Normative modelling of brain function abnormalities in complex pathology needs the whole brain

Janus R. L. Kobbersmed[1], Chetan Gohil[2], Andre Marquand[3,4], Diego Vidaurre[1,2]

[1]Department of Clinical Medicine, Center of Functionally Integrative Neuroscience, Aarhus University, 8000 Aarhus, Denmark
[2]Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, United Kingdom
[3]Department of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen 6525EN, the Netherlands
[4]Donders Institute for Cognition, Brain and Behavior, Radboud University, Nijmegen, The Netherlands

Correspondence to: Janus R. L. Kobbersmed
Center of Functionally Integrative Neuroscience, Aarhus University, Universitetsbyen 3, 8000 Aarhus C.
E-mail: janus@cfin.au.dk

Abstract

Many brain diseases and disorders lack objective measures of brain function as indicators of pathology. The search for brain function biomarkers is complicated by the fact that these conditions are often heterogenous and described as a spectrum from normal to abnormal rather than a sick-healthy dichotomy. Normative modelling addresses these challenges by characterizing the normal variation of brain function given sex and age and identifying abnormalities as deviations from this norm. Focusing on functional connectivity (FC) as a way to capture the network properties of the brain's activity, we here argue that the pathological effects of neurological or psychiatric disease lie at the systemic level, and that whole-brain normative models are more suitable to capture individual variations associated to these complex conditions than existing localized approaches that analyze one connection at a time. To be able to capture the whole-brain effects of disease, we thus propose Functional Connectivity Integrative Normative Modelling (FUNCOIN) as a novel whole-brain approach to normative modelling of FC. Using FUNCOIN and UK Biobank resting-state fMRI data from 46,000 healthy subjects across training and testing sets, we found that subjects with bipolar disorder and Parkinson's disease were significantly, and substantially, more likely than healthy subjects to exhibit abnormal FC patterns, which was not seen in localized models. Subjects with bipolar disorder divided into two distinct subgroups characterized by different brain function deviations. In Parkinson's disease subjects, abnormal FC patterns were significant even on scans up to 8 years before diagnosis.

Running title: Whole-brain normative modelling in fMRI

Keywords: Normative modelling; disease biomarkers; resting-state fMRI; functional connectivity; network models

Introduction

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[1]. 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[2]. However, recent studies suggest that brain disease would be better characterized by the consequences that such focal abnormalities have on wider brain function[3,4]. 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.

Meanwhile, as a way of identifying individuals that deviate from the healthy population, normative modelling has become increasingly popular particularly in structural, but also functional, brain imaging[5,6]. 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[7]. 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 model prediction. 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 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[8-15].

Functional connectivity (FC) from resting-state fMRI (rsfMRI) has been instrumental to our modern understanding of brain function[16-20]. In patients, rsfMRI studies have revealed complex, network-level, disease-related changes in FC over a range of neurological and psychiatric diagnoses[4], e.g. changes in thalamic, sensorimotor and cerebellar networks in subjects with Parkinson's disease[21,22] and a lack of anticorrelation between medial and dorsolateral prefrontal cortex in bipolar disorder[4,23]. Therefore, based on the 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 resting-state FC. Unfortunately, current normative models of rsfMRI FC are localized in the sense that they measure abnormalities at the level of individual connections (edges), and do not capture the whole-brain signatures of disease.

To fill this gap, we introduce Functional Connectivity Integrative Normative Modelling (FUNCOIN) as a method to evaluate how pathology pushes individuals from the normative range of brain function at a system level. FUNCOIN is an adaptation of a recently developed method from statistics literature, covariate-assisted principal regression[24], which performs simultaneous dimensionality reduction and regression of covariance matrices. In essence, FUNCOIN, which we released 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 semidefinite) 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. Consequently, we also avoid mass hypothesis testing, a major limitation of current normative modelling approaches on FC.

Using data from UK Biobank (UKB) and FUNCOIN, we create an integrative normative model of FC from 32000 individuals and test it on 14000 healthy subjects. We then use this model to characterize one psychiatric and one neurological brain disease: bipolar disorder (BD) and Parkinson's disease (PD). Strikingly, PD subjects were identifiable in our model years before being diagnosed. Subjects with BD, on the other hand, we found to stratify into two subgroups characterized by their different brain function deviations. None of these effects could be found by localized normative modelling approaches.

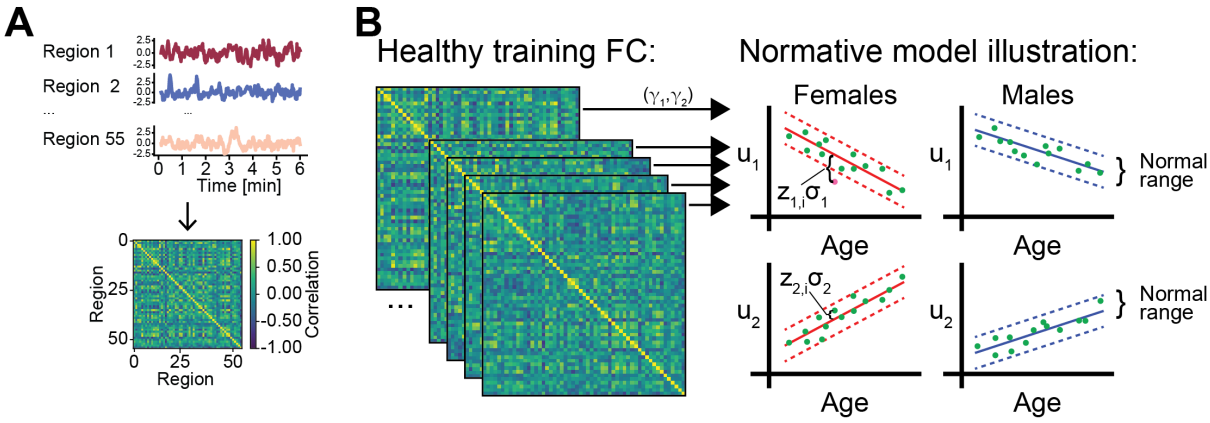

Figure 1: Schematic for whole-brain normative modelling. A. We aim at estimating a normative model of rsfMRI FC, calculated as the Pearson correlation of the time series from each pair of ICA components/brain regions. **B.** FUNCOIN identifies two projections (γ_1,γ_2) in the training group such that their magnitudes are modelled by a (2-dimensional) log-linear model (see Methods). Individual (out-of-sample) subject deviation is determined as *Z*-scores (green points and annotation). The pink point is an outlier example.

Results

Whole-brain normative modelling of FC

To perform normative modelling of rsfMRI FC, we need: 1) A large rsfMRI dataset of subjects representing the healthy background population as well as subjects that are expected to deviate from the healthy cohort in specific dimensions due to brain disease; and 2) a method to predict FC as a function of sex and age.

We use rsfMRI data from UKB (*N*>60,000 people)[25]. In this setting, we consider a subject healthy if they have no psychiatric or neurological diagnoses (group F or G diagnoses in ICD-10), which is the case for approximately 50000 of the subjects. We divide the subjects into three groups: 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 with psychiatric or neurological diagnoses. We focused on chronic brain diseases that have been addressed in the rsfMRI FC literature[4] and that have more than 100 subjects in the dataset: BD (*N*=101) and PD (*N*=132). The data consist of rsfMRI time series in the UKB-parcellation with 55 components from independent component analysis (ICA) (Fig. 1A, top, and Methods)[26]. Our measure of brain function is FC, determined in this case as the Pearson correlation matrix from all pairs of the parcellation components (Fig. 1A, bottom)[26].

In order to do network-level normative modelling of FC, we propose FUNCOIN (as outlined in Fig. 1B), which is a modified, unbiased version (see Methods, Supplementary Fig. 1) of a recently developed covariance regression method, covariate-assisted principal regression[24]. 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 (Fig. 1B left), we identify two projections ($\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$, Fig. 1B arrows) of the FC matrices and corresponding regression coefficients that relate these projections to age and sex (through a log-linear model; see Methods), which we can use to compute the normal range in the training data (illustrated in Fig. 1B right). With the out-of-sample healthy test subjects, we assess explanatory power and generalizability of the model, while the individual *Z*-scores of the healthy and diagnosed subjects are used to probe the model's ability to identify subjects with brain disease as outliers. 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 (Fig. 1B, pink point, see Methods).

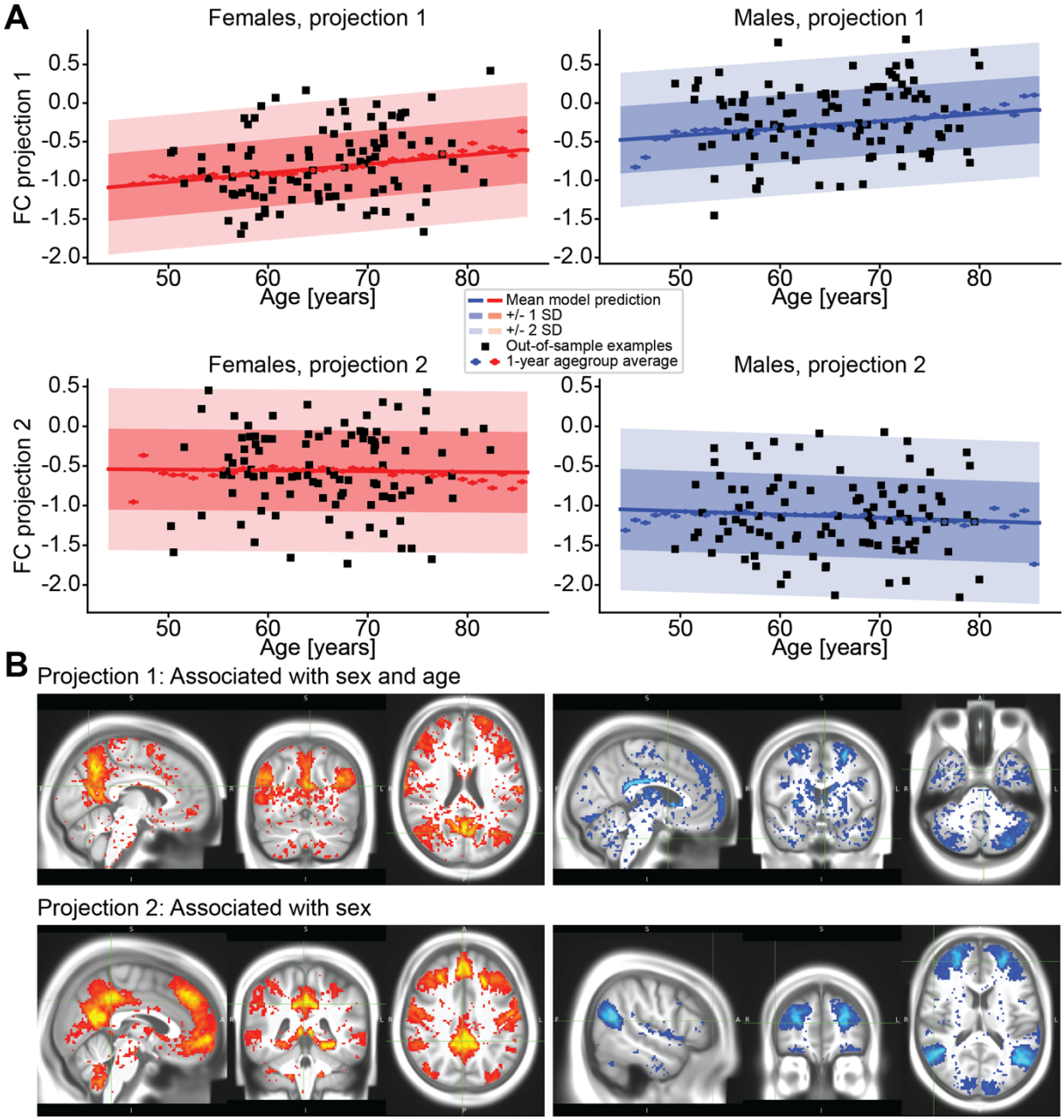

Figure 2: Whole-brain normative modelling with FUNCOIN on UK Biobank. A. Normative graphs from the identified projections found by fitting to the training data: For each sex (females red, males blue), the normative graphs of the two projections (top and bottom) are plotted as a function of age. 100 random out-of-sample subjects for each sex are plotted (black) to show the spread in the background population. Component 1 is associated with both sex and age, while component 2 is mainly associated with sex. See also Supplementary Figs. 1-2, 5 and Supplementary Tables 1-2. **B.** Brain maps showing the positive (warm colors) and negative (cool colors) parts of each component. Component 1 comprises default mode network areas, dorsolateral prefrontal cortex, thalamus, retrosplenial cortex, cerebellum, and striatal structures. Component 2 includes default mode network, temporoparietal junction, parahippocampal gyrus, and medial and dorsolateral prefrontal cortex. See also supplementary text.

Sex- and age-related whole-brain FC trajectories identified

The projected FC matrices from the 32,000 training subjects give rise to a linear model (Fig. 2A, see below for model validation and calibration), where the first projection is associated with both sex and age (Fig. 2A, top, R^2=0.32), and the second projection is associated mostly with sex (Fig. 2A, bottom, R^2=0.24). The shaded areas show ±2 SD. See Supplementary Fig. 2 for $\boldsymbol{\gamma}$ loadings and Supplementary Table 1 for β_0 and $\boldsymbol{\beta}$ coefficients and statistics. Having identified two components whose strength change with sex and age (component 1) and sex (component 2), we inspected the positive and negative parts of the components identified (Fig. 2B). 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 extent, 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). 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. However, a change in functional connectivity between brain regions pronounced in the component will have specific effects on the component strength. For example, for two regions of opposite sign, 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.

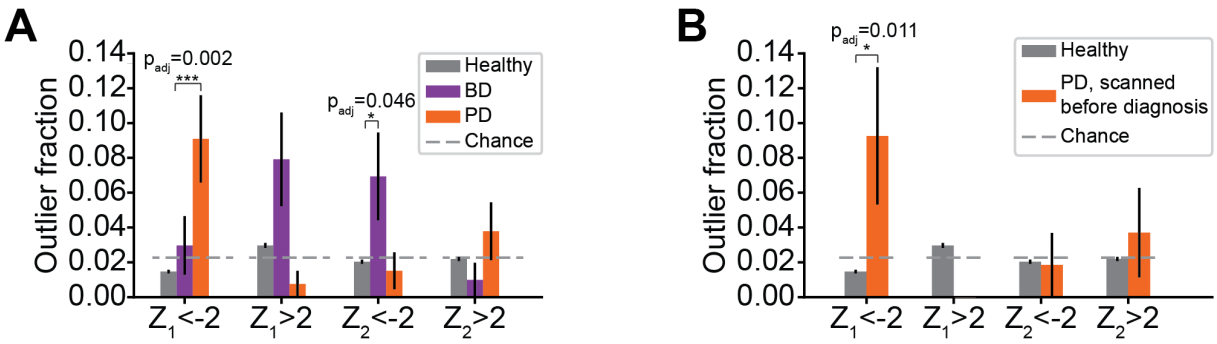

Figure 3: Whole-brain normative modelling with FUNCOIN on UK Biobank. A. Proportions of out-of-sample healthy and diagnosed subjects with extreme *Z*-scores in the two FUNCOIN projections in either direction ($Z<-2$ and $Z>2$). PD and BD subjects showed significantly more outliers after FDR correction: For PD, 9.1% had $Z_1<-2$; for BD, 7.9% had $Z_1>2$ and 6.9% had $Z_2<-2$. **B.** Focusing on subjects with PD scanned up to 7.8 years before diagnosis, 9.3% of the subjects with PD were outliers with $Z_1<-2$.

Whole-brain normative modelling can identify individual deviances in complex diseases

To test our hypothesis that normative modeling is best done on network level, we first tested how well our whole-brain model identified subjects with PD or BD. We did so by quantifying the proportion of outlier subjects in either direction of the two FUNCOIN projections ($Z<-2$ or $Z>2$) among out-of-sample subjects with and without brain diagnoses. Given that the healthy population *Z*-scores follow a standard normal distribution (see model validation below), we expect $p(Z>2)=p(Z<-2)\approx 2.3\%$ of healthy individuals to be outliers by chance for each direction (>2 or <-2) in each projection (Z_1 and Z_2)(Fig. 3A-B, dashed lines). For each diagnosis, we compared the proportion of outliers to the group of healthy subjects with permutation testing followed by FDR-correction with Benjamini-Yekutieli correction (see Methods)[27]. 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 healthy group (Fig. 3A, $p_{adj}=0.002$). This indicates that the disease tends to decrease this component of FC, which is also indicated by the lower proportion 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)[2] and at a network level (thalamic and cerebellar networks)[21,22]. Subjects with BD showed a significantly larger proportion of outliers in the second component with 6.9% having $Z_2<-2$ (Fig. 3A, $p_{adj}=0.046$). 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 in a smaller rsfMRI study (N_{BD}=14)[23]. In the first component, 7.9% of the subjects with BD had Z_1>2, which did not survive FDR-correction (p_{adj}=0.09). Interestingly, none of the subjects with BD had both Z_1>2 and Z_2<-2 indicating that the model detects 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%.

Parkinson's patients show whole-brain FC deviations from normal years earlier than diagnosis

Parkinson's disease is a progressing, neurodegenerative disease characterized by a long prodromal phase. The first noticeable symptoms occur on average 10 years before clinical diagnosis, and earlier diagnosis will probably be important when future disease-modifying treatments become available[28]. We therefore took advantage of the fact that some of the PD subjects in UKB were scanned before the time of diagnosis by testing the model's ability to identify pathology in these subjects (N=54, scanned up to 7.8 years before diagnosis). Remarkably, the proportion of outlier subjects with PD was significantly higher than among healthy subjects (p_{adj}=0.01) and approximately the same (9.3%, Fig. 3B) as when we considered the whole group of subjects with PD (9.1%, Fig. 3A). On average, the outliers in this group were scanned 3.2 years before clinical diagnosis. Altogether, this shows that the integrative normative modelling approach can identify brain function abnormalities in Parkinson's disease even years before clinical diagnosis, evidencing its potential for diagnosing early-stage brain disease or in screening of at-risk individuals e.g. due to genetically susceptibility and/or family history.

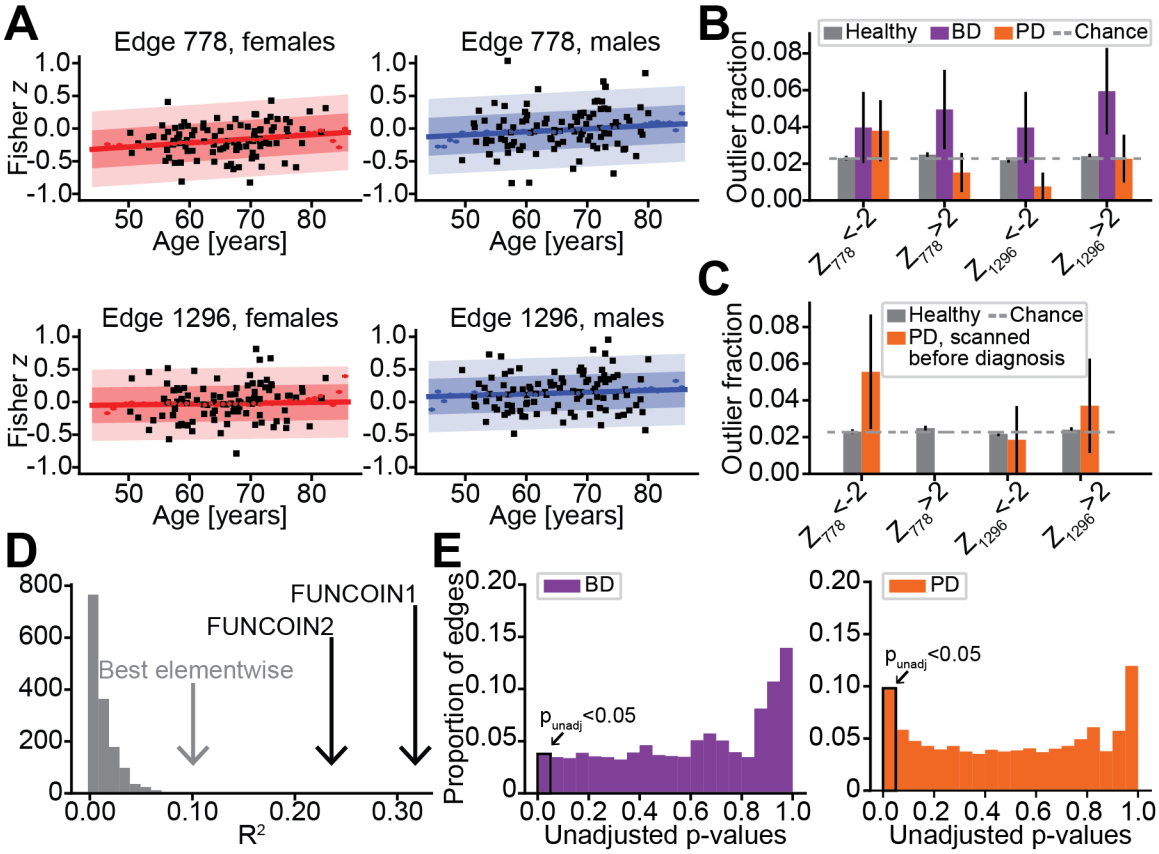

Figure 4: Localized models cannot reliably predict brain diseases, A. For each sex (females red, males blue), the normative graphs of the two edges in the edgewise model with highest R^2 values (0.10 and 0.09 respectively) (top and bottom) are plotted as a function of age. 100 random out-of-sample subjects for each sex are plotted (black) to show the spread in the background population. **B.** Same analysis as in 2C, but with the two selected edges from 3A. The diagnosed groups do not have significantly more outliers in these edges compared to the healthy subjects. **C.** Same analysis as in 2D, i.e. analysis restricted to subjects scanned before clinical diagnosis, but with the two selected edges from 3A. None of the differences are significant. **D.** Histogram of R^2 values in the edgewise model show that most of the individual connection models do not capture a substantial association with sex and age. **E.** Histograms of unadjusted p-values from the all the individual, edgewise models and in each direction. The p-values come from comparisons like in 2A-B,D-E, where BD (purple) and PD (orange) subjects were compared to the healthy group with Fisher's exact test.

Localized normative models cannot reliably find deviances in FC

To test our hypothesis that normative modelling of complex diseases from functional data should be tackled at an integrative level, we contrasted FUNCOIN to an edgewise approach, where multiple models predict each connection in the FC matrix independently. As it is common practice in the field, we predicted the Fisher z-transformed correlations values[6,29-31] by for each edge fitting a linear regression model with sex, age, and sex-age interaction as covariates.

A fundamental issue of all edgewise approaches is the massive number of *Z*-scores per subject (in this case 2970; $Z<-2$ and $Z>2$ for each connection). After correcting for multiple comparisons, very strong effects are needed to be able to draw meaningful conclusions. We therefore first tested the performance of the edgewise model by picking out the two best models based on goodness-of-fit as measured by the R^2 values (Fig. 4A). Assessment of model assumptions (normality and homogeneity of residuals) and out-of-sample *Z*-score distributions showed that the model fits were reasonable (Supplementary Fig. 3). However, in these two edges, the proportions of outliers among the test subjects with diagnoses were not significantly different from the outlier proportion among healthy test subjects (Fig. 4B). Restricting the analysis to subjects scanned before diagnosis (like in Fig. 3B) did not yield any significant difference either (Fig. 4C). The two models had R^2 values of 0.10 and 0.09 respectively, which was much higher than for most of the other edges, but substantially lower than the R^2 values obtained with FUNCOIN (Fig. 4D).

To illustrate the problem of picking out a few edges for analysis, we conducted the same analysis as above (comparing the proportion of outliers in each diagnosis groups to the proportion among healthy test subjects) in each direction for each of the 1485 edges and calculated the unadjusted p-values with Fisher's exact test. For the BD and PD subjects respectively, 3.8% and 9.8% of the p-values were below 0.05 (Fig. 4E), suggesting that most of the differences observed in the localized, edgewise model may be false positives. With no other criteria, it becomes impossible to sift the relevant models out of the 2970 hypotheses per subject group. After correcting for multiple comparisons with the Benjamini-Yekutieli procedure at level $\alpha=0.05$, all adjusted p-values were above 0.15 with only 1% being below 0.95, again illustrating the multiple comparisons problem of analyzing connection by connection.

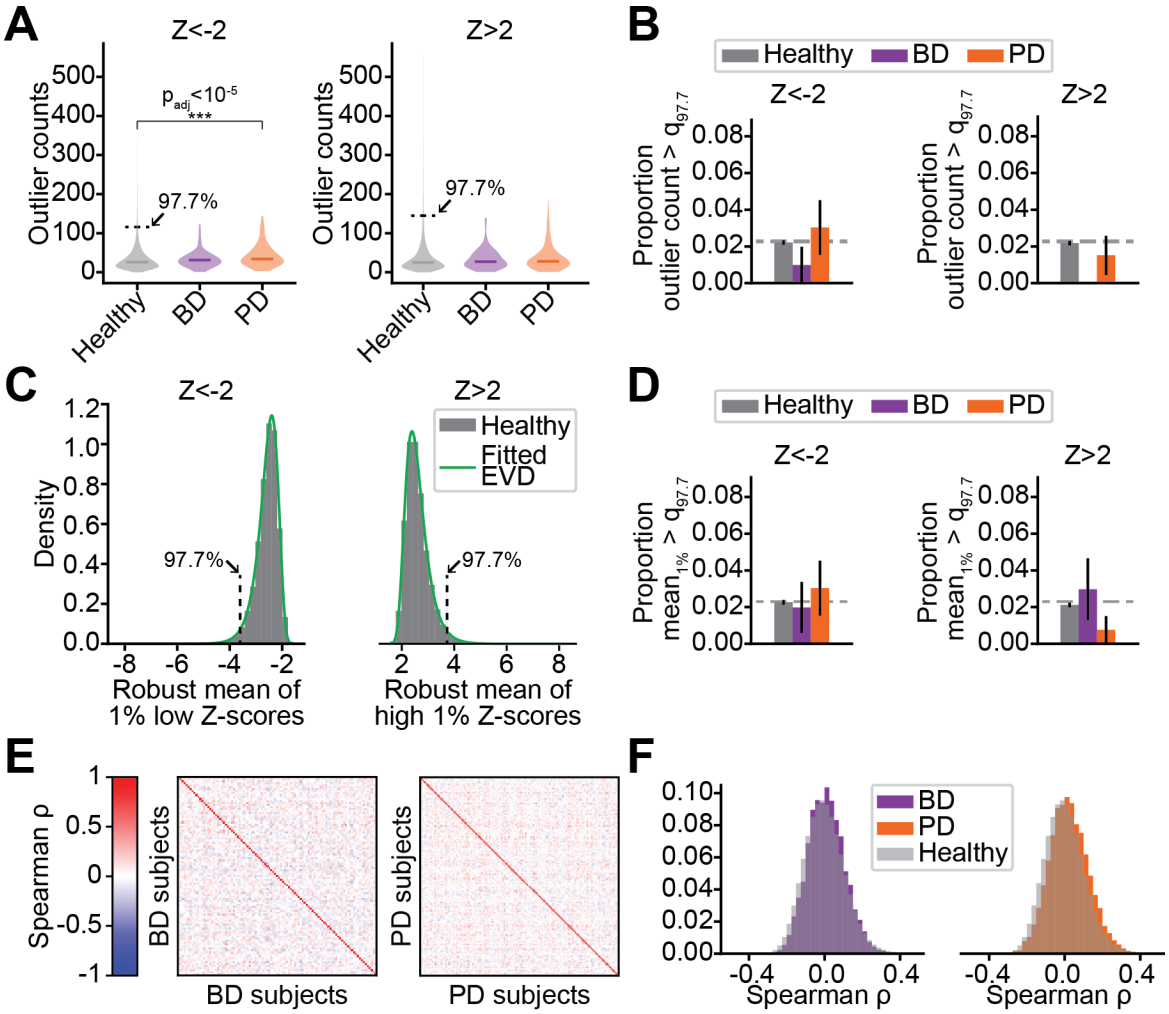

Figure 5: Identifying group, but not subject level, differences by summarizing edgewise *Z*-scores. A. Summarizing the *Z*-scores by for each subject counting the number of outlier edges for *Z*<-2 (left) and *Z*>2 (right). Violin plots show the density of count of outliers in each subject group. Mann-Whitney U test reveals significant group difference between healthy and PD subjects in the *Z*<-2 outlier count. Arrow points to the 97.7% percentile used as threshold in B. **B.** To identify outlier subjects, we threshold the outlier count in each direction by the 97.7% percentile in the training data. We find no significant difference in proportion of outlier subjects between healthy and BD/PD subjects. **C.** Histograms of robust means (i.e. 90% trimmed) of 1% lowest (left) and highest (right) *Z*-scores (grey bars) as well as the probability density functions of the fitted extreme value distributions (green). Arrows point to the 97.7% percentile of the fitted distribution in each direction. **D.** Thresholding at the 97.7 percentile of the robust means of the 1% most extreme *Z*-scores in each direction show no significant difference in proportion of outlier subjects between healthy and BD/PD subjects. **E.** Heatmap of Spearman correlations between outliers in BD (left) and PD (right) subjects reveal a large diversity in the distribution of outlier edges. **F.** Histograms of Spearman correlation values plotted in E reveal small correlation values centered around 0, which indicates large heterogeneity in which edges are outliers.

Integrating over multiple localized models can only capture deviances at the group level

Instead of choosing a few edges for analysis, we now summarized the *Z*-scores for each subject with two different methods proposed in previous normative modelling studies. First, we counted the number of outlier connections for each subject, as in the group comparison analyses from previous, localized, normative modelling studies[6,32]. Comparing the distribution of counts of outliers ($Z<-2$ and $Z>2$) in the healthy and PD groups, we found significant differences when counting $Z<-2$ ($p_{adj}<10^{-5}$ from Mann-Whitney U test, Fig. 5A). However, this is a group difference analysis which identifies differences in distribution between the two groups but cannot pick out individual outlier subjects —one of the fundamental goals of normative modelling. Therefore, to identify outlier subjects, i.e. subjects with more outlier edges than what was observed in healthy subjects, we took the 97.7% percentile of outlier edge counts in the training set, i.e. the percentile corresponding to one-sided *Z*-score of 2, as a threshold for identifying outlier subjects. We saw no significant differences in the proportion of outlier subjects between the healthy and the BD and PD groups respectively (Fig. 5B). Restricting this analysis to PD subjects scanned before diagnosis also showed a group difference but no significant results in the subject level prediction (Supplementary Fig. 4A-B). Counting deviations in both directions together (i.e. $|z|>2$) also led to the same conclusions (Supplementary Fig. 4C-D).

Previous studies have also suggested to summarize large numbers of *Z*-scores by, for each subject, taking the 1% most extreme *Z*-scores, removing the 10% most extreme of these, and computing the mean. This is followed by fitting an extreme value distribution (EVD) to the "robust means" obtained in this way from the training data. A given percentile of the fitted EVD can then be a threshold for identifying outlier subjects[13,33-37]. Like the counting of number of outlier edges (see above), this method has identified group level differences in various settings. In our edgewise model, the computed robust means for the most extreme *Z*-scores in each direction were well fitted by extreme value distributions (Fig. 5C). However, thresholding the fitted distributions at the 97.7 percentile, we saw no significant difference in proportion of outlier subjects (i.e. subjects with an extremely large robust average of extreme *Z*-scores) among the healthy subjects and subjects with BD or PD (Fig. 5D). Carrying out the same analysis on subjects scanned before their PD diagnosis (Supplementary Fig. 4E) as well as summarizing absolute *Z*-scores this way (Supplementary Fig. 4F-H) also showed no significant results.

While the methods for summarizing the *Z*-scores from all individual connections can find group differences, it is possible that they suffer from lack of specificity which basically makes the individual subjects in the groups incomparable at the subject level and prevents us from identifying the dimensions in which they tend to deviate. Therefore, to address the similarity of the edgewise *Z*-score patterns between subjects, we computed the Spearman correlation of *Z*-scores for each pair of subjects in the BD and PD groups (Fig. 5E). The average Spearman correlation values were 0.01 and 0.02 for the BD and PD groups respectively, which shows that essentially no common outlier patterns were found. The average Spearman correlation between healthy subjects was 0.0005. Comparing the distributions of Spearman correlation values in BD and PD subjects against healthy subjects, BD and PD subjects appeared somewhat right-shifted (towards *Z*-scores being more positively correlated), but the differences were small (Fig. 5F). While this could indicate that the diseases have highly individualized effects on brain function, it cannot be ruled out that this simply owes to statistical noise. With FUNCOIN, however, we find specific dimensions of deviation in subjects with brain diseases, even though the *Z*-scores from individual connections appear to be distributed as randomly as in the healthy subjects. This suggests that the heterogeneity in *Z*-scores may be caused by the fact that individual connections cannot capture the network level effects of brain pathology —supporting our claim that normative modelling should be performed at an integrative, whole-brain level to be able to characterize complex brain diseases.

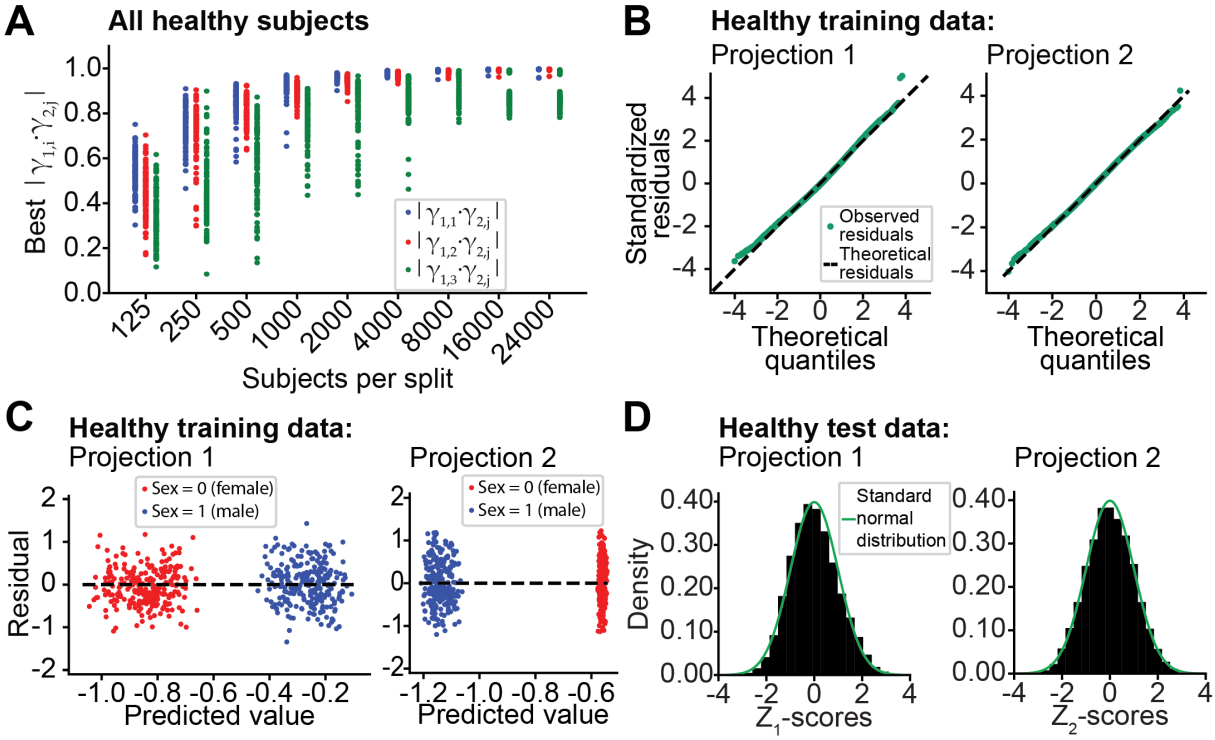

Figure 6: Model validation and sanity checks
A. Assessing consistency for different choices of number of subjects by repeatedly splitting the group in two random subgroups and estimating the first 5 γ components. Each of the first three γ in group 1, ($\gamma_{1,1}$, $\gamma_{1,2}$, $\gamma_{1,3}$), are compared to the 5 identified γ in group 2, ($\gamma_{2,1}$, …, $\gamma_{2,5}$), by taking the absolute dot product. A value of 1 indicates perfect alignment. The maximal value for each of the three first projections in group 1 are displayed for each of 100 random splits at different sample sizes. Consistency of the identification of the first two components (blue and red) is reached at around 2000 subjects, while component 3 is less consistently identified in all sample sizes up to 24000. **B.** QQ-plots from the training data give us no reason to suspect non-normality of residuals, which is assumed in model fitting and *Z*-score interpretation. **C.** Residuals vs. fitted value for the training data (250 random points for each sex and each projection are visualized) show variance homogeneity of the residuals, which is a necessary assumption for the model fitting and the *Z*-score interpretation. **D.** 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 shows good generalizability and calibration of the model to out-of-sample subjects.

Model validation and calibration

To make sure the FUNCOIN modelling framework is appropriate for normative modelling, we carefully assessed model assumptions and performance. We enumerate the following requirements: consistency, unbiasedness, statistical validity, calibration, and scalability.

We first evaluated consistency of the identified $\boldsymbol{\gamma}$ projections by for different sample sizes repeatedly dividing subsets of the healthy subjects into two random, equally sized

groups and fitting the model. This showed consistent estimation of the $\boldsymbol{\gamma}$ projections for sample sizes larger than 2000 subjects (see Methods, Fig. 6A). Second, normative modelling requires unbiased model predictions for the *Z*-scores to be valid, meaning that the model predicts the averages observed in the data, which was the case after our modification of the originally proposed regression method[24] (see Methods and Supplementary Fig. 1). With respect to statistical validity and calibration, Q-Q plots of training and test data showed that the residuals were normally distributed (Fig. 6B, Supplementary Fig. 5A), while plots of residuals against fitted values showed homogenous residual variance (Fig. 6C, Supplementary Fig. 5B). This ensures the validity of the model fit. The out-of-sample *Z*-scores were well calibrated in the sense that they followed standard normal distributions (Fig. 6D, skew values of 0.21 and 0.02, excess kurtosis of -0.14 and -0.21). The data come from 4 different scanning sites. In line with earlier normative modelling studies on structural MRI from UKB[37,38], we found no scanning site effects on the *Z*-scores of the healthy subjects (see Supplementary text and Supplementary Fig. 6). Finally, normative models rely on fitting to large datasets making scalability crucial. Our model 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 (Supplementary Fig. 7), which shows that the method is feasible for large-scale normative models.

Discussion

In this paper, we argue that normative modelling of rsfMRI FC is best done on an integrative level, and, to be able to do that, we propose a whole-brain FC normative modelling framework, FUNCOIN. We illustrate the usability of FUNCOIN in detecting brain diseases and contrast it to a common edgewise model. 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 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 up to 8 years before the time of diagnosis. The subjects with BD divided into two subgroups based on their different alterations in FC. Third, our model circumvents several challenges from edgewise models by simultaneously performing covariate-relevant dimensionality reduction and regression. Thereby, the relevant

information for FC regression and normative modelling is extracted while massive hypothesis testing is avoided and specificity of the found deviations is increased.

Subtyping heterogenous patient cohorts

Normative modelling is tailored for subject-level inference and well-suited to differentiate between subtypes of disease in heterogenous patient groups[13]. For example, earlier normative modelling studies in structural MRI have revealed large heterogeneity in patients with BD[35,39]. In some normative modelling studies, it is argued that disease heterogeneity was found by observing at least 1 deviating brain region in most subjects with disease, and most brain regions had at least 1 outlier subject[32,40]. This argument for disease heterogeneity can be problematic. With an a priori outlier chance of $P(|Z|>2)=0.046$ in any edge, the probability of observing at least one outlier subject by chance is 99.1% and 99.8% respectively in the BD (N=101) and PD (N=132) subjects respectively. Indeed, in our edgewise model with the BD and PD subjects, there was at least 1 outlier subject in 98.7% and 99.5% of the edges, respectively, and in both groups every subject had outlier values in several edges. This highlights the need of taking both a priori probability and number of subjects into account when examining disease heterogeneity. Based on our analysis in Figure 5, we argue that the observed heterogeneity may be due to coincidence as would be expected in independent models of each edge.

In contrast, the different Z-scores from FUNCOIN come from a few well-defined, orthogonal components of the FC and therefore convey information about distinct, well-defined parts of FC. By differentiating between subjects being outliers in different components, our method has the potential identify subtypes of brain diseases with different underlying pathologies in specific dimensions. In our results, we observed this in the group of subjects with BD, where no subjects had both $Z_1>2$ and $Z_2<-2$ (Fig. 3A). This indicates two different subgroups of BD. Future studies applying our method to more patients with a known disease subtype can explore this further.

Other normative models in rsfMRI

Previous normative models in rsfMRI are all based on predicting FC. Most studies have utilized edgewise methods such as versions of linear or polynomial regression[6,29,34,37],

Gaussian-Gamma mixture models[30], or generalized additive models[41]. Two studies used Gaussian processes to model the ROI-wise summed FC strength[32,40], and an earlier study applied dimensionality reduction of the FC with group-ICA before regression of the components[42].

As shown above, one challenge inherent to the edgewise approach is massive hypothesis testing. Strategies for dealing with this include correlation of extreme *Z*-scores to cognitive trait scores[37], normative probability maps[13], counting above-threshold *Z*-scores[6,32] and the use of extreme value distributions[13,33-37]. Here we investigated the latter two, which are commonly used. Like in earlier studies, we find that summarizing a large amount of *Z*-scores can reveal group differences, but, as we show, it comes at the cost of losing specificity and moving away from subject level inference, which are otherwise important advantages of normative modelling.

The inherent risk of very bad model fits in some edges pose another problem for the summarizing methods. 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[43]. Gaussian processes are extremely flexible and avoid the problem of bad model fits, but they do not scale well beyond a few thousand participants[37]. The use of likelihood warping in BLR can (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[37].

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 482 time points of 55 ICA components (Supplementary Fig. 7).

Finally, the approach of Kessler et al, applying dimensionality reduction with group-ICA followed by prediction of ADHD diagnosis from the *Z*-scores, is closer to our approach[42]. 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 is key for achieving a higher statistical power.

Limitations of our study

As in other normative modelling studies of FC, interpretation of the *Z*-scores can be challenging. First, we do not know a priori whether 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)[13]. 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, despite the impact of disease. Evaluating subjects before intervention might increase the disease prediction accuracy, and 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[44]. 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 for 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.

Materials and methods

Data

We use rsfMRI time series data from the UK-Biobank[25], which contains rsfMRI data from more than 60k subjects. 46,000 subjects without brain diagnoses (International Classification of Disease 10 (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 out-of-sample testing)[7]. For disease prediction, we used data from subjects with BD (ICD-10 F31, *N*=101) and PD (ICD-10 G20, *N*=132). The time of diagnosis used for calculating the time between diagnosis and scan (Figs. 3-5) is the time of first occurrence in health records.

The rsfMRI scans were performed using a Siemens Skyra 3T with acquisition parameters harmonized across the four scanning sites. We used the UKB ICA parcellation with 55 ICA components[26,45,46]. Details on data acquisition and preprocessing can be found in previous UKB papers[47,48].

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

The UKB has a Research Tissue Bank (RTB) approval from North West Multi-centre Research Ethics Committee (MREC). No separate ethics approval was needed for this study.

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. The methods differ in two ways: 1) The actual strengths of the projections found with covariate-assisted principal regression are biased, which is corrected in FUNCOIN (see below); 2) We make a correction to the rank-completion step originally proposed (see supplementary materials).

The FC matrix is high-dimensional and positive semidefinite. The method performs dimensionality reduction and regression in one step while respecting the definiteness of the FC by finding components, $\boldsymbol{\gamma} \in \mathbb{R}^p$, of a set of p-by-p FC matrices such that the components are shared among the subjects and their strengths are most associated with covariates:

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

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-1 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 but limited by the 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 from all subjects' data (see supplementary materials). In this study, the time series data were standardized to mean 0 and variance 1 before the model fitting. The components identified are

of the covariance matrix with T degrees of freedom, which for the standardized data is the Pearson correlation matrix.

Consistency of identified projections

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 identifying up to 5 components (Fig. 6A). 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 assessed consistency of the estimated projections 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 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 approximately $N \geq 2000$. The third projection was not consistently identified for $N \leq 24000$. The number of subjects required for consistency depends, among other factors, on data quality parameters, the number of time points, and number of brain regions, p.

Resolving bias in the model

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 yield skew residuals making the definition of outliers as $|Z|>2$ biased. We reproduced the simulation study from Zhao et al[24] (see Supplementary text) and found $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ in agreement with values reported by Zhao et al

(Supplementary Tables 4-5)[24]. However, the estimations of β_0 in both projections were biased and had low confidence interval coverage probabilities (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 confirmed the intercept bias. Average observed values in 1-year age groups were in general lower than predicted by the model (Supplementary Fig. 1). To remove bias on the coefficients, we therefore added another step to the algorithm: After having identified projections (and coefficients), we perform linear regression on the transformed FC values to get unbiased beta estimates. That is, for each projection, j, we fit β_0 and $\boldsymbol{\beta}$ in the following model:

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

with $\boldsymbol{\gamma}_j \in \mathbb{R}^p$, $\boldsymbol{\Sigma_i} \in \mathbb{R}^{pxp}$, $\boldsymbol{x_i} \in \mathbb{R}^{q-1}$, $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^{q-1}$ defined as above, and $\varepsilon_{j,i} \sim N(0, \sigma_j^2)$ being the residuals in projection j. This relies on the established consistency and low bias of the $\boldsymbol{\gamma}$ estimates and on training data residuals in each component having homogenous variance (see model validation below). This modified, unbiased version suited for normative modelling is FUNCOIN.

We also made a correction to the rank-completion step proposed in covariate-assisted principal regression (see text in supplementary materials).

Model fitting

We included sex, age, and sex-age interaction as covariates. The age variable was linearly transformed the interval [0,1], with 0 and 1 corresponding to the minimum and maximum age in the whole dataset (44.56 and 85.43 years). Sex was coded as a categorical variable with 0 denoting female, 1 denoting male. The data come from 4 different scanning sites. When exploring potential scanning site effects (Supplementary text and Supplementary Fig. 6), scanning site was included as an additive, categorical variable and was dummy coded with site 1 as reference. The model was fitted by for several choices of initial conditions optimizing the log-likelihood function and applying the bias-correction step (see above). See Supplementary Material for details.

Selecting the number of components

As $\boldsymbol{\gamma}$ components are identified sequentially, 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 $\boldsymbol{\gamma_2}$, but not $\boldsymbol{\gamma_3}$, were consistently identified in our study (see above). Second, when we fitted a model including site effects, the two first identified $\boldsymbol{\gamma}$ had site coefficients close to zero, while some scanning site effect was estimated in the third $\boldsymbol{\gamma}$ (see Supplementary material).

Statistical analysis

In Figures 3-4, to test whether each group of diagnosed subjects had significantly different proportion of outliers compared to the group of healthy out-of-sample subjects, we obtained *p*-values from permutation testing with 10,000 iterations. For each iteration in each diagnosis, we calculated Fisher's exact test (i.e. from a 2-by-2 contingency table) for each test (e.g. $|z|>2$, $z<-2$, etc.). To compare the distribution of the count of *Z*-scores between healthy and diagnosed subjects (Fig. 5A, Supplementary Fig. 4A,C-D), we applied Mann-Whitney U test. In Figure 5, permutation testing was used to test if diagnosed subjects had a larger proportion of outliers than healthy. Statistics of regression coefficients (Supplementary Table 1) was found using two-sided t-test as implemented in the FUNCOIN python package. All *p*-values in each figure panel were FDR-corrected with the Benjamini-Yekutieli procedure at level $\alpha=0.05$. See Supplementary Material and Supplementary Table 6 for implementation details, test statistics, and *p*-values.

Software

With this paper, a python package, 'FUNCOIN', is released (https://github.com/kobbersmed/funcoin), installable directly from GitHub or with pip. Features include fitting to data, applying the model to out-of-sample data, estimating confidence intervals parametrically or by bootstrapping, calculating *Z*-scores, performing statistical tests, and more. See the accompanying documentation and tutorial.

Data and code availability

The data used are part of the UK Biobank, an open-access data repository, and is available upon application to the UK Biobank.

Model training and testing, *Z*-score calculation, and calculation of model coefficient statistics were performed using the FUNCOIN Python package.

Code used for utilizing FUNCOIN and analyzing results as well as an instructive README file is available on GitHub (https://github.com/kobbersmed/Paper_Normative_modelling_of_brain_function).

Acknowledgements

We thank Yi Zhao, Indiana University School of Medicine, who first proposed the method that this paper builds on, for valuable correspondence and for readily answering any inquiries.

Funding

JRLK 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), an ERC Starting Grant (ERC-StG-2019-850404), and from the Independent Research Fund Denmark (2034-00054B). AM gratefully acknowledges support from an ERC consolidator grant (101001118) 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 Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Declaration of competing interest

The authors report no competing interests.

Supplementary material

Supplementary material is available online.

References

1. Calhoun VD, Pearlson GD, Sui J. Data-driven approaches to neuroimaging biomarkers for neurological and psychiatric disorders: emerging approaches and examples. *Curr Opin Neurol*. Aug 1 2021;34(4):469-479. doi:10.1097/WCO.0000000000000967
2. Lotharius J, Brundin P. Pathogenesis of parkinson's disease: dopamine, vesicles and α-synuclein. *Nature Reviews Neuroscience*. 2002/12/01 2002;3(12):932-942. doi:10.1038/nrn983
3. Fornito A, Zalesky A, Breakspear M. The connectomics of brain disorders. *Nat Rev Neurosci*. Mar 2015;16(3):159-72. doi:10.1038/nrn3901
4. Zhang J, Kucyi A, Raya J, et al. What have we really learned from functional connectivity in clinical populations? *Neuroimage*. Nov 15 2021;242:118466. doi:10.1016/j.neuroimage.2021.118466
5. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry*. Oct 2019;24(10):1415-1424. doi:10.1038/s41380-019-0441-1
6. Rutherford S, Barkema P, Tso IF, et al. Evidence for embracing normative modeling. *Elife*. Mar 13 2023;12doi:10.7554/eLife.85082
7. Rutherford S, Kia SM, Wolfers T, et al. The normative modeling framework for computational psychiatry. *Nat Protoc*. Jul 2022;17(7):1711-1734. doi:10.1038/s41596-022-00696-5
8. Cai N, Choi KW, Fried EI. Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. *Human Molecular Genetics*. 2020;29(R1):R10-R18. doi:10.1093/hmg/ddaa115
9. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine*. 2013/05/14 2013;11(1):126. doi:10.1186/1741-7015-11-126
10. Hyman SE. The Diagnosis of Mental Disorders: The Problem of Reification. *Annual Review of Clinical Psychology*. 2010;6(Volume 6, 2010):155-179. doi:https://doi.org/10.1146/annurev.clinpsy.3.022806.091532
11. Insel T, Cuthbert B, Garvey M, et al. Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*. 2010/07/01 2010;167(7):748-751. doi:10.1176/appi.ajp.2010.09091379
12. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. Dec 2012;17(12):1174-9. doi:10.1038/mp.2012.105
13. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry*. Oct 1 2016;80(7):552-61. doi:10.1016/j.biopsych.2015.12.023
14. Moriarity DP, Joyner KJ, Slavich GM, Alloy LB. Unconsidered issues of measurement noninvariance in biological psychiatry: A focus on biological phenotypes of psychopathology. *Mol Psychiatry*. Mar 2022;27(3):1281-1285. doi:10.1038/s41380-021-01414-5
15. Fraza C, Rutherford S, Buckova BR, Beckmann CF, Marquand AF. The promise of quantifying individual risk for brain disorders through normative modeling, a narrative

review. *Neurosci Biobehav Rev*. Jul 11 2025;176:106284. doi:10.1016/j.neubiorev.2025.106284
16. Margulies DS, Ghosh SS, Goulas A, et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc Natl Acad Sci U S A*. Nov 1 2016;113(44):12574-12579. doi:10.1073/pnas.1608282113
17. Power JD, Cohen AL, Nelson SM, et al. Functional network organization of the human brain. *Neuron*. Nov 17 2011;72(4):665-78. doi:10.1016/j.neuron.2011.09.006
18. Rasero J, Aerts H, Ontivero Ortega M, Cortes JM, Stramaglia S, Marinazzo D. Predicting functional networks from region connectivity profiles in task-based versus resting-state fMRI data. *PLoS One*. 2018;13(11):e0207385. doi:10.1371/journal.pone.0207385
19. Smith SM, Fox PT, Miller KL, et al. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*. 2009;106(31):13040-13045. doi:doi:10.1073/pnas.0905267106
20. Smith SM, Vidaurre D, Beckmann CF, et al. Functional connectomics from resting-state fMRI. *Trends Cogn Sci*. Dec 2013;17(12):666-82. doi:10.1016/j.tics.2013.09.016
21. Gratton C, Koller JM, Shannon W, et al. Emergent Functional Network Effects in Parkinson Disease. *Cereb Cortex*. Jun 1 2019;29(6):2509-2523. doi:10.1093/cercor/bhy121
22. Tinaz S, Lauro P, Hallett M, Horovitz SG. Deficits in task-set maintenance and execution networks in Parkinson's disease. *Brain Struct Funct*. Apr 2016;221(3):1413-25. doi:10.1007/s00429-014-0981-8
23. Chai XJ, Whitfield-Gabrieli S, Shinn AK, et al. Abnormal medial prefrontal cortex resting-state connectivity in bipolar disorder and schizophrenia. *Neuropsychopharmacology*. Sep 2011;36(10):2009-17. doi:10.1038/npp.2011.88
24. Zhao Y, Wang B, Mostofsky SH, Caffo BS, Luo X. Covariate Assisted Principal regression for covariance matrix outcomes. *Biostatistics*. Jul 17 2021;22(3):629-645. doi:10.1093/biostatistics/kxz057
25. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*. 2015;12(3)doi:10.1371/journal.pmed.1001779
26. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging*. Feb 2004;23(2):137-52. doi:10.1109/TMI.2003.822821
27. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*. 2001:1165-1188.
28. Bloem BR, Okun MS, Klein C. Parkinson's disease. *Lancet*. Jun 12 2021;397(10291):2284-2303. doi:10.1016/S0140-6736(21)00218-X
29. Jalbrzikowski M, Murty VP, Tervo-Clemmens B, Foran W, Luna B. Age-Associated Deviations of Amygdala Functional Connectivity in Youths With Psychosis Spectrum Disorders: Relevance to Psychotic Symptoms. *Am J Psychiatry*. Mar 1 2019;176(3):196-207. doi:10.1176/appi.ajp.2018.18040443
30. Looden T, Floris DL, Llera A, et al. Patterns of connectome variability in autism across five functional activation tasks: findings from the LEAP project. *Mol Autism*. Dec 27 2022;13(1):53. doi:10.1186/s13229-022-00529-y

31. Smith BB, Zhao Y, Lindquist MA, Caffo B. Regression models for partially localized fMRI connectivity analyses. *Front Neuroimaging*. 2023;2:1178359. doi:10.3389/fnimg.2023.1178359
32. Sun X, Sun J, Lu X, et al. Mapping Neurophysiological Subtypes of Major Depressive Disorder Using Normative Models of the Functional Connectome. *Biol Psychiatry*. Dec 15 2023;94(12):936-947. doi:10.1016/j.biopsych.2023.05.021
33. Kia SM, Marquand A. Normative Modeling of Neuroimaging Data Using Scalable Multi-task Gaussian Processes. Springer International Publishing; 2018:127-135.
34. Jiang A, Ma X, Li S, et al. Age-atypical brain functional networks in autism spectrum disorder: a normative modeling approach. *Psychol Med*. Jul 2024;54(9):2042-2053. doi:10.1017/S0033291724000138
35. Wolfers T, Rokicki J, Alnaes D, et al. Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. *Hum Brain Mapp*. Jun 1 2021;42(8):2546-2555. doi:10.1002/hbm.25386
36. Zabihi M, Oldehinkel M, Wolfers T, et al. Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. *Biol Psychiatry Cogn Neurosci Neuroimaging*. Jun 2019;4(6):567-578. doi:10.1016/j.bpsc.2018.11.013
37. Fraza CJ, Dinga R, Beckmann CF, Marquand AF. Warped Bayesian Linear Regression for Normative Modelling of
Big Data. 2021;doi:10.1101/2021.04.05.438429
38. Kia SM, Huijsdens H, Rutherford S, et al. Closing the life-cycle of normative modeling using federated hierarchical Bayesian regression. *PLoS One*. 2022;17(12):e0278776. doi:10.1371/journal.pone.0278776
39. Wolfers T, Doan NT, Kaufmann T, et al. Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry*. Nov 1 2018;75(11):1146-1155. doi:10.1001/jamapsychiatry.2018.2467
40. Liu L, Jia D, He Z, Wen B, Zhang X, Han S. Individualized functional connectome abnormalities obtained using two normative model unveil neurophysiological subtypes of obsessive compulsive disorder. *Prog Neuropsychopharmacol Biol Psychiatry*. Dec 20 2024;135:111122. doi:10.1016/j.pnpbp.2024.111122
41. Sun L, Zhao T, Liang X, et al. Human lifespan changes in the brain's functional connectome. *Nat Neurosci*. Apr 2025;28(4):891-901. doi:10.1038/s41593-025-01907-4
42. Kessler D, Angstadt M, Sripada C. Growth Charting of Brain Connectivity Networks and the Identification of Attention Impairment in Youth. *JAMA Psychiatry*. May 1 2016;73(5):481-9. doi:10.1001/jamapsychiatry.2016.0088
43. Savage HS, Mulders PCR, van Eijndhoven PFP, et al. Dissecting task-based fMRI activity using normative modelling: an application to the Emotional Face Matching Task. *Commun Biol*. Jul 20 2024;7(1):888. doi:10.1038/s42003-024-06573-z
44. Bučková BR, Fraza C, Rehák R, et al. Using normative models pre-trained on cross-sectional data to evaluate longitudinal changes in neuroimaging data. eLife Sciences Publications, Ltd; 2024.
45. Griffanti L, Salimi-Khorshidi G, Beckmann CF, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage*. Jul 15 2014;95:232-47. doi:10.1016/j.neuroimage.2014.03.034
46. Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. Automatic denoising of functional MRI data: combining independent component

analysis and hierarchical fusion of classifiers. *Neuroimage*. Apr 15 2014;90:449-68. doi:10.1016/j.neuroimage.2013.11.046
47. Alfaro-Almagro F, Jenkinson M, Bangerter NK, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage*. Feb 1 2018;166:400-424. doi:10.1016/j.neuroimage.2017.10.034
48. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. Nov 2016;19(11):1523-1536. doi:10.1038/nn.4393