Original Article

CAMBRIDGE
UNIVERSITY PRESS

Brain functional connectivity and anatomical features as predictors of cognitive behavioral therapy outcome for anxiety in youths

Andre Zugman[1] 🟢, Grace V. Ringlein[1], Emily S. Finn[2], Krystal M. Lewis[1],
Erin Berman[1], Wendy K. Silverman[3], Eli R. Lebowitz[3], Daniel S. Pine[1] and
Anderson M. Winkler[1,4]

[1]Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA; [2]Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA; [3]Child Study Center, Yale University, New Haven, CT, USA and [4]Division of Human Genetics, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA

Abstract

Background. Because pediatric anxiety disorders precede the onset of many other problems, successful prediction of response to the first-line treatment, cognitive-behavioral therapy (CBT), could have a major impact. This study evaluates whether structural and resting-state functional magnetic resonance imaging can predict post-CBT anxiety symptoms.

Methods. Two datasets were studied: (A) one consisted of $n = 54$ subjects with an anxiety diagnosis, who received 12 weeks of CBT, and (B) one consisted of $n = 15$ subjects treated for 8 weeks. Connectome predictive modeling (CPM) was used to predict treatment response, as assessed with the PARS. The main analysis included network edges positively correlated with treatment outcome and age, sex, and baseline anxiety severity as predictors. Results from alternative models and analyses are also presented. Model assessments utilized 1000 bootstraps, resulting in a 95% CI for R^2, r, and mean absolute error (MAE).

Results. The main model showed a MAE of approximately 3.5 (95% CI: [3.1–3.8]) points, an R^2 of 0.08 [-0.14–0.26], and an r of 0.38 [0.24–0.511]. When testing this model in the left-out sample (B), the results were similar, with an MAE of 3.4 [2.8–4.7], $R^2 - 0.65$ [-2.29–0.16], and r of 0.4 [0.24–0.54]. The anatomical metrics showed a similar pattern, where models rendered overall low R^2.

Conclusions. The analysis showed that models based on earlier promising results failed to predict clinical outcomes. Despite the small sample size, this study does not support the extensive use of CPM to predict outcomes in pediatric anxiety.

Introduction

As pediatric anxiety disorders precede the onset of most persistent adult emotional problems (Gregory et al., 2007; Nelemans et al., 2014; Pine, Cohen, Gurley, Brook, & Ma, 1998; Woodward & Fergusson, 2001), successful treatment could exert a long-term impact. However, cognitive behavior therapy (CBT), a first-line treatment, produces remission only in less than half of all cases (Ginsburg et al., 2011; Piacentini et al., 2014; Silverman, Pina, & Viswesvaran, 2008). Because CBT is time-consuming, identifying reliable predictors of treatment outcomes could markedly influence practice. Clinical features, such as comorbidity or severity, only partially predict outcomes (Kunas, Lautenbacher, Lueken, & Hilbert, 2021).

Magnetic resonance imaging (MRI) indices may be able to predict outcomes beyond such clinical features. Measurements derived from MRI are reliable, scalable, and already used in relatively large samples (Miller et al., 2016). This study applies a predictive framework with resting-state functional connectivity (rsFC) and structural MRI (sMRI) in medication-free children seeking treatment for anxiety disorders. Two samples are studied, each receiving CBT by trained experts, to support a three-step approach. This begins with model building, followed by cross-validation in the first, larger sample. The approach ends with model testing in the smaller, held-out sample.

This study extends considerable research (Dubois & Adolphs, 2016; Mueller et al., 2013) using rsFC to model an individual's 'connectome' computed by correlating signals among network 'nodes' (Sporns, 2011). Connectome predictive modeling (CPM) (Shen et al., 2017) generates clinical insights by correlating edgewise rsFC matrices with clinical measures and pooling associations in a second prediction stage. CPM can predict important constructs, such as intelligence (Finn et al., 2015; Gao, Greene, Constable, & Scheinost, 2019; Greene, Gao, Scheinost, & Constable, 2018), attention (Rosenberg, Finn, Scheinost, Constable, & Chun, 2017; Rosenberg et al., 2016), and anxiety (Ren et al., 2021; Wang et al., 2021). Although promising, CPM is still

understudied, as is rsFC for anxiety disorders more broadly (Zugman, Jett, Antonacci, Winkler, & Pine, 2023). One study and a follow-up replication study used rsFC analyzed with methods different from CPM to predict treatment outcomes in anxiety (Ashar et al., 2021; Whitfield-Gabrieli et al., 2016). Both studies focused on adults with a diagnosis of social anxiety disorder, and the predictive model failed to replicate. A recent study that included adults who underwent CBT treatment for anxiety disorders in two adult cohorts found no successful predictive model for treatment response using different machine learning pipelines (Hilbert et al., 2024). No studies used CPM to predict treatment response in pediatric anxiety, and the available rsFC studies were small. Of note, while the sample size in this study is also small, it is larger than either of the two past studies. Across the three studies, small sample sizes reflect the difficulty of delivering state-of-the-art treatment to medication-free subjects along with brain imaging investigations. The primary goal of this study is to predict CBT response using CPM in pediatric anxiety disorders.

The secondary goal considers aspects of imaging reliability. sMRI generates measures with higher reliability than rsFC. Hence, sMRI could have advantages in predicting treatment response. However, rsFC, while less reliable (Hedges et al., 2022; Noble, Scheinost, & Constable, 2019) may identify subsets of stable features that relate more consistently than sMRI to clinical measures (Mansour L, Tian, Yeo, Cropley, & Zalesky, 2021). We term the use of sMRI in this framework 'anatomical predictive modeling (APM)' since no connectome is involved. Within the CPM framework, we compare the ability of sMRI and rsFC to predict treatment response.

Recent literature describes idiosyncratic rsFC patterns related to subject identity as akin to 'fingerprints'. These patterns may predict variables of clinical interest (Amico & Goñi, 2018; Byrge & Kennedy, 2020; Finn et al., 2015; Lin, Baete, Wang, & Boada, 2020). Recent research and commentary, however, suggest otherwise (Finn & Rosenberg, 2021; Mantwill, Gell, Krohn, & Finke, 2022). Thus, a third objective of this study is to assess whether MRI features that are most unique to individuals are relevant in predicting response to treatment.

Methods

Participants and measures

Anxious youth and healthy volunteers (HV) were recruited through referral to participate in the study at the National Institute of Mental Health (NIMH), National Institutes of Health (NIH), Bethesda, Maryland, United States, and enrolled in a protocol (01-M-0192; Principal Investigator: D.S.P.) for an ongoing clinical trial. Patients were considered for enrollment if they had a diagnosis of any DSM-5 anxiety disorder established by a licensed clinician using the KSADS (Kiddie Schedule for Affective Disorders and Schizophrenia). Exclusion criteria for all participants were a history of psychotic disorder, bipolar disorder, developmental disorders, obsessive-compulsive disorder, post-traumatic stress disorder, substance use disorder, contraindication to MRI scan, use of medication, or an estimated IQ lower than 70 (as measured by the Wechsler Abbreviated Scale of Intelligence). HV also were excluded if they had any current psychiatric diagnoses. All parents and research participants provided written informed consent/assent in a protocol approved by the NIH Institutional Review Board (IRB).

Symptom severity and treatment response was assessed using the Pediatric Anxiety Rating Scale (PARS) (The Research Units on Pediatric Psychopharmacology Anxiety Study Group, 2002), the gold-standard clinician-administered assessment incorporating both child and parent reports. The PARS was administered at four time points before, during (Weeks 3 and 8), and after treatment. The total PARS score ranges from 0 to 25, with a clinical cut-off of nine or higher indicating a likely presence of an anxiety disorder. In addition to CBT administered by experts, all patients received either an active or sham version of attention-bias modification therapy (ABMT). To maximize sample sizes, groups were combined irrespective of randomization to either active or sham ABMT. CBT in this sample was delivered using a standardized protocol, consisting of 12 weekly sessions (Silverman & Ginsburg, 1998; Silverman, Rey, Marin, Jaccard, & Pettit, 2022). The first three treatment sessions entail an introduction to CBT, psychoeducation, and self-monitoring/tracking. Starting at session four, participants complete in-session exposures and learn cognitive restructuring strategies and coping mechanisms (Lebowitz, Marin, Martino, Shimshoni, & Silverman, 2019). For additional details, see (Haller et al., 2024).

The above forms our main dataset (Dataset A). To determine whether the results obtained are replicable, we used a small sample of $N = 15$ (Dataset B) individuals who participated in a previous 8-week clinical trial to study the effects of CBT on pediatric anxiety (White et al., 2017), also under protocol 01-M-0192. It used a different resting state sequence, and CBT in this study followed the Coping Cat protocol (Podell, Mychailyszyn, Edmunds, Puleo, & Kendall, 2010). The eight sessions aim to develop skills to recognize signs of anxiety and anxious thoughts, relaxation techniques, and coping. Again, we use only patients with resting-state fMRI from up to 90 days before or 30 days after initiating treatment and who had available PARS at baseline and at the end of treatment (8 weeks). Demographic characteristics and MRI acquisition parameters for both datasets are described in Table 1 and Table 2 and in the Supplementary Material.

Connectome predictive modeling

The CPM analysis included only patients who had an available resting state fMRI scan collected up to 90 days before or 30 days after treatment initiation. As the primary objective was to assess treatment response, the main analysis excluded those who did not have available PARS at baseline or at 12 weeks. The baseline PARS came from either the screening visit or the third week of treatment, prior to the exposure-based portion of the treatment, and closer in time to the date of the MRI scan acquisition. A full sample description is provided in Supplementary Table 1. We used the methods outlined in Shen et al. (2017). Before conducting CPM, we verified that motion (as assessed via average framewise displacement) would not be a good predictor of the PARS score (Dataset A: $r = -0.2030$, $p = 0.1380$; Dataset B: $r = 0.0877$, $p = 0.7544$; two-tailed p-values assessed with 10,000 permutations). The rsFC matrices for every subject were tested for their association with PARS at the end of treatment; significantly ($p < 0.01$) associated edges were selected, and their rsFC Fisher's r-to-z values summed. These sums were used as independent variables in a second linear regression.

Cross-validation

Regression coefficients from the second model were used to predict, in a leave-one-out cross-validation loop, the PARS score of a subject

unseen in the previous steps. This analysis used the patients from Dataset A.

External validation

Regression coefficients from the second model with Dataset A were used to predict the PARS score of all patients in Dataset B.

Selection of significant edges in the initial step of CPM can consider edges that are positively correlated with PARS, negatively correlated, or both; edges can also be selected using other criteria. We departed from (Shen et al., 2017) in two aspects: (1) we investigated the inclusion of age and sex in the second regression model as predictors of interest and, separately, as nuisance, as well as without any such additional regressors as in the original publication; and (2) in addition to investigating the performance of using only positively correlated, negatively correlated, and both sets of edges, we also investigated the performance of CPM when using the most discriminative edges identified using fingerprinting; details of these two departures from the original method are provided below, and results from these various models are presented in the Supplementary Material. Model performance was assessed using the mean absolute error (MAE), the simple correlation coefficient (r) between observed (y_i) and predicted values ($\hat{y}_i$), and a version of the coefficient of determination (R^2) that is suitable for cross-validation and is computed as (Kvålseth, 1985):

$$R^2 = 1 - \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 / \sum_{i=1}^{n} \left(y_i - \bar{y}_i\right)^2$$

Note that R^2 does not correspond to the square of the correlation coefficient r; for commentary on the merits of each metric, see (Chicco, Warrens, & Jurman, 2021; Poldrack, Huckins, & Varoquaux, 2020). Confidence intervals (95%) were computed for these three quantities using 1000 bootstraps (Davison & Hinkley, 1997).

Anatomical predictive modeling

We investigated how replacing rsFC in CPM with measurements of brain morphology, which we termed APM, would impact predictions. To facilitate comparison with CPM, the APM analysis considers the same individuals. Surface-based representations of the brain were obtained with FreeSurfer 6.0.1, as part of fMRIprep processing, and resampled into the 'fsaverage5' space (a brain mesh with the same topology of a geodesic sphere produced by 5 recursive subdivisions of an icosahedron), which contains 20,484 vertices spanning both hemispheres; we removed those with constant variance, thus masking out non-cortical regions, to a total of 18,742 vertices used for analysis (compare to 23,220 unique edges used for analysis in the rsFC-based models). We investigated 5 different cortical morphometric measurements (area, thickness, curvature, sulcal depth, and gray/white matter contrast) and two levels of smoothing (FWHM = 0 and 15 mm).

Nuisance variables

The prediction may make use of other variables, such as age and sex, or consider these as a nuisance. In the former case, they are included as additional regressors in both stages of CPM and, subsequently, as additional predictors (Rao, Monteiro, & Mourao-Miranda, 2017). In the latter case, these variables are likewise included in the first regression model of CPM (that identifies edges), whereas in the second regression, both data and model are residualized with respect to these variables in the training set; the estimated regression coefficients from the training set are then to residualize also the test set (Snoek, Miletić, & Scholte, 2019); prediction uses then residualized variables, with coefficients of variables of interest and of no interest estimated from the training set. Nuisance effects can be added back to the predicted values to ensure results are compatible with the quantities of interest. If the data used for testing contains a substantial number of subjects, an improved model consists of residualizing the test set using estimated nuisance effects from the test set itself, as opposed to from the training set, thus reducing the risk of covariate shift (Rao et al., 2017). We investigated models without nuisance variables, as well as with age and sex (and scanner where appropriate). For the cross-validation case, in which it is not possible to estimate nuisance effects from test samples (in a leave-one out cross-validation, the test set has only one subject), we used the regression coefficients for nuisance variables obtained from the training set (Snoek et al., 2019), whereas for external validation, nuisance effects were estimated directly from the test set.

Fingerprinting

The fingerprinting analysis included only participants with at least two fMRI sessions within one year of each other. Subjects included in the primary data were allowed to be included in the fingerprinting analysis if they had a follow-up rsfMRI available. The sample characteristics are described in Supplementary Table 2. Fingerprinting with sMRI used the same individuals to facilitate comparison between the two approaches. We followed the methods outlined in (Finn et al., 2015). Each subject had two resting state scans, collected on average 123 days apart. The rsFC matrix was unwrapped into a vector, and then the Pearson's correlation coefficient between every baseline rsFC of every subject with every follow-up rsFC was computed, providing an index of similarity. Subject identification was successful if, for every baseline rsFC matrix, the most similar follow-up rsFC matrix belongs to the same subject. As the most similar follow-up is allowed to be repeated (i.e. with replacement), the p-value for the number of correct identifications (k) can be computed using a binomial distribution with parameters n = number of subjects, $p = 1/n$ and location $k = 1$.

Correlations can be interpreted as the dot product of vectors normalized to unit variance (Rodgers & Nicewander, 1988). This provides an indicator of the contribution of each edge to the final correlation. Let the correlation be expressed as (Finn et al., 2015):

$$r_{ij} = \sum_{e=1}^{M} \varphi_{ij}(e)$$

where M is the number of edges, $\varphi_{ij}(e) = x_i^b(e) x_j^f(e)$, $x_i^b(e)$ is the normalized value of the rsFC at edge e for subject i at baseline, and $x_j^f(e)$ is the normalized rsFC at the same edge for subject j at follow-up. If $i = j$, r_{ii} is the correlation between a subject's own baseline and follow-up rsFC matrices. The quantity φ is interesting because, if $\varphi_{ii}(e) \geq \varphi_{ij}(e)$ and $\varphi_{ii}(e) \geq \varphi_{ji}(e)$, then edge e contributes to the identification of the subject's rsFC at the other time point. An estimator of the probability $P_i(e)$ that an edge makes such contribution by chance is given by:

$$\hat{P}_i(e) = \left[\sum_{j=1}^{n} I\left(\varphi_{ij}(e) \geq \varphi_{ii}(e)\right) + \sum_{j=1}^{n} I\left(\varphi_{ji}(e) \geq \varphi_{ii}(e)\right) - 1\right] /$$
$$(2n - 1)$$

Table 1. Descriptive statistics for Datasets A and B, as used for the CPM and APM analyses. Additional sample details can be found in the Supplementary Material

	Dataset A (n = 54)		Dataset B (n = 15)			
	Mean or *n*	SD or %	Mean or *n*	SD or %	Test statistic*	*p*-value
Age	13.06	2.70	13.19	3.37	−0.17	0.87
Framewise displacement (mean)	0.11	0.05	0.14	0.06	−2.15	0.03
Framewise displacement (%)	9.79	8.88	15.40	13.34	−1.93	0.06
IQ (WASI)	112.39	13.42	115.53	17.17	−0.75	0.45
PARS total at baseline	14.63	2.84	16.73	2.89	−2.53	0.01
PARS total at follow-up	8.93	4.44	13.73	3.37	−3.88	<0.01
PARS difference	−5.70	4.39	−3.00	3.98	−2.15	0.04
Sex					0.01	0.93
Female	33	61%	10	67%		
Male	21	39%	5	33%		
*Diagnosis**					7.50	0.76
Panic disorder	2.0	4%	0.0	0%		
Separation anxiety	17.0	31%	5.0	33%		
Social phobia	33.0	61%	10.0	67%		
Specific phobia	14.0	26%	3.0	20%		
Generalized anxiety disorder	46.0	85%	11.0	73%		

*Two-sample *t*-test or Chi-squared test when appropriate.
**Patients may have more than one diagnosis, thus the sum is higher than 100%.

Table 2. Descriptive statistics for the fingerprinting sample (from Dataset A). Additional sample details can be found in the Supplementary Material

	Patients (n = 42)		Controls (n = 24)			
	Mean or *n*	SD or %	Mean or *n*	SD or %	Test statistic*	*p*-value
Time between scans (days)	128.20	40.68	113.54	38.93	1.42	0.16
Age at baseline	13.67	2.67	14.19	2.47	−0.78	0.44
Age at follow-up	13.93	2.63	14.50	2.45	−0.86	0.39
FD (mean) at baseline	0.11	0.05	0.10	0.05	0.17	0.87
FD (%) at baseline	9.20	9.23	9.88	11.86	−0.26	0.79
FD (mean) at follow-up	0.12	0.05	0.11	0.05	0.49	0.62
FD (%) at follow-up	10.21	9.51	10.24	10.07	−0.01	0.99
IQ (WASI)	111.33	13.06	111.83	12.67	−0.15	0.88
Sex					0.77	0.38
Female	27	64%	12	50%		
Male	15	36%	12	50%		
*Diagnosis**						
Panic disorder	2%	5%				
Separation anxiety	11%	26%				
Social phobia	27%	64%				
Specific phobia	13%	31%				
Generalized anxiety disorder	36%	86%				

*Two-sample *t*-test or Chi-squared test when appropriate.
**Patients may have more than one diagnosis, thus the sum is higher than 100%.

where $I(\cdot)$ is the indicator (Kronecker) function, and n is the number of subjects. Note that the above formulation is different than the one originally proposed by (Finn et al., 2015); edges that are highly predictive can have $\hat{P}_i(e)$ as low as $1/(2n-1)$, as opposed to zero; edges that are not predictive can have $\hat{P}_i(e)$ as high as 1. A global estimate of the differential power (DP) of a given edge for subject identification can be computed as:

$$DP(e) = -2\sum_{i=1}^{n} ln\big(\hat{P}_i(e)\big)$$

where the quantity $-ln\big(\hat{P}_i(e)\big)$ follows an exponential distribution with rate parameter 1 if the true (unknown) $P_i(e)$ follows a uniform distribution. The constant 2 adjusts that rate to 1/2. An exponential distribution with rate parameter 1/2 is a Chi-squared distribution with 2 degrees of freedom; the sum of n random variables following this distribution also follows a Chi-squared distribution, now with $2n$ degrees of freedom. Thus, the hypothesis that an edge is more informative than could be expected by chance can be tested. This formulation also allows the selection of edges (e.g. for later analyses, such as in CPM) using a threshold based on the probability distribution under the null hypothesis of chance DP. Note that the above formulation of $DP(e)$ is also different from the original work by (Finn et al., 2015).

Anatomical fingerprinting

Following (Mansour L et al., 2021), we also investigate fingerprinting using measures of cortical morphology: area, thickness, curvature, sulcal depth, and gray/white matter contrast, as opposed to only unwrapped rsFC matrices. Fingerprinting methods are otherwise the same as for rsFC data. Surface-based cortical measurements were as with APM.

Results

Prediction of anxiety scores using CPM

We report the main results for CPM using GSR, full (not partial) rsFC, positive edges without weighting, and both a model in which age, sex, and baseline PARS are used as predictors, as well as a model in which

data are residualized in relation to these variables. These results are emphasized since they led to generally superior accuracy across multiple analyses. CPM was not able to predict post-treatment 12-week PARS scores at a level that exceeded chance (Table 3). While using the simple correlation coefficient might give the impression of statistical robustness, the magnitude of relations between predicted and expected scores was only moderate, with an MAE of approximately 3.5 points (PARS scores range between 0 and 25). This MAE does approach a level that would be clinically useful (Walkup et al., 2001, 2008), but the correlation between predicted and observed PARS did not exceed 0.4. Additionally, no model showed a high R^2, and its 95% CI indicated that no model was better than chance. Overall, the low R^2 indicates that the models fail to predict above the mean of the target value. A scatter plot showing observed and predicted values for one of these models appears in Figure 1, upper left panel.

Of note, as expected, results were weaker in some analyses appearing in Supplemental Material, where the complete set of results, with varying processing choices, are provided. This included analyses predicting change scores and analyses using the smaller Dataset B to build the model, which was then tested in the larger Dataset A (Figure 1, lower left panel).
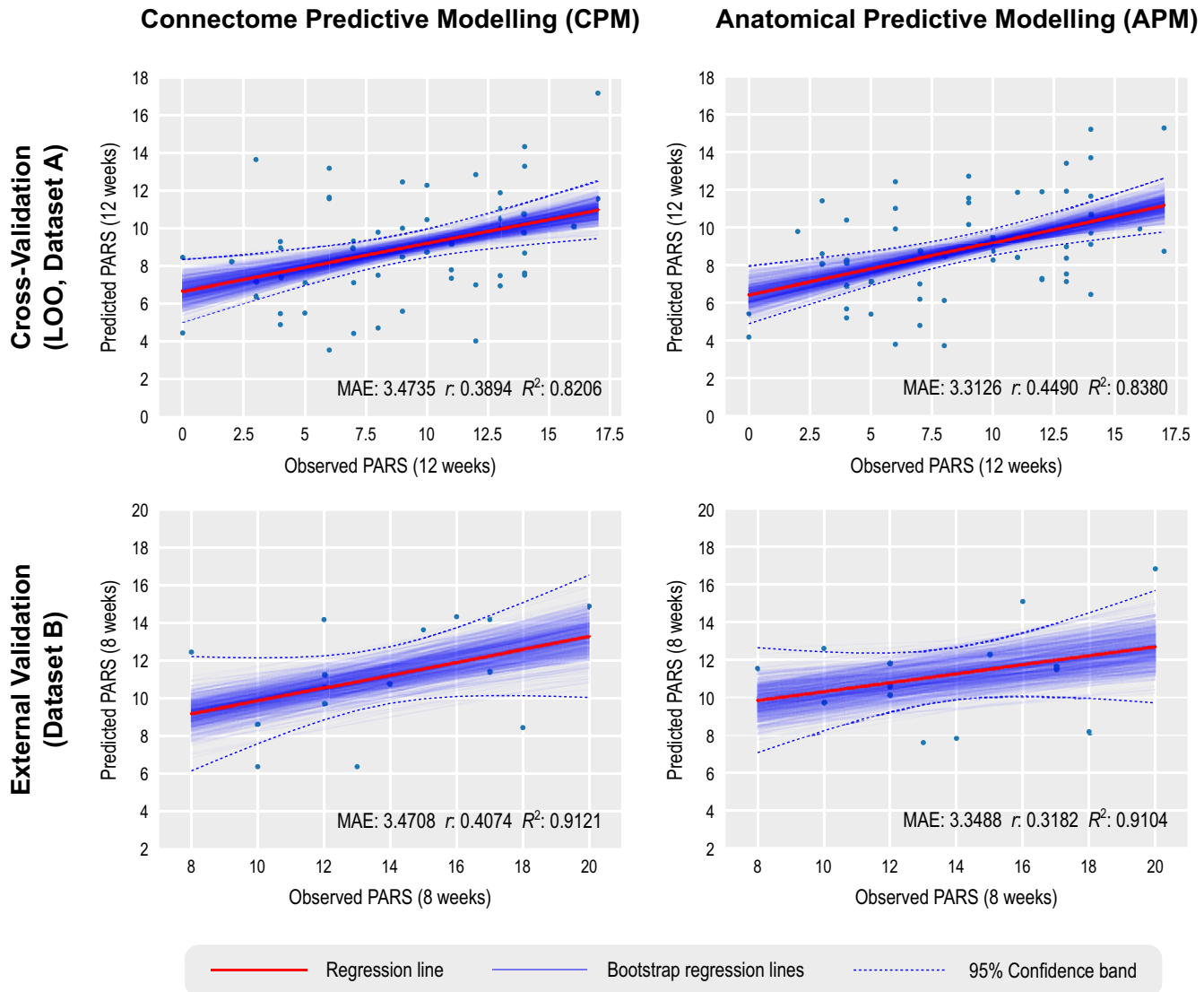
Prediction of anxiety scores using APM

For APM, the models that were generally better predictive were those that used the gray/white matter contrast, without smoothing, and that selected both positive and negative vertices in the first regression of APM. Moreover, this set of APM models also tended to produce higher correlations between predicted and observed PARS scores than CPM, with correlations above 0.4. However, as with CPM, the results show no model produced a strong R^2. A summary is presented in Table 4, and a scatter plot showing observed and predicted values for one of these models appears in Figure 1, upper right panel.

Unlike the cross-validation results, models for CPM generally produced stronger results for external validation than models for APM. Moreover, whereas results for CPM appeared generally comparable across cross-validation and external validation, for APM, external validation for Dataset B produced indices of

Table 3. Mean absolute error (MAE) of CPM-predicted vs. observed PARS at 12 weeks, using Dataset A for training and leave-one-out cross-validation and Dataset B for external validation. The corresponding correlation (r) and coefficient of determination (R^2) are also shown. Confidence intervals (95%), based on 1000 bootstraps, are between brackets. A scatter plot for the model marked with an asterisk (*) is in Figure 1 (left panels)

Predictors	Nuisance	MAE	r	R^2
Dataset A (leave-one-out cross-validation, prediction at 12 weeks):				
Positive edges	None	3.9116 [3.5795–4.2518]	0.0696 [(−0.1315)–0.2504]	−0.1551 [(−0.3319)–0.0054]
Positive edges, age, sex	None	3.4655 [3.0858–3.8348]	0.3621 [0.2133–0.508]	0.0757 [(−0.1449)–0.2700]
Positive edges	Age, Sex	3.5271 [3.0549–3.9932]	0.3283 [0.1499–0.4972]	0.0268 [(−0.2443)–0.2680]
Positive edges, age, sex*	Baseline PARS*	3.4735 [3.1024–3.8543]	0.3894 [0.2479–0.5117]	0.0824 [(−0.1428)–0.2617]
Positive edges	Baseline PARS, Age, Sex	3.5984 [3.1710–4.0292]	0.3836 [0.2250–0.5221]	0.0543 [(−0.2113)–0.2737]
Dataset B (external validation, prediction at 8 weeks):				
Positive edges	None	4.0386 [3.6969–4.5391]	0.3825 [0.0427–0.6226]	−1.1001 [(−2.0428)–(−0.3175)]
Positive edges, age, sex	None	3.8189 [3.3147–4.7429]	0.3689 [0.2186–0.5188]	−1.0566 [(−2.1579)–(−0.1737)]
Positive edges	Age, Sex	5.1804 [4.6118–6.2915]	0.2838 [0.1137–0.4696]	−2.2670 [(−3.3783)–(−1.2284)]
Positive edges, age, sex*	Baseline PARS*	3.4708 [2.8457–4.7787]	0.4074 [0.2460–0.5417]	−0.6533 [(−2.2945)–0.1641]
Positive edges	Baseline PARS, Age, Sex	4.2822 [3.6120–5.4734]	0.3662 [0.1813–0.5274]	−1.3648 [(−2.9771)–(−0.4045)]

Connectome Predictive Modelling (CPM)   **Anatomical Predictive Modelling (APM)**

Figure 1. Prediction of anxiety scores using CPM and APM; APM used gray/white matter contrast. The main regression line (red) is based on the observed and predicted values (represented by the dots). The bootstrap regression lines (faint blue) are based on the bootstrapped predictions used to construct the 95% confidence intervals given in Tables 3 and 4 (to avoid clutter, only 500 out of 1000 lines are shown in each panel). The 95% confidence bands were computed parametrically in relation to the main regression line and are presented merely as an additional reference. Observe that the slopes of the bootstrapped lines in the external validation are less variable, which is expected given the larger number of observations that are predicted in a single step (15 in this case) versus the single prediction in each step of the leave-one-out (LOO) cross-validation.

accuracy that were generally lower than for cross-validation. Figure 1, lower right panel shows the corresponding scatter plot for observed and predicted values for Dataset B. An extended set of results for APM with cortical thickness, surface area, curvature, and sulcal depth, with and without smoothing, are provided in the Supplementary Material.

Localization of predictive edges and vertices

For both CPM and APM, the predictive elements – edges or vertices, respectively – identified in the first regression were widely distributed throughout the brain. These topographies did not manifest patterns comparable to networks of known specific functions. We focus on the models that included age, sex, and baseline PARS as nuisance; these are highlighted with an asterisk (*) in Tables 3 and 4. Figure 2 shows the edges most frequently identified in the leave-one-out cross-validation using Dataset A
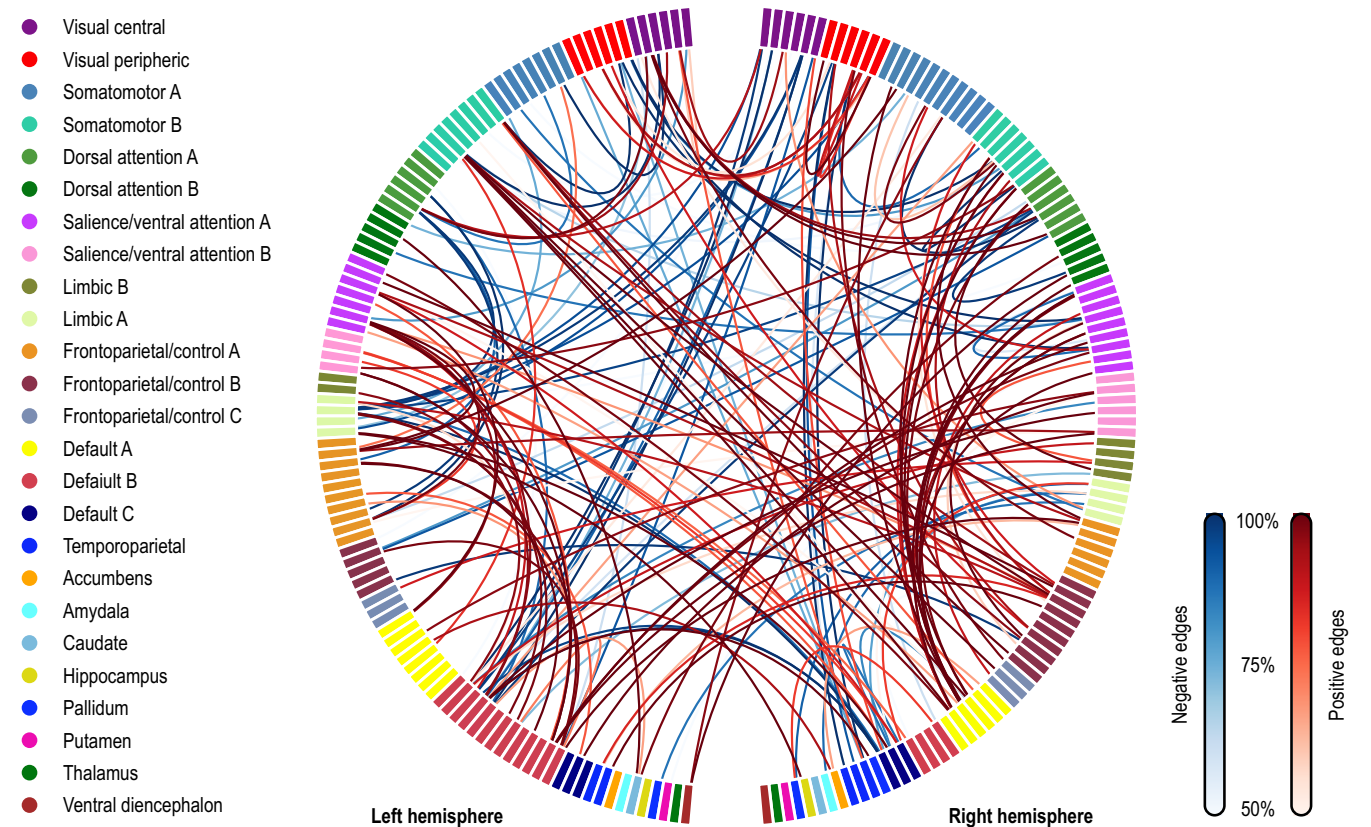
with CPM; Figure 3 provides a similar depiction for APM, using gray/white matter contrast. For both CPM and APM, the number of elements found as significant in the first stage of the respective predictive model was relatively small, about two orders of magnitude smaller than the number of edges or vertices available for a given model.

Uniqueness and its predictive value

Fingerprinting using rsFC features and sMRI features led to strong accuracy for subject identification. Using rsFC features from the baseline scan, the correct follow-up scan was correctly identified for 53 of 66 subjects (80.3%, $p = 6.2 \times 10^{-84}$), whereas doing the reverse produced correct identifications for 56 of 66 (84.9%, $p = 2.3 \times 10^{-91}$); these results are based on full (not partial) correlations and with GSR. Anatomical fingerprinting led to even higher rates of correct identification, with a near 100%

Table 4. Mean absolute error (MAE) of APM-predicted (with gray/white contrast) vs. observed PARS, using Dataset A for training and leave-one-out cross-validation and Dataset B for external validation. The corresponding correlation (*r*) and coefficient of determination (R^2) are also shown. Confidence intervals (95%), based on 1000 bootstraps, are between brackets. A scatter plot for the model marked with an asterisk (*) is in Figure 1 (right panels)

Predictors	Nuisance	MAE	*r*	R^2
Dataset A (leave-one-out cross-validation, prediction at 12 weeks):				
Both pos/neg vtx	None	3.3416 [3.0283–3.6602]	0.4074 [0.2052–0.5714]	0.1539 [(−0.0183)–0.3104]
Both pos/neg vtx, age, sex	None	3.2228 [2.8900–3.5997]	0.4242 [0.2846–0.5386]	0.1515 [(−0.0648)–0.3301]
Both pos/neg vtx	Age, sex	2.9331 [2.1735–3.7533]	0.5421 [0.3514–0.6926]	0.2823 [(−0.2671)–0.7636]
Both pos/neg vtx, age, sex*	Baseline PARS*	3.3126 [2.9656–3.6979]	0.4490 [0.3108–0.5619]	0.1714 [(−0.0462)–0.3692]
Both pos/neg vtx	Baseline PARS, age, sex	3.9326 [3.1805–4.7093]	0.4579 [0.2765–0.6103]	−0.1029 [(−0.6948)–0.3560]
Dataset B (external validation, prediction at 8 weeks):				
Both pos/neg vtx	None	3.3873 [2.9435–4.19478]	0.2376 [(−0.2450)–0.6356]	−0.7177 [(−1.9185)–0.2394]
Both pos/neg vtx, Age, Sex	None	4.5185 [3.8783– 5.7451]	0.1709 [0.0177–0.3664]	−1.7479 [(−3.0675)–(−0.6599)]
Both pos/neg vtx	Age, sex	5.1874 [4.5683–6.3311]	0.2283 [0.0544–0.4180]	−2.4419 [−3.6649–(−1.4652)]
Both pos/neg vtx, age, sex*	Baseline PARS*	3.3488 [2.6322–4.7357]	0.3182 [0.1532–0.4607]	−0.6847 [(−2.197)–0.2572]
Both pos/neg vtx	Baseline PARS, age, sex	4.2541 [3.5672–5.4248]	0.3517 [0.1782–0.5139]	−1.4368 [−2.8349–(−0.5182)]

Legend:
- Visual central
- Visual peripheric
- Somatomotor A
- Somatomotor B
- Dorsal attention A
- Dorsal attention B
- Salience/ventral attention A
- Salience/ventral attention B
- Limbic B
- Limbic A
- Frontoparietal/control A
- Frontoparietal/control B
- Frontoparietal/control C
- Default A
- Defaiult B
- Default C
- Temporoparietal
- Accumbens
- Amydala
- Caudate
- Hippocampus
- Pallidum
- Putamen
- Thalamus
- Ventral diencephalon

Left hemisphere — Right hemisphere

Negative edges / Positive edges — 100% / 75% / 50%

Figure 2. Edges most frequently identified as positively (red) or negatively (blue) associated with the PARS score at 12 weeks in Dataset A, as found in the first stage of CPM. The frequency refers to the number of iterations of the leave-one-out cross-validation in which a significant association was found; edges found in at least 50% of the iterations are shown (128 positive, 73 negative, out of 23,220 edges). The connections shown are for the model marked with an asterisk (*) in Table 3 (only the positive edges were used in the second stage of CPM; the negative edges are depicted for completeness). Named networks are those identified by Yeo et al. (2011); the set of nodes also includes 8 subcortical regions. Note that despite the seemingly large number of connections, only a small fraction of the total number of edges is used, in a pattern mostly diffuse and unstructured.

success rate for most of the measurements studied (cortical area, thickness, curvature, sulcal depth, and gray/white matter contrast). Differential power for edges and for the gray/white matter contrast are shown in Figure 4; Figure 5 shows DP for the other anatomical measurements. DP was found substantially higher for every anatomical measurement studied compared to connectivity measurements: while DP for edges was found generally weak and scattered, for gray/white contrast, DP was found stronger and
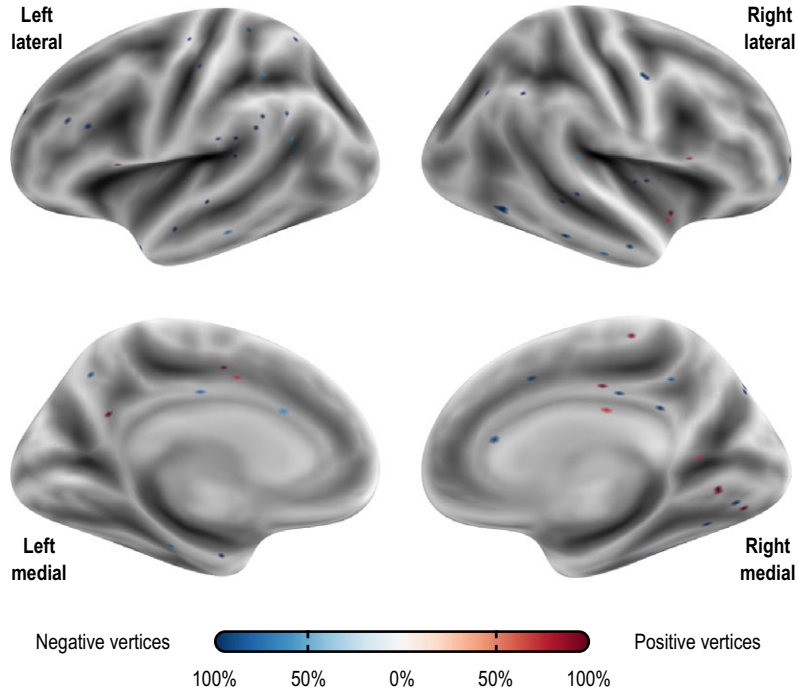
Figure 3. Vertices most frequently identified as positively (red) or negatively (blue) associated with the PARS score at 12 weeks in Dataset A. Observe that the pattern is mostly scattered, with isolated vertices (representing tiny regions) diffusely spread throughout the cortex. These results are as found in the first stage of APM using gray/white matter contrast. The percentage refers to the number of iterations of the leave-one-out cross-validation in which a significant association was found over all iterations; vertices found in at least 50% of the iterations are shown (16 positive and 55 negative, out of 18,742 vertices). The vertices shown are for the model marked with an asterisk (*) in Table 4.
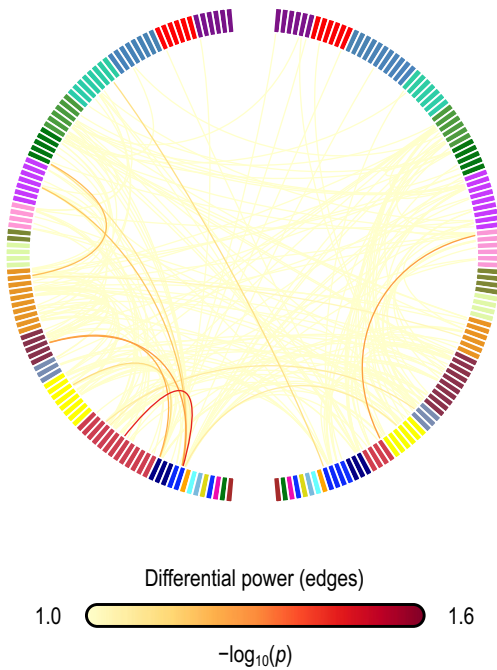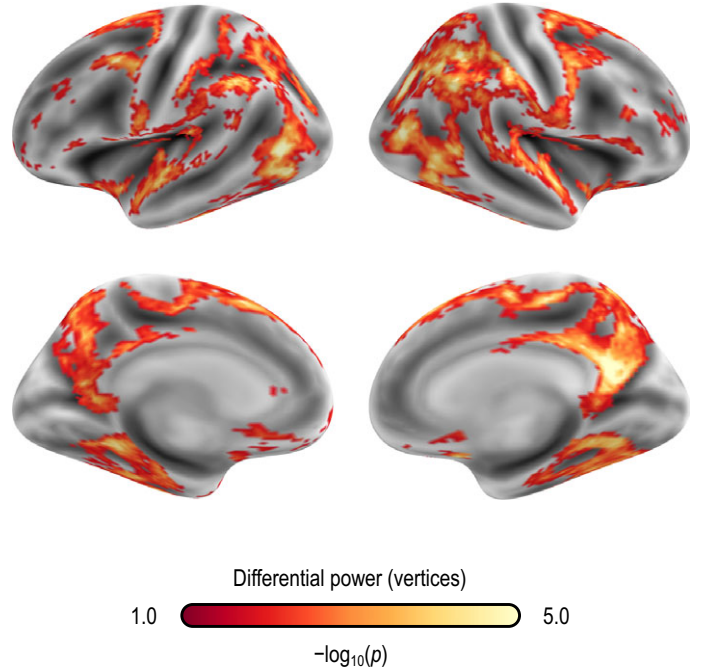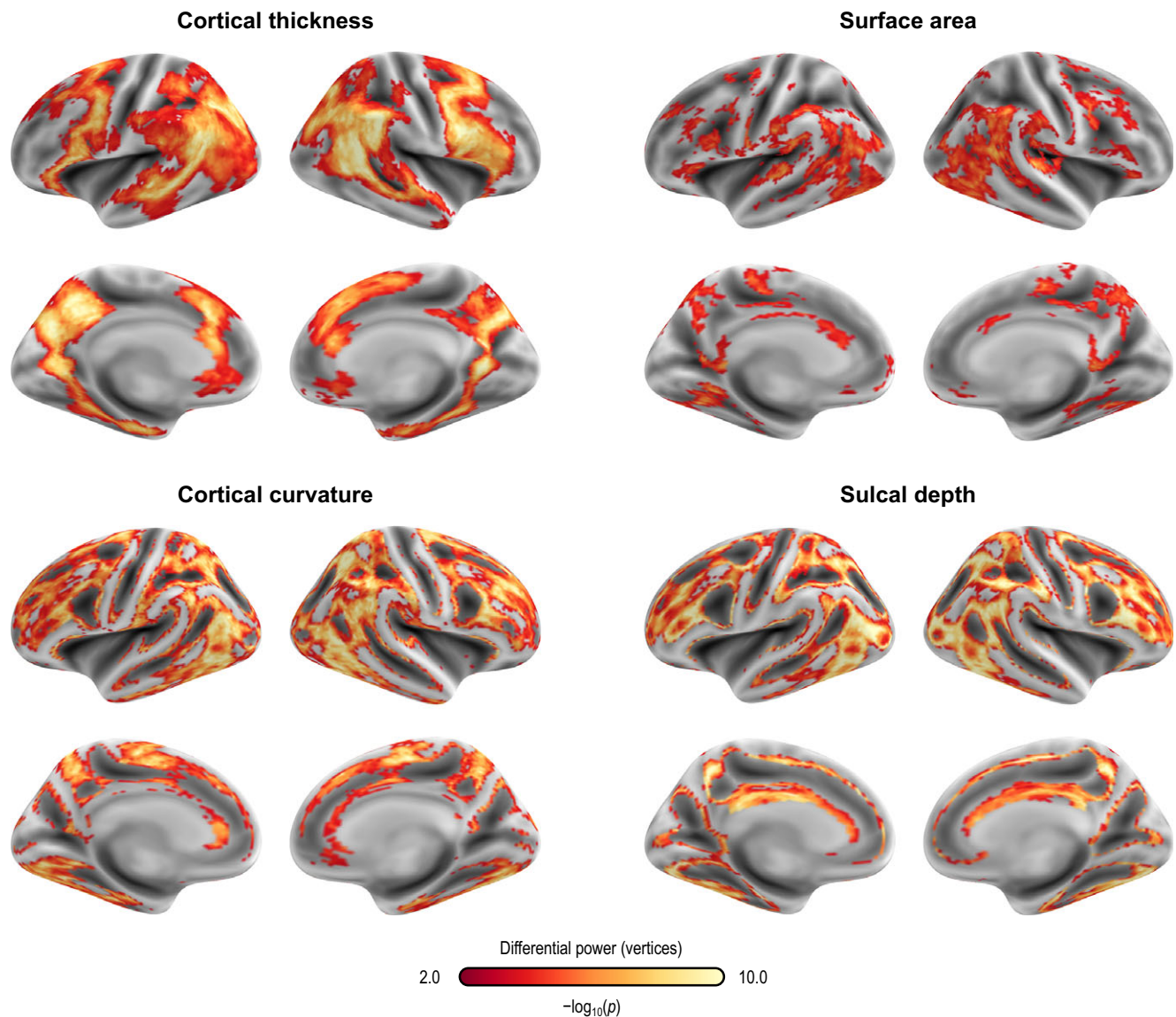

Figure 4. Differential power (DP) for edges using connectome fingerprinting (left), and for vertices using anatomical fingerprinting with gray/white matter contrast (right), in logarithmic scale based on their *p*-values (i.e. −log10(p), where p is the p-value for DP, thus allowing scales to be comparable). Network names for the left panel are the same as for Figure 2, and name views are the same as for Figure 3. While anatomical fingerprinting without smoothing was slightly more accurate, the smoothed version includes the same regions and is more informative; hence, it is the one shown. Higher values for the DP indicate features that are more unique. DP is much higher for anatomical measurements than for connectivity measurements (note the different color scales); DP for connectivity features (edges) is generally weak and scattered, whereas for gray/white contrast (vertexwise), DP is stronger and with better-defined localization. DP for cortical area, thickness, curvature, and sulcal depth are shown in Figure 5.

Cortical thickness

Surface area

Cortical curvature

Sulcal depth

Differential power (vertices)

2.0 — 10.0

$-\log_{10}(p)$

Figure 5. Differential power (DP) for vertices using anatomical fingerprinting with cortical thickness, cortical surface area, cortical curvature, and sulcal depth, in logarithmic scale based on their p-values (i.e. $-\log10(p)$, where p is the *p*-value for DP, thus allowing scales to be comparable). While anatomical fingerprinting without smoothing was slightly more accurate, the smoothed version includes the same regions and is more informative; hence, it is the one shown. Higher values for the differential power indicate features that are more unique. As with the gray/white contrast, DP is for other anatomical measurements.

with well-defined locations, covering mostly parietal cortex, precuneus, inferior temporal lobe, and caudal portions of the frontal lobe before reaching the precentral gyrus, and preserving central sulcus, pre- and postcentral gyri, insula, and cuneus, all of which are regions of known lower variability among individuals. The relation between DP in different modalities can be seen in Supplementary Figure 1. There is little relation between the DP between structural measurements and rsFC.

The edges or vertices with higher DP derived from fingerprinting, that is, those more 'unique', yielded slightly lower correlation to PARS scores compared to those found by model fitting in the first stage of CPM/APM. There was no overlap between the edges selected by using fingerprinting, compared to the edges found in the first stage of the CPM approach; the same was observed for APM.

Discussion

Prediction of response to treatment

This work applied CPM to predict response to CBT in pediatric anxiety disorders. The study used expert clinicians and a gold-standard measure of treatment outcome, in medication-free subjects recruited using criteria from past large-scale randomized controlled trials of pediatric anxiety disorders, that is, RUPP (Walkup et al., 2001) and CAMS (Walkup et al., 2008). Three main findings emerged. First, no model produced clearly significant results when using R^2. Second, sMRI outperformed rsFC for fingerprinting, where it achieved excellent accuracy. Finally, both CPM and APM had limitations; no single model emerged as consistently better than all other models, and prediction arose from hundreds of regions that did not cohere into networks identified in other studies.

An advantage of the CPM methodology over other predictive models concerns its capacity to generate interpretable findings that might prove useful in a clinical context. Nevertheless, the current findings suggest the need for improvements before clinically useful approaches can emerge. For example, the edges that drove successful prediction varied across cross-validation loops. Such patterns complicate attempts to identify one set of robustly predictive edges. Findings in this study also failed to reveal patterns closely overlapping with regions previously associated with anxiety. As in Linke et al. (2021), the current findings could reflect a 'many-to-one' pattern, where complex collections of connections in the brain interact to shape pediatric psychopathology.

The prediction of therapeutic response in this study went beyond a mere exercise of rating unseen data; it related functional connectivity to a gold-standard, clinically relevant outcome variable, over and above baseline levels of severity, as well as demographic factors such as age and sex. Models that used imaging data to predict the posttreatment PARS with baseline PARS as a nuisance resulted in higher quality models than without. This is somewhat to be expected as the subject baseline symptom level might be related to brain measures and treatment outcomes.

Prediction offers the potential to stratify subjects according to the likelihood that treatment is successful, to indicate those who may need additional support, as well as to use data to support mechanistic theories of psychopathology and their links to novel therapeutics, although some have warned caution (Mitchell, Potash, Barocas, D'Amour, & Lum, 2021). In effect, precision medicine and personalized clinical predictions have been garnering increased attention in recent years (Fair & Yeo, 2020; Laumann, Zorumski, & Dosenbach, 2023). However, a recent systematic review of 308 prediction models for psychiatry outcomes reported that 95% of studies were at high risk of bias primarily due to overfitting and biased variable selection methods; only 20% performed external validation on an independent sample (Meehan et al., 2022), highlighting the need for robust methodology and validation in clinical models.

Resting-state fMRI is frequently criticized as an inaccurate picture of what would be brain resting-state activity, given that rsFC has been shown to relate to numerous uncontrolled variables such as mood (Harrison et al., 2008) or alertness (Chang et al., 2016), albeit more consistent results can be found using rsFC in predictive models than with other analytical approaches (Taxali, Angstadt, Rutherford, & Sripada, 2021). In effect, analysis of task-based fMRI from the same trial shows baseline differences and a return to normality after CBT (Haller et al., 2024). Additional use of tasks meant to draw out individual differences in the measure of interest may provide additional predictive power in CPM and reduce confounding effects due to the lack of engagement during rest (Finn et al., 2017), and as demonstrated in recent research (Barron et al., 2021; Greene et al., 2018; Rosenberg et al., 2016). Using a general functional connectivity (GFC) measure based on multiple fMRI tasks may also have advantages over single-task CPM. Elliott et al. (2019) show increased test-retest reliability and higher heritability in GFC than rsFC. GFC may also improve prediction over single-task FC, both when computed using averaged connectomes or concatenation of time series (Gao et al., 2019).

Our findings that neither CPM nor APM reliably predicted PARS score after treatment illustrate the difficulty in applying potential imaging-based measurements to improve treatment outcome predictions in a clinical sample. Additional research is necessary to explore how integrating different imaging modalities might benefit predictive algorithms.

Individual uniqueness

As in a previous study, there was no overlap between the edges selected during CPM and the edges used for subject identification (Mantwill et al., 2022). In this study, edge selection from fingerprinting using the most discriminatory features led to comparable results to the CPM. The edges selected in each case possibly represent different sources of variability that are not related; both approaches to edge selection might contain relevant, yet distinct information (Finn & Rosenberg, 2021). However, in our study, no approach appeared particularly promising over the other.

Limitations

Our study presents a relatively small sample size. The limited number of subjects included in each dataset led us to perform leave-one-out cross-validation, which can yield unstable estimates of accuracy (Varoquaux et al., 2017). This is somewhat countered by the fact that we were able to benefit from another dataset with similar inclusion criteria and study design for validation.

A limiting applicability concerned model selection: we successfully found models that appeared to be successful when the correlation coefficient is considered, but that are in fact, predicting the target value worse than the mean. The results highlight the need to use more than one metric when assessing predictive models. Another limiting factor to the application of CPM in our sample is that the best models show an MAE of approximately 3.4 (PARS ranges between 0 and 25). Using predictive models in clinical practice is an emerging science. The added value that these models can bring to clinical practice remains uncertain and needs to be assessed objectively. In a study to detect the risk of mental health crisis stratified according to an automated model based on health registry, most clinical teams found the measure useful, although leading to relatively few additional actions (Garriga et al., 2022). Finally, there is a need to train and test such predictive algorithms in more diverse settings. Our sample was mostly comprised of white Americans with high family incomes. Future studies should test if predictive models are generalizable to populations of diverse ethnic and cultural backgrounds.

Conclusion

This study applied a predictive model approach to data from children and adolescents with anxiety. Limited success was obtained in the prediction of outcomes of CBT treatment. While some models showed moderate correlations between predicted and observed anxiety scores, the overall predictive power was weak, with low model performance. The methods used here have shown promising results when applied in other populations, including when predicting anxiety symptoms, and in line with that of other recent work (Hilbert et al., 2024), in which the authors fail to predict response to CBT in a population of adults with anxiety disorders using different machine learning methods based on rs-fMRI. We found mostly diffuse patterns of edges being selected for prediction, limiting the interpretation of the findings. Despite using expert clinicians, gold-standard outcome measures, and medication-free subjects, the models failed to consistently identify robust patterns associated with treatment response. This study, therefore, does not

support evidence for the use of CPM to predict treatment outcomes in pediatric anxiety.

References

Amico, E., & Goñi, J. (2018). The quest for identifiability in human functional connectomes. *Scientific Reports*, **8**(1), 8254. https://doi.org/10.1038/s41598-018-25089-1.

Ashar, Y. K., Clark, J., Gunning, F. M., Goldin, P., Gross, J. J., & Wager, T. D. (2021). Brain markers predicting response to cognitive-behavioral therapy for social anxiety disorder: An independent replication of Whitfield-Gabrieli et al. 2015. *Translational Psychiatry*, **11**, 260. https://doi.org/10.1038/s41398-021-01366-y.

Barron, D. S., Gao, S., Dadashkarimi, J., Greene, A. S., Spann, M. N., Noble, S., … Scheinost, D. (2021). Transdiagnostic, connectome-based prediction of memory constructs across psychiatric disorders. *Cerebral Cortex*, **31**(5), 2523–2533. https://doi.org/10.1093/cercor/bhaa371.

Byrge, L., & Kennedy, D. P. (2020). Accurate prediction of individual subject identity and task, but not autism diagnosis, from functional connectomes. *Human Brain Mapping*, **41**(9), 2249–2262. https://doi.org/10.1002/hbm.24943.

Chang, C., Leopold, D. A., Schölvinck, M. L., Mandelkow, H., Picchioni, D., Liu, X., … Duyn, J. H. (2016). Tracking brain arousal fluctuations with fMRI. *Proceedings of the National Academy of Sciences*, **113**(16), 4518–4523. https://doi.org/10.1073/pnas.1520613113.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, **7**, e623. https://doi.org/10.7717/peerj-cs.623.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, **20**(6), 425–443. https://doi.org/10.1016/j.tics.2016.03.014.

Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., … Hariri, A. R. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *NeuroImage*, **189**, 516–532. https://doi.org/10.1016/j.neuroimage.2019.01.068.

Fair, D. A., & Yeo, B. T. T. (2020). Precision neuroimaging opens a new chapter of neuroplasticity experimentation. *Neuron*, **107**(3), 401–403. https://doi.org/10.1016/j.neuron.2020.07.017.

Finn, E. S., & Rosenberg, M. D. (2021). Beyond fingerprinting: Choosing predictive connectomes over reliable connectomes. *NeuroImage*, **239**, 118254. https://doi.org/10.1016/j.neuroimage.2021.118254.

Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., & Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *NeuroImage*, **160**, 140–151. https://doi.org/10.1016/j.neuroimage.2017.03.064

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, **18**(11), 1664–1671. https://doi.org/10.1038/nn.4135.

Gao, S., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic measures. *NeuroImage*, **201**, 116038. https://doi.org/10.1016/j.neuroimage.2019.116038.

Garriga, R., Mas, J., Abraha, S., Nolan, J., Harrison, O., Tadros, G., & Matic, A. (2022). Machine learning model to predict mental health crises from electronic health records. *Nature Medicine*, **28**(6), 1240–1248. https://doi.org/10.1038/s41591-022-01811-5.

Ginsburg, G. S., Kendall, P. C., Sakolsky, D., Compton, S. N., Piacentini, J., Albano, A. M., … March, J. (2011). Remission after acute treatment in children and adolescents with anxiety disorders: Findings from the CAMS. *Journal of Consulting and Clinical Psychology*, **79**(6), 806–813. https://doi.org/10.1037/a0025933.

Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, **9**(1), 2807. https://doi.org/10.1038/s41467-018-04920-3.

Gregory, A. M., Caspi, A., Moffitt, T. E., Koenen, K., Eley, T. C., & Poulton, R. (2007). Juvenile mental health histories of adults with anxiety disorders. *The American Journal of Psychiatry*.

Haller, S. P., Linke, J. O., Grassie, H. L., Jones, E. L., Pagliaccio, D., Harrewijn, A., … Brotman, M. A. (2024). Normalization of fronto-parietal activation by cognitive-behavioral therapy in unmedicated pediatric patients with anxiety disorders. *American Journal of Psychiatry*, appi.ajp.20220449. https://doi.org/10.1176/appi.ajp.20220449.

Harrison, B. J., Pujol, J., López-Solà, M., Hernández-Ribas, R., Deus, J., Ortiz, H., … Cardoner, N. (2008). Consistency and functional specialization in the default mode brain network. *Proceedings of the National Academy of Sciences*, **105**(28), 9781–9786. https://doi.org/10.1073/pnas.0711791105.

Hedges, E. P., Dimitrov, M., Zahid, U., Brito Vega, B., Si, S., Dickson, H., … Kempton, M. J. (2022). Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. *NeuroImage*, **246**, 118751. https://doi.org/10.1016/j.neuroimage.2021.118751.

Hilbert, K., Böhnlein, J., Meinke, C., Chavanne, A. V., Langhammer, T., Stumpe, L., … Lueken, U. (2024). Lack of evidence for predictive utility from resting state fMRI data for individual exposure-based cognitive behavioral therapy outcomes: A machine learning study in two large multi-site samples in anxiety disorders. *NeuroImage*, **295**, 120639. https://doi.org/10.1016/j.neuroimage.2024.120639.

Kunas, S. L., Lautenbacher, L. M., Lueken, U., & Hilbert, K. (2021). Psychological predictors of cognitive-behavioral therapy outcomes for anxiety and depressive disorders in children and adolescents: A systematic review and meta-analysis. *Journal of Affective Disorders*, **278**, 614–626. https://doi.org/10.1016/j.jad.2020.09.092.

Kvålseth, T. O. (1985). Cautionary note about R^2. *The American Statistician*, **39**(4), 279–285. https://doi.org/10.1080/00031305.1985.10479448.

Laumann, T. O., Zorumski, C. F., & Dosenbach, N. U. F. (2023). Precision neuroimaging for localization-related psychiatry. *JAMA Psychiatry*, **80**(8), 763. https://doi.org/10.1001/jamapsychiatry.2023.1576.

Lebowitz, E. R., Marin, C., Martino, A., Shimshoni, Y., & Silverman, W. K. (2019). Parent-based treatment as efficacious as cognitive behavioral therapy for childhood anxiety: A randomized noninferiority study of supportive parenting for anxious childhood emotions. *Journal of the American Academy of Child and Adolescent Psychiatry*, S0890-8567(19)30173-X. https://doi.org/10.1016/j.jaac.2019.02.014.

Lin, Y., Baete, S. H., Wang, X., & Boada, F. E. (2020). Mapping brain–behavior networks using functional and structural connectome fingerprinting in the HCP dataset. *Brain and Behavior*, **10**(6). https://doi.org/10.1002/brb3.1647.

Linke, J. O., Abend, R., Kircanski, K., Clayton, M., Stavish, C., Benson, B. E., … Pine, D. S. (2021). Shared and anxiety-specific pediatric psychopathology dimensions manifest distributed neural correlates. *Biological Psychiatry*, **89**(6), 579–587. https://doi.org/10.1016/j.biopsych.2020.10.018.

Mansour, L. S., Tian, Y., Yeo, B. T. T., Cropley, V., & Zalesky, A. (2021). High-resolution connectomic fingerprints: Mapping neural identity and behavior. *NeuroImage*, **229**, 117695. https://doi.org/10.1016/j.neuroimage.2020.117695.

Mantwill, M., Gell, M., Krohn, S., & Finke, C. (2022). Brain connectivity fingerprinting and behavioural prediction rest on distinct functional systems of the human connectome. *Communications Biology*, **5**(1), 261. https://doi.org/10.1038/s42003-022-03185-3.

Meehan, A. J., Lewis, S. J., Fazel, S., Fusar-Poli, P., Steyerberg, E. W., Stahl, D., & Danese, A. (2022). Clinical prediction models in psychiatry: A systematic review of two decades of progress and challenges. *Molecular Psychiatry*, **27**(6), 2700–2708. https://doi.org/10.1038/s41380-022-01528-4.

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., … Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, **19**(11), 1523–1536. https://doi.org/10.1038/nn.4393.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, **8**(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., … Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron*, **77**(3), 586–595. https://doi.org/10.1016/j.neuron.2012.12.028.

Nelemans, S. A., Hale, W. W., Branje, S. J. T., Raaijmakers, Q. A. W., Frijns, T., Van Lier, P. A. C., & Meeus, W. H. J. (2014). Heterogeneity in development of adolescent anxiety disorder symptoms in an 8-year longitudinal community study. *Development and Psychopathology*, **26**(1), 181–202. https://doi.org/10.1017/S0954579413000503.

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, **203**, 116157. https://doi.org/10.1016/j.neuroimage.2019.116157.

Piacentini, J., Bennett, S., Compton, S. N., Kendall, P. C., Birmaher, B., Albano, A. M., … Walkup, J. (2014). 24- and 36-week outcomes for the child/adolescent anxiety multimodal study (CAMS). *Journal of the American Academy of Child & Adolescent Psychiatry*, **53**(3), 297–310. https://doi.org/10.1016/j.jaac.2013.11.010.

Pine, D. S., Cohen, P., Gurley, D., Brook, J., & Ma, Y. (1998). The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. *Archives of General Psychiatry*, **55**(1), 56. https://doi.org/10.1001/archpsyc.55.1.56.

Podell, J. L., Mychailyszyn, M., Edmunds, J., Puleo, C. M., & Kendall, P. C. (2010). The coping cat program for anxious youth: The fear plan comes to life. *Cognitive and Behavioral Practice*, **17**(2), 132–141. https://doi.org/10.1016/j.cbpra.2009.11.001.

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry*, **77**(5), 534. https://doi.org/10.1001/jamapsychiatry.2019.3671.

Rao, A., Monteiro, J. M., & Mourao-Miranda, J. (2017). Predictive modelling using neuroimaging data in the presence of confounds. *NeuroImage*, **150**, 23–49. https://doi.org/10.1016/j.neuroimage.2017.01.066

Ren, Z., Daker, R. J., Shi, L., Sun, J., Beaty, R. E., Wu, X., … Qiu, J. (2021). Connectome-based predictive modeling of creativity anxiety. *NeuroImage*, **225**, 117469. https://doi.org/10.1016/j.neuroimage.2020.117469.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, **42**(1), 59. https://doi.org/10.2307/2685263.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Constable, R. T., & Chun, M. M. (2017). Characterizing attention with predictive network models. *Trends in Cognitive Sciences*, **21**(4), 290–302. https://doi.org/10.1016/j.tics.2017.01.011.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, **19**(1), 165–171. https://doi.org/10.1038/nn.4179.

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, **12**(3), 506–518. https://doi.org/10.1038/nprot.2016.178.

Silverman, W. K., & Ginsburg, G. S. (1998). Anxiety Disorders. In T. H. Ollendick & M. Hersen (Eds.), *Handbook of child psychopathology* (pp. 239–268). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-5905-4_9.

Silverman, W. K., Pina, A. A., & Viswesvaran, C. (2008). Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, **37**(1), 105–130. https://doi.org/10.1080/15374410701817907.

Silverman, W. K., Rey, Y., Marin, C. E., Jaccard, J., & Pettit, J. W. (2022). Does training parents in reinforcement skills or relationship skills enhance individual youths' cognitive behavioral therapy for anxiety? Outcome, specificity, and mediation. *Clinical Psychological Science*, **10**(2), 355–373. https://doi.org/10.1177/21677026211016402.

Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage*, **184**, 741–760. https://doi.org/10.1016/j.neuroimage.2018.09.074.

Sporns, O. (2011). *Networks of the Brain*. Cambridge, MA, USA: MIT Press.

Taxali, A., Angstadt, M., Rutherford, S., & Sripada, C. (2021). Boost in test-retest reliability in resting state fMRI with predictive modeling. *Cerebral Cortex (New York, N.Y.: 1991)*. https://doi.org/10.1093/cercor/bhaa390.

The Research Units on Pediatric Psychopharmacology Anxiety Study Group. (2002). The Pediatric Anxiety Rating Scale (PARS): Development and psychometric properties. *Journal of the American Academy of Child & Adolescent Psychiatry*, **41**(9), 1061–1069. https://doi.org/10.1097/00004583-200209000-00006.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage*, **145**, 166–179. https://doi.org/10.1016/j.neuroimage.2016.10.038.

Walkup, J. T., Albano, A. M., Piacentini, J., Birmaher, B., Compton, S. N., Sherrill, J. T., … Kendall, P. C. (2008). Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *New England Journal of Medicine*, **359**(26), 2753–2766. https://doi.org/10.1056/NEJMoa0804633.

Walkup, J. T., Labellarte, M. J., Riddle, M. A., Pine, D. S., Greenhill, L., Klein, R., … Roper, M. (2001). Fluvoxamine for the treatment of anxiety disorders in children and adolescents. *New England Journal of Medicine*, **344**(17), 1279–1285. https://doi.org/10.1056/NEJM200104263441703.

Wang, Z., Goerlich, K. S., Ai, H., Aleman, A., Luo, Y., & Xu, P. (2021). Connectome-based predictive modeling of individual anxiety. *Cerebral Cortex*, **31**(6), 3006–3020. https://doi.org/10.1093/cercor/bhaa407.

White, L. K., Sequeira, S., Britton, J. C., Brotman, M. A., Gold, A. L., Berman, E., … Pine, D. S. (2017). Complementary features of attention bias modification therapy and cognitive-behavioral therapy in pediatric anxiety disorders. *American Journal of Psychiatry*, **174**(8), 775–784. https://doi.org/10.1176/appi.ajp.2017.16070847.

Whitfield-Gabrieli, S., Ghosh, S. S., Nieto-Castanon, A., Saygin, Z., Doehrmann, O., Chai, X. J., … Gabrieli, J. D. E. (2016). Brain connectomics predict response to treatment in social anxiety disorder. *Molecular Psychiatry*, **21**(5), 680–685. https://doi.org/10.1038/mp.2015.109.

Woodward, L. J., & Fergusson, D. M. (2001). Life course outcomes of young people with anxiety disorders in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, **40**(9), 1086–1093. https://doi.org/10.1097/00004583-200109000-00018.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, **106**(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011.

Zugman, A., Jett, L., Antonacci, C., Winkler, A. M., & Pine, D. S. (2023). A systematic review and meta-analysis of resting-state fMRI in anxiety disorders: Need for data sharing to move the field forward. *Journal of Anxiety Disorders*, **99**, 102773. https://doi.org/10.1016/j.janxdis.2023.102773.