zhao, Wang, et al. 2021) to assess bias on the β_0s. Table S1-S2 show that the estimation of γ and β is in agreement with values reported in (Zhao, Wang, et al. 2021). However, the estimations of β_0 (the intercept) in both projections, which were not reported by Zhao et al, were biased with β_0 of the second direction having the largest bias. The coverage probabilities of the β_0 confidence intervals were 0,27 and 0.69 respectively.

Applying covariate-assisted principal regression on our training data (n=32,000) with sex, age, and sex-age interaction as covariates, we also observed bias on the intercept. The average values in 1-year age groups of the training data were in general lower than predicted by the model with coefficients β_0 and β identified by the algorithm (Figure S1B). To remove any bias on the coefficients, we therefore added another step to the algorithm: After having identified projections (and coefficients), we perform ordinary linear regression on the transformed FC values to get unbiased beta estimates. That is, for each projection, *j*, we fit β_0 and β in the following model:

$$\mathrm{u}_{j,i} = \log\left(\boldsymbol{\gamma}_j^T \Sigma_i \boldsymbol{\gamma}_j\right) = \beta_0 + \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_{j,i}$$

where $\boldsymbol{\gamma}_j \in \mathbb{R}^p$ is projection *j*; $\boldsymbol{\Sigma_i}$ is the covariance matrix (size *p*-by-*p*) of subject *i*, $\boldsymbol{x_i}$ is the covariate vector (length *q*) of subject *i*; $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^{\boldsymbol{q}}$ are the intercept and coefficients of the model; and $\varepsilon_{j,i} \sim N(0, \sigma_j^2)$ is the residual in projection *j* which is normally distributed with mean 0 and projection-specific variance σ_j^2. This relies on the already established consistency and low bias of the γ estimates (see above, Table S2, and Figure S1A). It also relies on the residuals of the transformed training data in each component being having homogenous variance, i.e. not depending on the predicted value. Residual vs fitted plots (Figure S1D) of the training data show that these assumptions are reasonable for each of the components. This modified version suited for normative modelling is FUNCOIN.

Model fitting

To identify $\boldsymbol{\gamma}, \beta_0$ and $\boldsymbol{\beta}$ in FUNCOIN, the model is first fitted by maximizing the likelihood function (i.e. minimizing the negative log-likelihood, ignoring constants)

using a block coordinate decent algorithm (Zhao, Wang, et al. 2021). The Python package (FUNCOIN, https://github.com/kobbersmed/funcoin and installable with pip) allows for setting the number of initial conditions tried, the maximal number of iterations, and tolerance. The fitting procedure is repeated with the specified number of random or prespecified initial conditions. Each run terminates if the maximal number of iterations is reached or if no coordinate in the $\boldsymbol{\gamma}, \beta_0,$ and $\boldsymbol{\beta}$ being optimized changes more than the specified tolerance. From all runs with different initial conditions, the best fit in terms of minimizing the objective function is kept. If the unbiased version (FUNCOIN) is selected (which is default), the identification of a component concludes with determining β_0 and $\boldsymbol{\beta}$ as described above to resolve bias on the model coefficients. All model fitting in our study used the unbiased version, FUNCOIN, with default values of number of initial conditions (20), max iterations (1000) and tolerance (1e-4).

If more projections are associated with the covariates, additional projections can be found with the same algorithm after subtracting the already identified components and doing a rank-completion step. See supplementary information and the original paper for details (Zhao, Wang, et al. 2021).

To keep the model as simple and general as possible, we only included sex, age, and sex-age interaction as covariates. The data come from 4 different scanning sites. We investigated the possibility of scanning site effects by comparing to another model with sex, age, sex-age interaction, and scanning site as covariates. Scanning site was included as a categorical variable assumed to have an additive effect and was dummy coded with site 1 as reference.

Z-scores

Each subject's deviation from the normative model presented in Figures 2 and 4 is determined by computing the Z-scores of each component, for which we need an estimate of the standard deviation of the residuals. In our model, the residuals in the log-linear model of each component, i.e. the transformed FC values, were normally distributed and homogenous (see Figure S1C-D). The homogeneity allows us to robustly estimate the variance of the model because we base it on all training subjects and not in e.g. age groups as in previous normative models (Bethlehem et al. 2020). For subject *i*, the Z-score of component *j* is determined as

$$Z_{j,i} = \frac{u_{j,i} - \hat{u}_{j,i}}{\sigma_j}$$

where $\mathrm{u}_{j,i} = \log\left(\boldsymbol{\gamma}_j^T \Sigma_i \boldsymbol{\gamma}_j\right)$ is subject *i*'s value of component *j*, $\hat{u}_{ji}$ is the predicted value given subject *i*'s sex and age, and σ_j is the standard deviation of the residuals in component *j*. Given that the residuals are normally distributed around the model prediction with standard deviation σ_j, the Z-scores follow a standard normal distribution. We consider $|z_j| > 2$ an outlier, which corresponds to the ~2.3 % most extreme observations in each direction.

For the element-wise model (Figure 4), the same formula for the Z-scores was used but with element-wise prediction and standard deviation.

Statistical testing and FDR correction

In Figure 3 and 4, the proportion of outliers in the healthy out-of-sample subjects as well as in subjects with different brain diagnoses was determined. To test whether each group of diagnosed subjects was significantly different from the group of healthy out-of-sample subjects, we obtained *p*-values from a χ^2 test with 1 degree of freedom (i.e. from a 2-by-2 contingency table). All *p*-values in each figure were FDR-corrected with the Benjamini-Yekutieli procedure at level α=0.05. Statistics of regression coefficients (Table S3) was found using two-sided t-test as implemented in the FUNCOIN python package. Note that this relies on the established normally distributed and homogenous residuals.

Selecting the number of components in our study

As $\boldsymbol{\gamma}$ components are identified sequentially (like in PCA), it is necessary to settle on how many components to analyze. In our study we kept the two first $\boldsymbol{\gamma}$ components identified, for two reasons. First and most important, assessment of consistency of the identified $\boldsymbol{\gamma}$ components revealed that $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$ are consistently identified for N≥2000. For datasets up to N=24000 (which was the highest possible N in the half-split test with healthy subjects), $\boldsymbol{\gamma_3}$ was not consistently identified (Figure S1A). Second, when we fitted a model including (additive) site effects, the two first identified $\boldsymbol{\gamma}$ had site coefficients close to zero, while some scanning site effect was estimated in the third $\boldsymbol{\gamma}$.

Brain maps

The rsfMRI data used comes with the UKB ICA-parcellation with a 55 network parcellation (ICA100 after removal of artifactual components). These components are provided in MNI152 space, 2 mm resolution, as a 4D array of shape 91x109x91x55 in the NIFTI image format. Brain maps shown in Figure 2 show the combination of ICA components for the identified $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$. These are obtained by right-multiplying the ICA array with $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$ respectively, which are of shape 55x1. The brain maps were created using fsleyes version 1.12.6 from the FMRIB software library (FSL) (https://zenodo.org/records/11047709). Labelling of the brain areas and networks in the identified components was done by visual inspection assisted by the Neurosynth decoder (https://neurosynth.org/decode/)(Yarkoni et al. 2011).

Simulation study

In order to assess bias and variance on $\boldsymbol{\gamma}$, β_0 and $\boldsymbol{\beta}$, we reproduced the simulation study reported in main results in the covariate-assisted principal regression paper (Zhao, Wang, et al. 2021). In brief, we simulate a scenario with *p*=5 and *q*= 2 (including intercept). For the simulation, we define a 5-by-5 matrix, Γ, whose columns are the 5 true eigenvectors of the covariance matrices, such that $\Sigma_i = \Gamma\Lambda_{\mathrm{i}}\Gamma^{\mathrm{T}}$. Λ_{i} is a diagonal matrix whose elements are the strengths/eigenvalues of each eigenvector for subject *i*. Two of the eigenvalues depend on a binary, categorical covariate. The binary variable is simulated from a Bernoulli distribution with probability 0.5, and the eigenvalue is defined as $\lambda_{j,i} = \exp\,(x_i\beta_j)$. From this, we simulate multivariate normal time series (*p*=5) for 100 time points for 100 subjects. We repeat the simulation 200 times. In each iteration, we estimate $\boldsymbol{\gamma}$, β_0 and $\boldsymbol{\beta}$ with the (Zhao, Wang, et al. 2021) algorithm. Confidence intervals of β_0 and $\boldsymbol{\beta}$ are estimated from 500 bootstrap samples after having identified $\boldsymbol{\gamma}$. To replicate the simulation study in (Zhao, Wang, et al. 2021) we only kept components as long as the average DfD value was below 2 (see supplementary material paragraph B). To deal with differences in the order of the identified components, we reordered the vectors according to best match with the truth. See Table S1-S2 for the true and estimated parameters as well as statistics, and see (Zhao, Wang, et al. 2021) for details on the data simulation.

Model formulation of the edgewise linear model

We compared our model to edgewise linear regression of the FC edges after Fisher transformation. That is, for each subject *i* we vectorized the 1485 unique elements of the correlation matrix, Σ_i, and applied Fisher's z-transformation on each correlation value. Fisher's z-transformation is a standard way of making the correlation values closer to normally distributed. Z-scores in this model were calculated (edgewise) with the same formula as in FUNCOIN.

Software

With this paper, a python package, 'FUNCOIN', is released (https://github.com/kobbersmed/funcoin), installable directly from GitHub or with pip. The package is an implementation of the above method to identify covariate-dependent components of covariance matrices based on the multivariate time series data of each subject. Features include fitting to data, applying the model to held-out test data, estimating confidence intervals by bootstrapping or parametrically, calculating Z-scores, performing statistical tests, and more. See the accompanying documentation and tutorial. Covariate-assisted principal regression, on which FUNCOIN is based, is available in R (Zhao, Wang, et al. 2021).

Data availability

The data used is part of UK Biobank, an open-access data repository.

Code availability

All model training and testing, Z-score calculation, and calculation of statistics were performed using the FUNCOIN Python package released with this paper (https://github.com/kobbersmed/funcoin and installable with pip). Any code used for utilizing the FUNCOIN package as well as analyzing and visualizing data will be shared upon request.

Bibliography

Abbas, K., M. Liu, M. Wang, D. Duong-Tran, U. Tipnis, E. Amico, A. D. Kaplan, M. Dzemidzic, D. Kareken, B. M. Ances, J. Harezlak, and J. Goni. 2023. 'Tangent functional connectomes uncover more unique phenotypic traits', *iScience*, 26: 107624.

Alfaro-Almagro, F., M. Jenkinson, N. K. Bangerter, J. L. R. Andersson, L. Griffanti, G. Douaud, S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, E. Vallee, D. Vidaurre, M. Webster, P. McCarthy, C. Rorden, A. Daducci, D. C. Alexander, H. Zhang, I. Dragonu, P. M. Matthews, K. L. Miller, and S. M. Smith. 2018. 'Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank', *Neuroimage*, 166: 400-24.

Beckmann, C. F., and S. M. Smith. 2004. 'Probabilistic independent component analysis for functional magnetic resonance imaging', *IEEE Trans Med Imaging*, 23: 137-52.

Benjamini, Yoav, and Daniel Yekutieli. 2001. 'The control of the false discovery rate in multiple testing under dependency', *Annals of statistics*: 1165-88.

Bethlehem, R. A. I., J. Seidlitz, R. Romero-Garcia, S. Trakoshis, G. Dumas, and M. V. Lombardo. 2020. 'A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder', *Commun Biol*, 3: 486.

Bi, S., Y. Guan, and L. Tian. 2024. 'Prediction of individual brain age using movie and resting-state fMRI', *Cereb Cortex*, 34.

Biswal, B. B., M. Mennes, X. N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, A. M. Dogonowski, M. Ernst, D. Fair, M. Hampson, M. J. Hoptman, J. S. Hyde, V. J. Kiviniemi, R. Kotter, S. J. Li, C. P. Lin, M. J. Lowe, C. Mackay, D. J. Madden, K. H. Madsen, D. S. Margulies, H. S. Mayberg, K. McMahon, C. S. Monk, S. H. Mostofsky, B. J. Nagel, J. J. Pekar, S. J. Peltier, S. E. Petersen, V. Riedl, S. A. Rombouts, B. Rypma, B. L. Schlaggar, S. Schmidt, R. D. Seidler, G. J. Siegle, C. Sorg, G. J. Teng, J. Veijola, A. Villringer, M. Walter, L. Wang, X. C. Weng, S. Whitfield-Gabrieli, P. Williamson, C. Windischberger, Y. F. Zang, H. Y. Zhang, F. X. Castellanos, and M. P. Milham. 2010. 'Toward discovery science of human brain function', *Proc Natl Acad Sci U S A*, 107: 4734-9.

Bučková, Barbora Rehák, Charlotte Fraza, Rastislav Rehák, Marián Kolenič, Christian Beckmann, Filip Španiel, Andre Marquand, and Jaroslav Hlinka. 2024. "Using normative models pre-trained on cross-sectional data to evaluate longitudinal changes in neuroimaging data." In.: eLife Sciences Publications, Ltd.

Cai, Na, Karmel W. Choi, and Eiko I. Fried. 2020. 'Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies', *Human Molecular Genetics*, 29: R10-R18.

Calhoun, V. D., G. D. Pearlson, and J. Sui. 2021. 'Data-driven approaches to neuroimaging biomarkers for neurological and psychiatric disorders: emerging approaches and examples', *Curr Opin Neurol*, 34: 469-79.

Chai, X. J., S. Whitfield-Gabrieli, A. K. Shinn, J. D. Gabrieli, A. Nieto Castanon, J. M. McCarthy, B. M. Cohen, and D. Ongur. 2011. 'Abnormal medial prefrontal cortex resting-state connectivity in bipolar disorder and schizophrenia', *Neuropsychopharmacology*, 36: 2009-17.

Cuthbert, Bruce N., and Thomas R. Insel. 2013. 'Toward the future of psychiatric diagnosis: the seven pillars of RDoC', *BMC Medicine*, 11: 126.

de Boer, Augustijn A. A., Johanna M. M. Bayer, Seyed Mostafa Kia, Saige Rutherford, Mariam Zabihi, Charlotte Fraza, Pieter Barkema, Lars T. Westlye, Ole A. Andreassen, Max Hinne, Christian F. Beckmann, and Andre Marquand. 2024. 'Non-Gaussian normative modelling with hierarchical Bayesian regression', *Imaging Neuroscience*, 2: 1-36.

Deery, H. A., R. Di Paolo, C. Moran, G. F. Egan, and S. D. Jamadar. 2023. 'The older adult brain is less modular, more integrated, and less efficient at rest: A systematic review of large-scale resting-state functional brain networks in aging', *Psychophysiology*, 60: e14159.

Dosenbach, N. U., B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar, K. A. Barnes, J. W. Dubis, E. Feczko, R. S. Coalson, J. R. Pruett, Jr., D. M. Barch, S. E. Petersen, and B. L. Schlaggar. 2010. 'Prediction of individual brain maturity using fMRI', *Science*, 329: 1358-61.

Edde, M., G. Leroux, E. Altena, and S. Chanraud. 2021. 'Functional brain connectivity changes across the human life span: From fetal development to old age', *J Neurosci Res*, 99: 236-62.

Eshaghi, A., A. L. Young, P. A. Wijeratne, F. Prados, D. L. Arnold, S. Narayanan, C. R. G. Guttmann, F. Barkhof, D. C. Alexander, A. J. Thompson, D. Chard, and O. Ciccarelli. 2021. 'Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data', *Nat Commun*, 12: 2078.

Finn, E. S., X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. T. Constable. 2015. 'Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity', *Nat Neurosci*, 18: 1664-71.

Fornito, A., A. Zalesky, and M. Breakspear. 2015. 'The connectomics of brain disorders', *Nat Rev Neurosci*, 16: 159-72.

Fraza, Charlotte J., Richard Dinga, Christian F. Beckmann, and Andre F. Marquand. 2021. 'Warped Bayesian Linear Regression for Normative Modelling of Big Data'.

Gratton, C., J. M. Koller, W. Shannon, D. J. Greene, B. Maiti, A. Z. Snyder, S. E. Petersen, J. S. Perlmutter, and M. C. Campbell. 2019. 'Emergent Functional Network Effects in Parkinson Disease', *Cereb Cortex*, 29: 2509-23.

Griffanti, L., G. Salimi-Khorshidi, C. F. Beckmann, E. J. Auerbach, G. Douaud, C. E. Sexton, E. Zsoldos, K. P. Ebmeier, N. Filippini, C. E. Mackay, S. Moeller, J. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K. L. Miller, and S. M. Smith. 2014. 'ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging', *Neuroimage*, 95: 232-47.

Griffin, Ben, Christine Ahrends, Chetan Gohil, Fidel Alfaro-Almagro, Mark W. Woolrich, Stephen M. Smith, and Diego Vidaurre. 2024. 'Stacking models of brain dynamics to improve prediction of subject traits in fMRI', *Imaging Neuroscience*, 2: 1-22.

Hyman, Steven E. 2010. 'The Diagnosis of Mental Disorders: The Problem of Reification', *Annual Review of Clinical Psychology*, 6: 155-79.

Insel, Thomas, Bruce Cuthbert, Marjorie Garvey, Robert Heinssen, Daniel S. Pine, Kevin Quinn, Charles Sanislow, and Philip Wang. 2010. 'Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders', *American Journal of Psychiatry*, 167: 748-51.

Jalbrzikowski, M., V. P. Murty, B. Tervo-Clemmens, W. Foran, and B. Luna. 2019. 'Age-Associated Deviations of Amygdala Functional Connectivity in Youths With Psychosis Spectrum Disorders: Relevance to Psychotic Symptoms', *Am J Psychiatry*, 176: 196-207.

Jiang, A., X. Ma, S. Li, L. Wang, B. Yang, S. Wang, M. Li, and G. Dong. 2024. 'Age-atypical brain functional networks in autism spectrum disorder: a normative modeling approach', *Psychol Med*, 54: 2042-53.

Kapur, S., A. G. Phillips, and T. R. Insel. 2012. 'Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it?', *Mol Psychiatry*, 17: 1174-9.

Kessler, D., M. Angstadt, and C. Sripada. 2016. 'Growth Charting of Brain Connectivity Networks and the Identification of Attention Impairment in Youth', *JAMA Psychiatry*, 73: 481-9.

Kia, S. M., H. Huijsdens, S. Rutherford, A. de Boer, R. Dinga, T. Wolfers, P. Berthet, M. Mennes, O. A. Andreassen, L. T. Westlye, C. F. Beckmann, and A. F. Marquand. 2022. 'Closing the life-cycle of normative modeling using federated hierarchical Bayesian regression', *PLoS One*, 17: e0278776.

Kia, Seyed Mostafa, and Andre Marquand. 2018. "Normative Modeling of Neuroimaging Data Using Scalable Multi-task Gaussian Processes." In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, edited by Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López and Gabor Fichtinger, 127-35. Cham: Springer International Publishing.

Liu, L., D. Jia, Z. He, B. Wen, X. Zhang, and S. Han. 2024. 'Individualized functional connectome abnormalities obtained using two normative model unveil neurophysiological subtypes of obsessive compulsive disorder', *Prog Neuropsychopharmacol Biol Psychiatry*, 135: 111122.

Looden, T., D. L. Floris, A. Llera, R. J. Chauvin, T. Charman, T. Banaschewski, D. Murphy, A. F. Marquand, J. K. Buitelaar, C. F. Beckmann, and Aims-Trials group. 2022. 'Patterns of connectome variability in autism across five functional activation tasks: findings from the LEAP project', *Mol Autism*, 13: 53.

Lopez-Larson, M. P., J. S. Anderson, M. A. Ferguson, and D. Yurgelun-Todd. 2011. 'Local brain connectivity and associations with gender and age', *Dev Cogn Neurosci*, 1: 187-97.

Lotharius, Julie, and Patrik Brundin. 2002. 'Pathogenesis of parkinson's disease: dopamine, vesicles and α-synuclein', *Nature Reviews Neuroscience*, 3: 932-42.

Lund, M. J., D. Alnaes, A. G. de Lange, O. A. Andreassen, L. T. Westlye, and T. Kaufmann. 2022. 'Brain age prediction using fMRI network coupling in youths and associations with psychiatric symptoms', *Neuroimage Clin*, 33: 102921.

Margulies, D. S., S. S. Ghosh, A. Goulas, M. Falkiewicz, J. M. Huntenburg, G. Langs, G. Bezgin, S. B. Eickhoff, F. X. Castellanos, M. Petrides, E. Jefferies, and J. Smallwood. 2016. 'Situating the default-mode network along a principal gradient of macroscale cortical organization', *Proc Natl Acad Sci U S A*, 113: 12574-79.

Marquand, A. F., S. M. Kia, M. Zabihi, T. Wolfers, J. K. Buitelaar, and C. F. Beckmann. 2019. 'Conceptualizing mental disorders as deviations from normative functioning', *Mol Psychiatry*, 24: 1415-24.

Marquand, A. F., I. Rezek, J. Buitelaar, and C. F. Beckmann. 2016. 'Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies', *Biol Psychiatry*, 80: 552-61.

Miller, K. L., F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. Andersson, L. Griffanti, G. Douaud, T. W. Okell, P. Weale, I. Dragonu, S. Garratt, S. Hudson, R. Collins, M. Jenkinson, P.

M. Matthews, and S. M. Smith. 2016. 'Multimodal population brain imaging in the UK Biobank prospective epidemiological study', *Nat Neurosci*, 19: 1523-36.
Moriarity, D. P., K. J. Joyner, G. M. Slavich, and L. B. Alloy. 2022. 'Unconsidered issues of measurement noninvariance in biological psychiatry: A focus on biological phenotypes of psychopathology', *Mol Psychiatry*, 27: 1281-85.
Nour, M. M., Y. Liu, and R. J. Dolan. 2022. 'Functional neuroimaging in psychiatry and the case for failing better', *Neuron*, 110: 2524-44.
Power, J. D., A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen. 2011. 'Functional network organization of the human brain', *Neuron*, 72: 665-78.
Rasero, J., H. Aerts, M. Ontivero Ortega, J. M. Cortes, S. Stramaglia, and D. Marinazzo. 2018. 'Predicting functional networks from region connectivity profiles in task-based versus resting-state fMRI data', *PLoS One*, 13: e0207385.
Rasero, J., A. I. Sentis, F. C. Yeh, and T. Verstynen. 2021. 'Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability', *PLoS Comput Biol*, 17: e1008347.
Rutherford, S., P. Barkema, I. F. Tso, C. Sripada, C. F. Beckmann, H. G. Ruhe, and A. F. Marquand. 2023. 'Evidence for embracing normative modeling', *Elife*, 12.
Rutherford, S., C. Fraza, R. Dinga, S. M. Kia, T. Wolfers, M. Zabihi, P. Berthet, A. Worker, S. Verdi, D. Andrews, L. K. Han, J. M. Bayer, P. Dazzan, P. McGuire, R. T. Mocking, A. Schene, C. Sripada, I. F. Tso, E. R. Duval, S. E. Chang, B. W. Penninx, M. M. Heitzeg, S. A. Burt, L. W. Hyde, D. Amaral, C. Wu Nordahl, O. A. Andreasssen, L. T. Westlye, R. Zahn, H. G. Ruhe, C. Beckmann, and A. F. Marquand. 2022. 'Charting brain growth and aging at high spatial precision', *Elife*, 11.
Rutherford, S., S. M. Kia, T. Wolfers, C. Fraza, M. Zabihi, R. Dinga, P. Berthet, A. Worker, S. Verdi, H. G. Ruhe, C. F. Beckmann, and A. F. Marquand. 2022. 'The normative modeling framework for computational psychiatry', *Nat Protoc*, 17: 1711-34.
Salimi-Khorshidi, G., G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith. 2014. 'Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers', *Neuroimage*, 90: 449-68.
Savage, H. S., P. C. R. Mulders, P. F. P. van Eijndhoven, J. van Oort, I. Tendolkar, J. N. Vrijsen, C. F. Beckmann, and A. F. Marquand. 2024. 'Dissecting task-based fMRI activity using normative modelling: an application to the Emotional Face Matching Task', *Commun Biol*, 7: 888.
Scheinost, D., E. S. Finn, F. Tokoglu, X. Shen, X. Papademetris, M. Hampson, and R. T. Constable. 2015. 'Sex differences in normal age trajectories of functional brain networks', *Hum Brain Mapp*, 36: 1524-35.
Schulz, M. A., D. Bzdok, S. Haufe, J. D. Haynes, and K. Ritter. 2024. 'Performance reserves in brain-imaging-based phenotype prediction', *Cell Rep*, 43: 113597.
Smith, B. B., Y. Zhao, M. A. Lindquist, and B. Caffo. 2023. 'Regression models for partially localized fMRI connectivity analyses', *Front Neuroimaging*, 2: 1178359.
Smith, S. M., D. Vidaurre, F. Alfaro-Almagro, T. E. Nichols, and K. L. Miller. 2019. 'Estimation of brain age delta from brain imaging', *Neuroimage*, 200: 528-39.
Smith, S. M., D. Vidaurre, C. F. Beckmann, M. F. Glasser, M. Jenkinson, K. L. Miller, T. E. Nichols, E. C. Robinson, G. Salimi-Khorshidi, M. W. Woolrich, D. M. Barch, K.

Ugurbil, and D. C. Van Essen. 2013. 'Functional connectomics from resting-state fMRI', *Trends Cogn Sci*, 17: 666-82.

Smith, Stephen M., Peter T. Fox, Karla L. Miller, David C. Glahn, P. Mickle Fox, Clare E. Mackay, Nicola Filippini, Kate E. Watkins, Roberto Toro, Angela R. Laird, and Christian F. Beckmann. 2009. 'Correspondence of the brain's functional architecture during activation and rest', *Proceedings of the National Academy of Sciences*, 106: 13040-45.

Sudlow, Cathie, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. 2015. 'UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age', *PLOS Medicine*, 12.

Sun, X., J. Sun, X. Lu, Q. Dong, L. Zhang, W. Wang, J. Liu, Q. Ma, X. Wang, D. Wei, Y. Chen, B. Liu, C. C. Huang, Y. Zheng, Y. Wu, T. Chen, Y. Cheng, X. Xu, Q. Gong, T. Si, S. Qiu, C. P. Lin, J. Cheng, Y. Tang, F. Wang, J. Qiu, P. Xie, L. Li, Y. He, Dida-Mdd Working Group, and M. Xia. 2023. 'Mapping Neurophysiological Subtypes of Major Depressive Disorder Using Normative Models of the Functional Connectome', *Biol Psychiatry*, 94: 936-47.

Tinaz, S., P. Lauro, M. Hallett, and S. G. Horovitz. 2016. 'Deficits in task-set maintenance and execution networks in Parkinson's disease', *Brain Struct Funct*, 221: 1413-25.

Wolfers, T., N. T. Doan, T. Kaufmann, D. Alnaes, T. Moberget, I. Agartz, J. K. Buitelaar, T. Ueland, I. Melle, B. Franke, O. A. Andreassen, C. F. Beckmann, L. T. Westlye, and A. F. Marquand. 2018. 'Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models', *JAMA Psychiatry*, 75: 1146-55.

Wolfers, T., J. Rokicki, D. Alnaes, P. Berthet, I. Agartz, S. M. Kia, T. Kaufmann, M. Zabihi, T. Moberget, I. Melle, C. F. Beckmann, O. A. Andreassen, A. F. Marquand, and L. T. Westlye. 2021. 'Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder', *Hum Brain Mapp*, 42: 2546-55.

Yarkoni, Tal, Russell A. Poldrack, Thomas E. Nichols, David C. Van Essen, and Tor D. Wager. 2011. 'Large-scale automated synthesis of human functional neuroimaging data', *Nature Methods*, 8: 665-70.

Zhang, J., A. Kucyi, J. Raya, A. N. Nielsen, J. S. Nomi, J. S. Damoiseaux, D. J. Greene, S. G. Horovitz, L. Q. Uddin, and S. Whitfield-Gabrieli. 2021. 'What have we really learned from functional connectivity in clinical populations?', *Neuroimage*, 242: 118466.

Zhao, Y., B. Caffo, X. Luo, and Initiative Alzheimer's Disease Neuroimaging. 2021. 'Principal regression for high dimensional covariance matrices', *Electron J Stat*, 15: 4192-235.

Zhao, Y., B. Wang, S. H. Mostofsky, B. S. Caffo, and X. Luo. 2021. 'Covariate Assisted Principal regression for covariance matrix outcomes', *Biostatistics*, 22: 629-45.

Funding and acknowledgements:

We thank Yi Zhao, Indiana University School of Medicine, who first proposed the method that this paper builds on, for readily answering any inquiries and for valuable correspondence on the details of the method.
JK is supported by a DFF1 project from the Independent Research Fund Denmark (2034-00054B). DV is supported by a Novo Nordisk Foundation Emerging Investigator Fellowship (NNF19OC-0054895) and an ERC Starting Grant (ERC-StG-2019-850404).
This work was based on data from UK Biobank under application 8107. This research was funded in part by the Wellcome Trust (215573/Z/19/Z). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

J.K. and D.V conceived the study. J.K., A.M, C.G, and D.V.: Contributed to methodological considerations and interpretation of results. J.K. and C.G.: Developed the software package. J.K.: Formal analysis, visualizations, and writing (original draft). J.K. and D.V: Wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Materials and correspondence

Correspondence and materials requests should be addressed to Janus R. L. Kobbersmed (janus@cfin.au.dk or januslind@gmail.com).