RESEARCH

Group-level comparison of brain connectivity networks

Fatemeh Pourmotahari[1], Hassan Doosti[2], Nasrin Borumandnia[3], Seyyed Mohammad Tabatabaei[4] and Hamid Alavi Majd[1*]

Abstract

Background: Functional connectivity (FC) studies are often performed to discern different patterns of brain connectivity networks between healthy and patient groups. Since many neuropsychiatric disorders are related to the change in these patterns, accurate modelling of FC data can provide useful information about disease pathologies. However, analysing functional connectivity data faces several challenges, including the correlations of the connectivity edges associated with network topological characteristics, the large number of parameters in the covariance matrix, and taking into account the heterogeneity across subjects.

Methods: This study provides a new statistical approach to compare the FC networks between subgroups that consider the network topological structure of brain regions and subject heterogeneity.

Results: The power based on the heterogeneity structure of identity scaled in a sample size of 25 exhibited values greater than 0.90 without influencing the degree of correlation, heterogeneity, and the number of regions. This index had values above 0.80 in the small sample size and high correlation. In most scenarios, the type I error was close to 0.05. Moreover, the application of this model on real data related to autism was also investigated, which indicated no significant difference in FC networks between healthy and patient individuals.

Conclusions: The results from simulation data indicated that the proposed model has high power and near-nominal type I error rates in most scenarios.

Keywords: Connectivity, Subject heterogeneity, fMRI, Statistical power, Type I error rate

Background

The non-invasive method of functional magnetic resonance imaging (fMRI) identifies the changes of functional connectivity (FC) patterns between healthy and patient individuals by measuring blood-oxygen-level-dependent (BOLD) [1]. Studies indicated that altered topological patterns of brain connectivity are related to many neurological disorders, including autism, Parkinson's disease, and Alzheimer's disease. Therefore, functional connectivity analysis is performed to investigate the connectivity patterns among the brain regions of patients to determine the biomarkers of neurodegenerative disorders [2–4].

Group comparison of brain connectivity patterns is usually performed using two graph theory-based techniques. In these methods, the network is represented by a graph so that the brain regions are defined as nodes, and the correlation between them is defined as the edge [5–7]. The first method is based on comparing the edges individually. In this approach, the correlation coefficient across time series of all-region pairs is calculated to estimate the edges, then the hypothesis of equal FC patterns between healthy and patient groups is assessed using a

*Correspondence: alavimajd@gmail.com

[1] Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
Full list of author information is available at the end of the article

group-level statistic. Since there are a large number of pairwise edge comparisons, multiple testing procedures such as false discovery rate (FDR) are required to control false positive noise [8, 9]. In the second method, the comparison of brain connectivity is based on the network of edges. Some models investigate the changes in the FC patterns as a subnetwork of the edges, and others investigate them as an entire network of the edges. However, fitting a proper model to determine differentially expressed FC characteristics between subgroups faces some problems due to the high dimension of connectivity data and their complex correlation structure [10–12].

Generally, the fit of the FC model based on the edge requires estimating the covariance matrix among the edges. Since the structure of dependency between the edges is related to the network topological structure of brain regions, considering this feature can accurately estimate the model parameters. In conventional statistical models, the structure of parameters dependency is often determined by the distance between brain regions. For example, spatial dependencies are defined based on the distance of voxels that belong to the same region of interest (ROI) [13–16].

However, because the dependency among the edges is defined based on the network topological structure and is not necessarily limited to spatial adjacency, the results of these models may not be appropriate. On the other hand, because of the high-dimensional FC data, it is difficult to estimate a large number of parameters in the sample covariance matrix between edges, especially if the number of regions in the brain is large. In this regard, there are several optimization techniques for estimating parameters when the number of variables is more than the sample size [17–21]; however, these methods may overlook the network topological features to control false positive and false negative noises [22, 23].

Another feature of FC data is the heterogeneity, which indicates the existence of differences in the connectivity structure of brain regions across subjects within the same group. By considering the heterogeneity in the model, the power of tests can be increased to diagnose neurodegenerative disorders associated with brain connectivity correctly [24]. The heterogeneity across subjects can indicate the changes in the performance of cognitive and behavioural domains in healthy individuals, as well as the severity of symptoms and different responses to clinical interventions in patients [25–28]. In this regard, using the penalized model-based clustering, DiLernia et al. (2020) reported distinct FC patterns for each individual in the healthy and patient [29].

Many studies have been conducted on FC modelling, which has considered some of the FC data properties. For example, Fiecas et al. (2017) accounted the heterogeneity across subjects by introducing a variance component model. However, in addition to ignoring the spatial features of the brain network, their proposed model cannot estimate the parameters of the edges when the number of ROIs is large. Furthermore, given the large number of ROIs in many brain atlases, applying this model for more than 20 ROIs faces computational challenges [24]. On the other hand, Chen et al. (2020) proposed a nonparametric Bayesian model to estimate the massive parameters of the covariance matrix among edges, which takes the spatial network feature in terms of its topological structure. However, the effect of between-subject heterogeneity is not considered in this model, and equalization of the whole FC network is not examined [23]. Given these challenges, the present study proposes a more comprehensive model for examining differentially expressed FC characteristics between subgroups by considering the network topological structure and the heterogeneity across subjects.

Materials and methods
Estimating covariance matrix between edges

Connectivity data for each subject ($n = 1, ..., N$) is formulated by a matrix $M_{v \times v}^n = \left\{ M_{i,j}^n \right\}$ so that the elements of $M_{i,j}^n$ represents the Fisher's Z-transformed correlations between brain regions i and j. The entire brain connectivity network can be displayed by a graph $M^n = \{V, E\}$, where V is the set of nodes and $E = V(1 - V)/2$ is the set of edges. In this setting, the degree is determined by the number of edges, and the correlation coefficient is determined by the edges. The corrections have been made. The matrix M^n is transformed into a vector $Y_{1 \times E}^n$. Suppose that the vector $\mathbf{Y}$ follows a normal multivariate distribution $Y_{1 \times E}^n \sim MVN\left(X_n^T \beta_{p \times E}, \sum_{E \times E}\right)$, which X_n^T indicates the design matrix with p covariates, $\beta_{p \times E}$ is the effect of covariates on the vector $\mathbf{Y}$, and $\sum_{E \times E}$ is the covariance matrix. The model has the following form:

$$Y_{N \times E} = X_{N \times p}^T \hat{\beta}_{p \times E} + R_{N \times E}^o \qquad (1)$$

where $R_{N \times E}^o$ is the residual matrix. In the nonparametric Bayesian model, $R_{N \times E}^o$ is used as input data to estimate the covariance matrix by considering the characteristics of the network topological structure.

Let $R_{1 \times E}^n | \Lambda_{E \times E} \sim MVN(\mathbf{0}, \Lambda_{E \times E})$. The edges correlation matrix $\Lambda_{E \times E} = f(G, \boldsymbol{\rho})$ is a function of the unknown network structure G and the correlation factors $\boldsymbol{\rho} = (\rho_0, \rho_1, ..., \rho_k)$. G is a probability measurement of the latent K networks that follows a Dirichlet process, with parameters G_o and α:

$$G \sim DP(\alpha, G_o) \qquad (2)$$

The correlation matrix based on the network topological structure is given by the equation:

$$\Lambda_{e_{i,j} \times e_{i',j'}} = \begin{cases} \rho_k & if \quad \omega_i = \omega_j = \omega_i' = \omega_j' = C_k \\ \rho_0 & otherwise \end{cases} \quad (3)$$

$\Lambda_{e_{i,j} \times e_{i',j'}}$ is indeed an element of the matrix $\Lambda_{E \times E}$ which is determined by the correlation between the edges $e_{i,j}$ and $e_{i',j'}$. Let $e_{i,j}$ denote connectivity between regions i and j, $i \neq j$. Based on whether region i belongs to the cluster k or not, the term $\omega_i = C_k$ is used as an indicator variable. If two edges are in the same cluster, it can be supposed that:

$$e_{i,j} \cong e_{i',j'} \in C_k \left(e_{i,j} \in C_k, e_{i',j'} \in C_k\right) \quad if \ and \ only \ if \quad \omega_i = \omega_j = \omega_{i'} = \omega_{j'} = C_k$$

The nonparametric Bayesian model is used to detect neighbourhood networks based on the input data and correlation structure in Eq. (3).

A discrete distribution is employed to assign each region in the K networks with probability $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$.

The Dirichlet process in Eq. (2) is therefore equivalent to the following equations:

$$\omega_i = C_k \mid \boldsymbol{\pi} \sim Discrete(\boldsymbol{\pi}), i = 1, \ldots, V$$

$$\boldsymbol{\pi} \mid \alpha \sim Dirichlet\left(^{\alpha}/_K, \ldots, ^{\alpha}/_K\right), k \to \infty$$

The distributions of ρ_o and ρ_k are assumed to be normal with parameters μ_o, μ_k, τ_o^2 and τ_k^2:

$$\rho_o \mid \mu_o, \tau_o^2 \sim N\left(\mu_o, \tau_o^2\right)$$

$$\rho_k \mid \mu_k, \tau_k^2 \sim N\left(\mu_k, \tau_k^2\right), k = 1, \ldots, K$$

The posterior probability of $R_{N \times E}$ is calculated by the prior distribution ρ, ω and function f:

The number of nodes within the network k is showed by $m_{-ik} = \sum_{j \neq i} I(\omega_j = C_k)$.

$$p\left(\omega_i \neq \omega_j \text{ for all } j \neq i \mid \boldsymbol{\omega}_{-i}, \boldsymbol{\rho}, R_{N \times E}\right)$$

$$\propto \exp\{-N/2(\log\left(\det\left(f\left((\boldsymbol{\omega}_{-1}, \omega_i = C_{k+1}), \boldsymbol{\rho}\right)\right)\right)$$

$$+ tr\left(Hf\left((\boldsymbol{\omega}_{-1}, \omega_i = C_{k+1}), \boldsymbol{\rho}\right)^{-1}\right))\} \times \frac{\alpha}{v - 1 + \alpha}$$

The Sherman–Morrison formula is used in the computation of $(\Lambda_{E \times E})^{-1}$ [30]:

$$f(\boldsymbol{\omega}, \boldsymbol{\rho})^{-1} = \left(\Lambda_{E \times E}\right)^{-1} = \left(A + \sqrt{\rho_o}\mathbf{1}_{E \times 1}\left(\sqrt{\rho_o}\mathbf{1}_{E \times 1}\right)^T\right)^{-1}$$

$$= A^{-1} - \frac{\rho_o A^{-1} \mathbf{1}_{E \times E} A^{-1}}{1 + \rho_o \mathbf{1}_{E \times 1}^T A^{-1} \mathbf{1}_{E \times 1}}$$

$$p\left(\rho_o \mid \boldsymbol{\omega}, H, \boldsymbol{\rho}_{-0}\right)$$

$$\propto \exp\left\{-N/2\left(\log\left(\det\left(f(\boldsymbol{\omega}, \rho_o, \boldsymbol{\rho}_{-0})\right)\right) + tr\left(Hf(\boldsymbol{\omega}, \rho_o, \boldsymbol{\rho}_{-0})^{-1}\right)\right)\right\} \times p(\rho_o \mid \mu_o, \tau_o^2)$$

$$p\left(\rho_k \mid \boldsymbol{\omega}, H, \boldsymbol{\rho}_{-k}\right)$$

$$\propto \exp\left\{-N/2\left(\log\left(\det\left(f(\boldsymbol{\omega}, \rho_k, \boldsymbol{\rho}_{-k})\right)\right) + tr\left(Hf(\boldsymbol{\omega}, \rho_o, \boldsymbol{\rho}_{-k})^{-1}\right)\right)\right\} \times p(\rho_k \mid \mu_k, \tau_k^2)$$

Finally, given the posterior distributions of $\boldsymbol{\omega}$ and $\boldsymbol{\rho}$, the covariance matrix between the edges is estimated. A detailed description of the theoretical methods of the

$$p\left(R_{N \times E} \mid \Lambda_{E \times E}, G, \boldsymbol{\rho}\right) p(G) p(\boldsymbol{\rho}) \propto \left|2\pi \Lambda_{E \times E}\right|^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \sum_s R_n^T \left(\Lambda_{E \times E}\right)^{-1} R_n\right) p(G) p(\boldsymbol{\rho})$$

$$= \exp\left\{-^N/_2 \left(\log\left|\Lambda_{E \times E}\right|^{-1} + tr\left(H\left(\Lambda_{E \times E}\right)^{-1}\right)\right)\right\} p(G) p(\boldsymbol{\rho})$$

Where $H^o = (R^o)^T R^o/(N-p)$ and $H = (Diag(H^o))^{-1/2} H^o (Diag(H^o))^{-1/2}$.

Then the sampling of the conditional posterior probabilities $\boldsymbol{\omega}$ and $\boldsymbol{\rho}$ is obtained by the Markov chain Monte Carlo (MCMC) algorithm:

$$p(\omega_i = C_k \mid \boldsymbol{\omega}_{-i}, \boldsymbol{\rho}, R_{N \times E})$$

Bayesian nonparametric model is available in the original paper [23].

Estimating between-subject variability

This section describes the estimation of the covariance matrix $\boldsymbol{\Psi}$ to control subject heterogeneity and the estimation of edges parameters $\boldsymbol{\beta}$ by considering the

$$\propto \exp\{-N/2(\log\left(\det\left(f\left((\boldsymbol{\omega}_{-1}, \omega_i = C_k), \boldsymbol{\rho}\right)\right)\right)) + tr\left(Hf((\boldsymbol{\omega}_{-1}, \omega_i = C_k), \boldsymbol{\rho})^{-1}\right))\} \times \frac{m_{-ik}}{v - 1 + \alpha}$$

dependence structure between edges $\sum_{E \times E}$. The terms $\psi_1, ..., \psi_N$ are considered to apply between-subject variability. It is assumed that $\psi_N(1, ..., N)$ is the scaled diagonal identity matrix with $E \times E$ dimensions. The Ψ covariance is a block-diagonal matrix with $\psi_1, ..., \psi_N$ matrices along its diagonal. Each entry of ψ_N indicates the variability that can be assigned to subject sampling.

Estimation of both β and Ψ parameters are obtained using the iterative method. In this process, an initial value of $\hat{\Psi}$ is replaced in Eq. (4), and then the residuals $R = Y - X\beta$ are calculated.

$$\hat{\beta} = \left(X' \left(\hat{\Sigma}_{E \times E} + \hat{\Psi}_{E \times E} \right)^{-1} X \right)^{-1} X' \left(\hat{\Sigma}_{E \times E} + \hat{\Psi}_{E \times E} \right)^{-1} Y \tag{4}$$

The estimate $\hat{\Psi}$ to be updated using R. The process is repeated until convergence is achieved. To estimate the parameter Ψ, the outer product of the $E \times 1$ residual vector is obtained for each subject, then averaged over these vectors to create a residual covariance matrix $\hat{\Omega}_{E \times E}$. For ψ_N, the structures of a scaled identity matrix and compound symmetry with the forms $\psi_N = \sigma^2 I$ and $\psi_N = \sigma^2 + b$ are assumed. Where σ^2 is set as the average of all diagonal elements of $\hat{\Omega}_{E \times E} - \hat{\Sigma}_{E \times E}$, I is the identity matrix with dimensions of $E \times E$ and b is considered as the average of the off-diagonal elements of $\hat{\Omega}_{E \times E} - \hat{\Sigma}_{E \times E}$.

Group-level statistic to detect differentially expressed FC patterns.

The basic hypotheses in this study are examined using the following test statistics.

Hypothesis 1: The entire functional connectivity networks are significantly different between case and control groups. The test statistic is:

$$\left(C(\hat{\beta}_1 - \hat{\beta}_2) \right)' \left(C(\text{vâr}(\hat{\beta}_1) + \text{vâr}(\hat{\beta}_2)) C' \right)^{-1} \left(C(\hat{\beta}_1 - \hat{\beta}_2) \right) \tag{5}$$

Where the contrast matrix of C is the $E \times E$ identity matrix. The parameters β_1 and β_2 are estimated according to Eq. (4). The covariance matrix for each group is estimated $\text{vâr}\left(\hat{\beta} \right) = \left(X' \left(\hat{\Sigma}_{E \times E} + \hat{\Psi}_{E \times E} \right)^{-1} X \right)^{-1}$.

Hypothesis 2: The functional connectivity of the ROI pairwise is significantly different between the case and the control groups. The following test statistic is:

algorithm of section. The values of the statistics for each permutation are then derived by replacing them with Eq. (5) and (6). In addition, the FDR method was used to adjust the *p*-value of the permutation test since there was the high number of multiple comparisons in the second hypothesis. In both the simulation and real data assessments, the nominal level of statistical significance was set at $\alpha = 0.05$. Figure 1 shows a process of the proposed method to determine differentially expressed FC patterns between two groups.

Simulation settings

In this section, simulation data were applied to assess the performance of the proposed method in terms of statistical power and type I error. The number of nodes $V = 20$, 25, and 30 with two latent clusters were considered to simulate the dependence structure between the edges. The correlation of the edges inside the cluster was $\rho_1 = \rho_2$ equal to the values 0.3, 0.5, and 0.7 as well as the correlation of the edges outside the cluster was $\rho_0 = 0$. In the next step, the *Uniform*$(-\delta, \delta)$ distribution was employed to create subject-specific effects. The values for each subject were separately generated from this distribution and were added to the diagonal entries of the edges covariance matrix. Given that the δ parameter controls the degree of heterogeneity between subjects, $\delta = 0.15$ and 0.3 were considered as low and high heterogeneity, respectively.

In the following, connectivity data of each group is generated using a multivariate normal distribution according to the features of the covariance matrix. For example, Fisher's Z-transformed correlation is considered a connectivity metric.

$$Y_{N \times E} \sim \begin{cases} MVN\left(0, \sum_{E \times E} + \Psi \right) + d & \text{Control group} \\ MVN\left(0, \sum_{E \times E} + \Psi \right) & \text{Case group} \end{cases}$$

The d vector reflects the differentially expressed FC network between groups so that an $E \times 1$ zero vector is considered, and then $d = 0.8$ is randomly replaced for about 5% of zero values.

The results of simulation data were based on 100 iterations for each setting to estimate power and type I errors under the null hypothesis. For all tests, the permutation number

$$\left(\left(\hat{\beta}_1(e) - \hat{\beta}_2(e) \right) \right)' \left(\left(\text{vâr}\left(\hat{\beta}_1(e) \right) + \text{vâr}\left(\hat{\beta}_2(e) \right) \right) \right)^{-1} \left(\left(\hat{\beta}_1(e) - \hat{\beta}_2(e) \right) \right), e = 1, .., E \tag{6}$$

A permutation test has been used to check the hypotheses while controlling type I error. In this regard, with each resampling of subjects between two groups, the interest parameters are estimated using the iterative

was 500. The model performance was evaluated based on the amount of correlation between edges, heterogeneity, and the dimensions of communication data in the sample size of 10 and 25. The proposed methods were implemented with R

Fig. 1 Process of the proposed approach to determine differentially expressed FC patterns between case and control groups
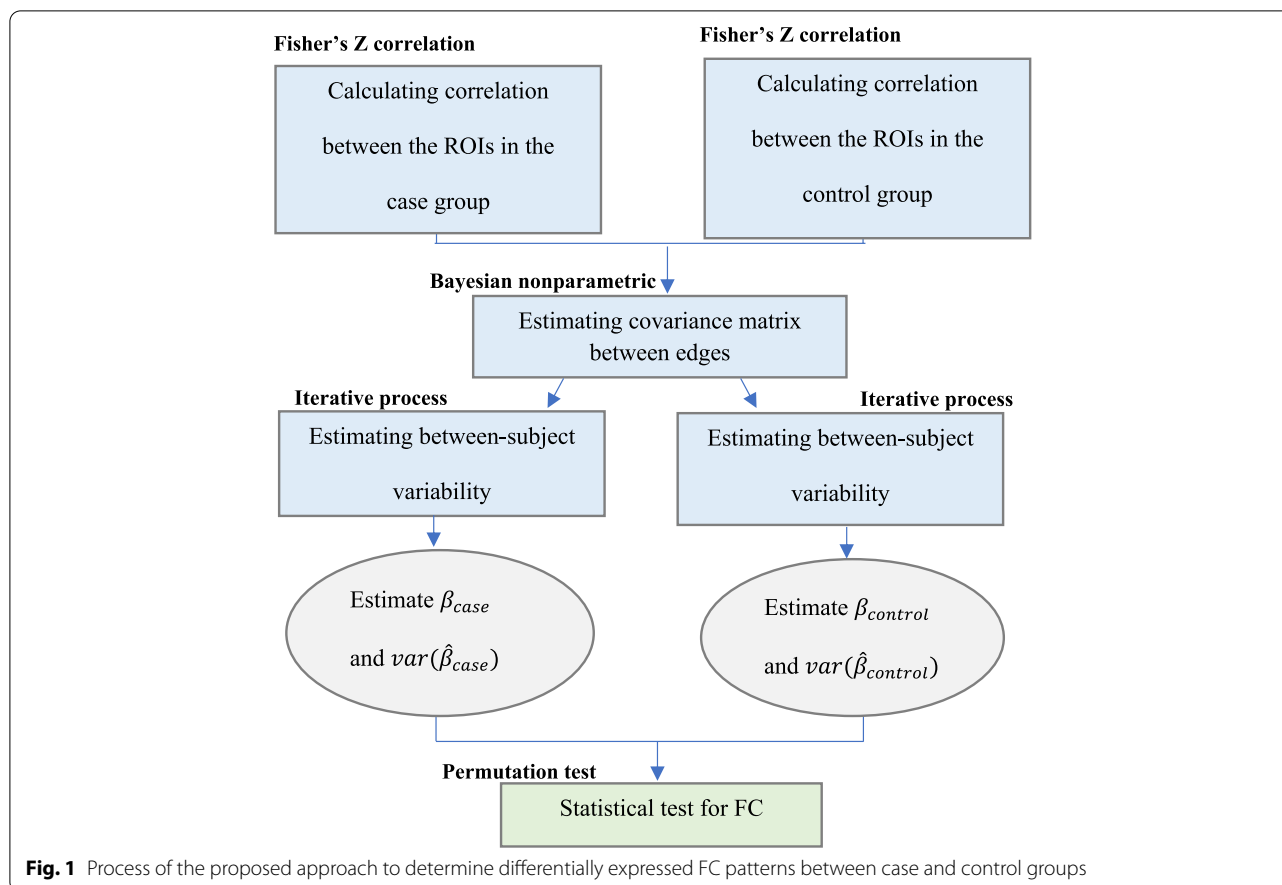
software version 4.0.5 and MATLAB R2021b. The codes of the method are available upon request.

Application to autism data
Subjects
This study examined resting-state fMRI data of 25 men with autism)7.29–15.66 years old) and 25 healthy men (7.26–14.66 years old) with a full intelligence quotient (FIQ) score above 87. The data were matched on age ($p = 0.272$) and FIQ score ($p = 0.659$).

Data acquisition
The resting-state fMRI data of subjects were obtained from the ABIDE database gathered by NYU Langone Medical Center [31]. fMRI images were collected using a 3 T Siemens Allegra scanner for 6 minutes. Participants were asked to relax and stare at the white cross displayed in the middle of a screen with a black background. The following is the scan acquiring procedure: TR = 2000 ms, TE = 15 ms, 33 slices with thickness = 4.0 mm, FOV = 240 mm, flip angle = 90.

Data processing
The pre-processed data were selected from the ABIDE dataset at http://fcon_1000.projects.nitrc.org/indi/abide/

[32]. Pre-processing of fMRI scans was performed using SPM8 software. The first ten values of each time series were eliminated due to the correction of the initial scan heterogeneity and the adaptability of individuals to the surrounding states; hence, a total number of 170 values per individual was considered. Normalization of scans was performed to the MNI (Montreal Neurological Institute) space with a resolution of $3 \times 3 \times 3 \, mm^3$ and correction of head movement according to Friston 24-parameter model [33] .According to the AAl (automated anatomical labelling) atlas, the pre-processed data were partitioned into 116 regions. [34]. The number of regions $V = 30$ affected by autism, including regions in the default mode network (DMN), were selected to compare data characteristics with simulation scenarios [35, 36].

Results
Simulation data
Table 1 displays the results of the simulation scenarios with $\rho = 0.3$. Based on the identity scaled structure, the power was above 0.90 in sample size $N = 25$, regardless of the heterogeneity amount and the number of regions. In the number of regions above 20 and the sample size of

Table 1 Type I error and power of the model to check differentially expressed FC patterns between two groups in term of simulation setting: sample size N, structure of between-subject variability, number of regions *V*, degree of heterogeneity and dependence among the edges $\rho = 0.3$

p=0.3			Compound Symmetry		Scaled	
			Heterogeneity		Heterogeneity	
	V	N	Low	High	Low	High
Type I Error	20	10	0.090	0.060	0.090	0.090
		25	0.050	0.050	0.030	0.050
	25	10	0.020	0.000	0.050	0.050
		25	0.050	0.010	0.040	0.040
	30	10	0.060	0.060	0.050	0.050
		25	0.070	0.060	0.070	0.070
Power	20	10	0.290	0.280	0.510	0.510
		25	0.550	0.500	0.980	0.960
	25	10	0.470	0.480	0.610	0.590
		25	0.920	0.890	1.000	1.000
	30	10	0.650	0.650	0.760	0.750
		25	0.990	0.960	1.000	1.000

10 and 25, the type I error rates formed on identities scaled and compound symmetry structures were close to 0.05. Based on the compound symmetry structure, the model had acceptable power values without affecting the degree of heterogeneity in the sample size $N=25$ and $V>20$.

Table 2 shows the results of the simulation scenarios with $\rho = 0.5$. In a sample size N=25, the power based on both structures of between-subject variability was greater than 0.8, regardless of the number of regions and heterogeneity. The model demonstrated acceptable values of

power in the identity scaled structure when $N=10$ and $V=30$. The type I error was close to 0.05 in most scenarios for the region numbers of 20 and 25.

Table 3 shows the results of the simulation scenarios with $\rho = 0.7$. Based on the identity scaled structure, the power was above 0.80, regardless of the number of regions, heterogeneity, and sample size. However, according to the compound symmetry structure, the model had lower power in most scenarios. In more than half of the simulation items, the type I error was less than or equal to 0.05.

Table 2 Type I error and power of the model to check differentially expressed FC patterns between two groups in term of simulation setting: sample size N, structure of between-subject variability, number of regions *V*, degree of heterogeneity and dependence among the edges $\rho = 0.5$

p=0.5			Compound Symmetry		Scaled	
			Heterogeneity		Heterogeneity	
	V	N	Low	High	Low	High
Type I Error	20	10	0.050	0.020	0.050	0.070
		25	0.070	0.050	0.040	0.040
	25	10	0.030	0.030	0.050	0.050
		25	0.020	0.030	0.040	0.040
	30	10	0.050	0.060	0.050	0.050
		25	0.100	0.090	0.070	0.070
Power	20	10	0.340	0.340	0.660	0.650
		25	0.870	0.960	0.990	0.990
	25	10	0.420	0.430	0.750	0.730
		25	0.980	0.990	1.000	1.000
	30	10	0.640	0.680	0.900	0.880
		25	0.990	1.000	1.000	1.000

Table 3 Type I error and power of the model to check differentially expressed FC patterns between two groups in term of simulation setting: sample size N, structure of between-subject variability, number of regions *V*, degree of heterogeneity and dependence among the edges $\rho = 0.7$

$p = 0.7$			Compound Symmetry			Scaled	
			Heterogeneity			Heterogeneity	
	V	N	Low	High		Low	High
Type I Error	20	10	0.050	0.020		0.070	0.070
		25	0.040	0.010		0.040	0.030
	25	10	0.030	0.060		0.050	0.050
		25	0.080	0.060		0.040	0.040
	30	10	0.030	0.050		0.070	0.080
		25	0.080	0.050		0.070	0.070
Power	20	10	0.070	0.050		0.910	0.900
		25	0.150	0.250		1.000	1.000
	25	10	0.100	0.110		0.880	0.870
		25	0.190	0.400		1.000	1.000
	30	10	0.200	0.260		1.000	1.000
		25	0.600	0.770		1.000	1.000

In summary, the power based on identity scaled structure in a sample size of 25 exhibited values greater than 0.90 without influencing the degree of correlation, heterogeneity, and the number of regions. This index had values above 0.80 in the small sample size and $\rho = 0.7$. In most scenarios, the type I error was near to the nominal level. With increasing the number of regions in the sample $N = 25$ and $\rho = 0.3, 0.5$, the power based on compound symmetry structure was greater than 0.80. When there was a high correlation and a small sample size, however, the model yielded lower power. In most simulation items with high heterogeneity, the type I error based on compound symmetry structure was near to the nominal level.

Autism data

The proposed model in the current study was used to compare brain connectivity patterns between healthy individuals and patients with autism. Firstly, the estimation correlation coefficient and clustering of brain regions were calculated using the nonparametric Bayesian method to consider the network topological features. In this method, the brain network regions were divided into four clusters, $K = 4$, which the largest cluster consisted of 14 regions. Then the covariance matrix with dimensions of 435×435 was obtained using the dependency structure of clusters. Following that, the FC patterns between the patient and healthy groups were compared by considering the effect of heterogeneity. Based on compound symmetry and identity scaled structures, the test statistic values were $20.28(p = 0.517)$ and $21.68(p = 0.457)$, respectively. These results indicate no significant difference in
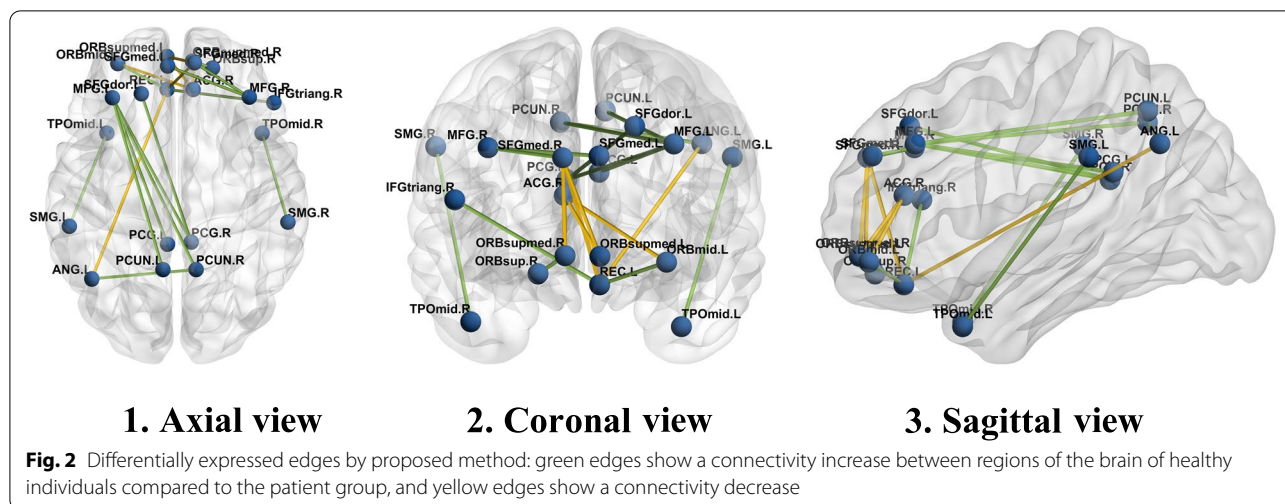
entire FC networks between patients and healthy individuals. Additionally, the Pard algorithm proposed by Chen et al. (2015) and the aSPU method proposed by Pan et al. (2014) were used to examine the differences in FC patterns [10, 37]. These models also displayed no significant FC networks between the two groups.

The FC comparisons of the pair ROIs were performed using test statistics of Eq. (6). Figure 2 depicts the different edges between the healthy and patient groups in three forms: axial, coronal, and sagittal views. In this chart, the increase of functional relationship of the healthy individuals compared to autism patients is manifested with a green edge, and the decrease of functional relationship is manifested with a yellow edge. The results showed out of 435, 18 edges were significant using both structures of variability. However, due to the high number of multiple comparisons, the *p*-value was corrected using the FDR approach. Therefore, after *p*-values adjustment, there was no substantial difference in connectivity between the two groups. Table 4 provides more details of connectivity changes between brain regions.

Discussion
Simulation data
To perform statistical inference on FC data, it seems necessary to consider the properties of heterogeneity between subjects and the dependency structure among edges. In previous studies, the spatial closeness between the regions defines the spatial structure [13–16, 38]. Accordingly, given that four regions define a pair of edges, it is difficult to consider the

Pourmotahari *et al. BMC Medical Research Methodology*   (2022) 22:273

Page 8 of 12

Fig. 2 Differentially expressed edges by proposed method: green edges show a connectivity increase between regions of the brain of healthy individuals compared to the patient group, and yellow edges show a connectivity decrease

distance between the edges in statistical models. On the other hand, the dependency between the edges is not necessarily limited to the spatial adjacency of the four regions. It can be related to the network topological structure. Recently, Chen et al. (2020) proposed a nonparametric Bayesian model based on network topological features to model the dependency structure of edges. However, this model does not include the heterogeneity effect between subjects which leads to accurate estimation of the parameters. The statistical model proposed in the present study investigates the FC network equality between the patient and healthy groups by combining the nonparametric Bayesian model for estimating the dependency structure between the edges and the algorithm proposed by Fiecas et al. (2017) for considering the heterogeneity of subjects [23, 24].

Simulation data showed that the power improves by increasing the correlation between the edges, based on identity scaled matrix structure. Accordingly, the Bayesian model proposed by Chen et al. (2020) had high accuracy when the correlation was above 0.3 [23]. These results indicate that model performance increases when detecting the network topological structure at higher correlations.

In the compound symmetry structure, the power was affected by the number of regions, sample size, and the edges correlation. Based on compound symmetry and identity scaled structures, the type I error was near to the nominal level when data dimensions and heterogeneity were high. In this regard, Fiecas et al. (2017) presented a variance component model by applying heterogeneity between subjects and temporal dependency, whose performance was similar to the proposed model based on compound symmetry structure in terms of statistical

power and type I error. However, in the proposed model, the power based on identity scaled structure yielded a higher value than the variance component model study without limiting the amount of heterogeneity and the number of regions in the sample size of 25 [24]. Accordingly, correct modelling of the dependency structure between the edges can improve the accuracy of estimating the model parameters and the power of the statistical test.

In FC studies, the scale of inference (i.e., at the edge, cluster or network scale) can significantly affect the results of statistical tests. It was recently shown that network-based analysis improves the power to capture average-size effects [39]. Consequently, it is important that the proposed model includes levels of inference from both the network and individual edges.

Another advantage of the proposed model compared to Fiecas et al.'s (2017) model is the improvement of computational power from 190 to 435 edges. However, since the interactions between the brain regions in high dimensions often show network topological features, applying more than 30 regions is one of the model's limitations due to the high computational cost and time. Figure 3 shows the computational time of the permutation test in different dimensions and the sample size of 10 and 25 on a computer with an i8 CPU and 16G RAM. In order to avoid computing difficulties, structures with a small number of parameters, including the identity scaled and compound symmetry, were used in this study to estimate the between-subject covariance. Therefore, any covariates were not considered in the model. On the other hand, Pard and aSPU models are some of the existing methods for evaluating FC networks, with no limitations in the dimensions of connectivity data [10, 37]. Since there are different

Table 4 Differentially expressed edges by proposed method; the ⇑ symbol shows a connectivity increase between brain regions in healthy individuals compared to the patients group. The ⇓ symbol shows a connectivity decrease.

	Different expressed edges			Control*	Autism*	Increase/ decrease
1	SFGdor.L	⟺	PCG.R	0.147	0.045	⇑
2	ORBsup.R	⟺	ORBsupmed.R	0.329	0.156	⇑
3	MFG.L	⟺	PCG.L	0.039	0.073-	⇓
4	MFG.L	⟺	PCG.R	0.094	-0.046	⇑
5	MFG.L	⟺	PCUN.R	0.058	0.072-	⇓
6	MFG.R	⟺	SFGmed.L	0.077	0.084-	⇓
7	MFG.R	⟺	SFGmed.R	0.246	0.095	⇑
8	ORBmid.L	⟺	REC.L	0.247	0.121	⇑
9	ORBmid.L	⟺	ACG.R	0.147-	0.007-	⇑
10	IFGtriang.R	⟺	REC.L	-0.073	-0.190	⇓
11	SFGmed.R	⟺	ORBsupmed.L	0.308	0.509	⇓
12	SFGmed.R	⟺	REC.L	0.178	0.352	⇓
13	ORBsupmed.L	⟺	ACG.R	0.127	0.269	⇓
14	ORBsupmed.R	⟺	ACG.R	0.243	0.401	⇓
15	REC.L	⟺	ANG.L	0.087	0.204	⇓
16	SMG.L	⟺	TPOmid.L	0.026-	0.142-	⇓
17	SMG.R	⟺	TPOmid.R	-0.006	0.109-	⇓
18	ANG.L	⟺	PCUN.R	0.283	0.129	⇑

* Parameter estimation of the edges between regions in patient and healthy group

Further details on the full names of the regions are available in the appendix.

patterns of functional connectivity for each subject [29], the heterogeneity effect between subjects is not included in these models.

Autism data

Estimating the FC patterns can provide a better understanding of the pathologies related to neurodegenerative diseases such as autism, and thus is essential in clinical research. The proposed statistical model in the current study was applied to resting-state fMRI data obtained on 25 patients with autism and 25 healthy individuals. There was no substantial difference in entire FC networks between the patient and healthy groups, which was consistent with the findings of the models proposed by Pan et al. (2014) and Chen et al. (2015) [10, 37]. These results contradict the results of previous studies examining the functional connectivity patterns of autistic patients compared to healthy individuals, which could be due to a large number of regions and the high sample size investigated in these studies [35, 40, 41]. The proposed model is applicable to the analysis of whole-brain connectivity data in practice. However, when there were more than 30 regions, the model's only shortcoming was the high computation time (Fig. 3). The model can be used with large
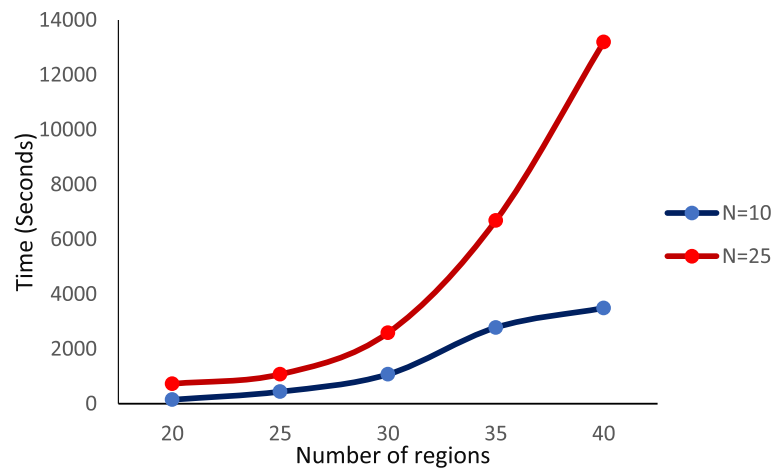
Fig. 3 Computation time for the permutation test in various dimensions and the sample sizes of 10 and 25

sample sizes as well. Nevertheless, in order to compare real data characteristics with the simulated scenarios, a sample size of 25 was considered for each patient and healthy group.

Another hypothesis in the proposed statistical model is the FC comparisons of pair regions. In this regard, the permutation test showed the difference in the functional connectivity of several regions between patients and healthy groups regardless of *p*-value adjustment. Most of these changes were related to DMN network regions, including frontal superior medial, cingulum (anterior and posterior), praecuneus, and angular. For example, in the patient group, there was a connectivity decrease of the right posterior cingulate region with the left superior frontal (dorsolateral), left middle frontal regions, and the right anterior cingulate region with the left middle frontal (orbital part). On the other hand, there was an increase in the dependency amount of the right frontal superior medial with the left superior frontal gyrus (medial orbital), and the left angular with the left rectus. Accordingly, several studies have reported functional connectivity changes of DMN regions in autism patients. [42–44]. DMN is one of the most important brain networks whose function changes under the influence of neurological disorders such as autism. These functional changes have a substantial impact on cognitive functions. [45].

Conclusion

This study presents a new approach for determining differentially expressed FC patterns between two groups, in which heterogeneity between subjects and the structure of dependency among edges are simultaneously considered. Simulation data indicated the high power and near-nominal type I error rates in the proposed model. Additionally, the application of the model on

real data related to autism was also evaluated, and there was no significant difference in the functional connectivity network between the patient and healthy groups.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-022-01712-8.

> **Additional file 1.**

Authors' contributions
FP, HA developed the statistical method. FP and SMT wrote the simulation computer code and performed the analysis of autism data. FP and NB interpreted the results. FP wrote the draft of the manuscript. HA, HD and NB reviewed and edited the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The corresponding author can provide the codes of the method upon reasonable request. The data presented in this study are available from the ABIDE database at http://preprocessed-connectomes-project.org/abide/.

Declarations

Ethics approval and consent to participate
Written informed consent was obtained for all participants, and experimental protocols were approved by the Institutional Review Board at the NYU Langone Medical Center. The autism dataset was de-identified in accordance with HIPAA (Health Insurance Portability and Accountability) guidelines, and no protected health information included. All methods were carried out in

accordance with relevant guidelines and regulations. The Ethics Committee of Shahid Beheshti University of Medical Sciences accepted this research (IR. SBMU.RETECH.REC.1399.1140).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [2]Department of Mathematics and Statistics, Macquarie University, Macquarie, Australia. [3]Urology and Nephrology Research Centre, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [4]Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran.

References
1. Greicius M. Resting-state functional connectivity in neuropsychiatric disorders. Curr Opin Neurol. 2008;21(4):424–30.
2. Weng S-J, Wiggins JL, Peltier SJ, Carrasco M, Risi S, Lord C, et al. Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders. Brain Res. 2010;1313:202–14.
3. Supekar K, Menon V, Rubin D, Musen M, Greicius MD. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput Biol. 2008;4(6):e1000100.
4. Wu T, Wang L, Chen Y, Zhao C, Li K, Chan P. Changes of functional connectivity of the motor network in the resting state in Parkinson's disease. Neurosci Lett. 2009;460(1):6–10.
5. He Y, Evans A. Graph theoretical modeling of brain connectivity. Curr Opin Neurol. 2010;23(4):341–50.
6. Fornito A, Zalesky A, Breakspear M. Graph analysis of the human connectome: promise, progress, and pitfalls. Neuroimage. 2013;80:426–44.
7. Simpson SL, Moussa MN, Laurienti PJ. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. Neuroimage. 2012;60(2):1117–26.
8. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci. 2009;10(3):186–98.
9. Simpson SL, Lyday RG, Hayasaka S, Marsh AP, Laurienti PJ. A permutation testing framework to compare groups of brain networks. Front Comput Neurosci. 2013;7:171.
10. Chen S, Kang J, Xing Y, Wang G. A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. Hum Brain Mapp. 2015;36(12):5196–206.
11. Kim J, Wozniak JR, Mueller BA, Pan W. Testing group differences in brain functional connectivity: using correlations or partial correlations? Brain Connect. 2015;5(4):214–31.
12. Zalesky A, Fornito A, Bullmore ET. Network-based statistic: identifying differences in brain networks. Neuroimage. 2010;53(4):1197–207.
13. Castruccio S, Ombao H, Genton MG. A scalable multi-resolution spatio-temporal model for brain activation and connectivity in fMRI data. Biometrics. 2018;74(3):823–33.
14. Kang H, Ombao H, Fonnesbeck C, Ding Z, Morgan VL. A Bayesian double fusion model for resting-state brain connectivity using joint functional and structural data. Brain Connect. 2017;7(4):219–27.
15. Wang R, Albert KM, Taylor WD, Boyd BD, Blaber J, Lyu I, et al. A bayesian approach to examining default mode network functional connectivity and cognitive performance in major depressive disorder. Psychiatry Res Neuroimaging. 2020;301:111102.
16. Li F, Zhang T, Wang Q, Gonzalez MZ, Maresh EL, Coan JA. Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. Ann Appl Stat. 2015;9(2):687–713.
17. Rothman AJ, Levina E, Zhu J. Generalized thresholding of large covariance matrices. J Am Stat Assoc. 2009;104(485):177–86.
18. Shen X, Pan W, Zhu Y. Likelihood-based selection and sharp parameter estimation. J Am Stat Assoc. 2012;107(497):223–32.
19. Cui Y, Leng C, Sun D. Sparse estimation of high-dimensional correlation matrices. Comput Stat Data Anal. 2016;93:390–403.
20. Bickel PJ, Levina E. Regularized estimation of large covariance matrices. Ann Stat. 2008;36(1):199–227.
21. Ravikumar P, Wainwright MJ, Raskutti G, Yu B. High-dimensional covariance estimation by minimizing ℓ1-penalized log-determinant divergence. Electron J Stat. 2011;5:935–80.
22. Chen S, Kang J, Xing Y, Zhao Y, Milton DK. Estimating large covariance matrix with network topology for high-dimensional biomedical data. Comput Stat Data Anal. 2018;127:82–95.
23. Chen S, Xing Y, Kang J, Kochunov P, Hong LE. Bayesian modeling of dependence in brain connectivity data. Biostatistics. 2020;21(2):269–86.
24. Fiecas M, Cribben I, Bahktiari R, Cummine J. A variance components model for statistical inference on functional connectivity networks. Neuroimage. 2017;149:256–66.
25. Mueller S, Wang D, Fox MD, Yeo BTT, Sepulcre J, Sabuncu MR, et al. Individual variability in functional connectivity architecture of the human brain. Neuron. 2013;77(3):586–95.
26. Wang D, Liu H. Functional connectivity architecture of the human brain: not all the same. Neurosci. 2014;20(5):432–8.
27. Nawaz U, Lee I, Beermann A, Eack S, Keshavan M, Brady R Jr. Individual variation in functional brain network topography is linked to schizophrenia symptomatology. Schizophr Bull. 2021;47(1):180–8.
28. Price RB, Lane S, Gates K, Kraynak TE, Horner MS, Thase ME, et al. Parsing heterogeneity in the brain connectivity of depressed and healthy adults during positive mood. Biol Psychiatry. 2017;81(4):347–57.
29. DiLernia A, Quevedo K, Camchong J, Lim K, Pan W, Zhang L. Penalized model-based clustering of fMRI data. arXiv Prepr arXiv201006408. 2020.
30. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes with source code CD-ROM 3rd edition: the art of scientific computing: Cambridge University Press; 2007.
31. Di Martino A, Yan C-G, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry. 2014;19(6):659–67.
32. Craddock C, Benhajali Y, Chu C, Chouinard F, Evans A, Jakab A, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. Front Neuroinform. 2013;7.
33. Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R. Movement-related effects in fMRI time-series. Magn Reson Med. 1996;35(3):346–55.
34. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002;15(1):273–89.
35. Chen S, Xing Y, Kang J. Latent and abnormal functional connectivity circuits in autism spectrum disorder. Front Neurosci. 2017;11:125.
36. Xu S, Li M, Yang C, Fang X, Ye M, Wei L, et al. Altered functional connectivity in children with low-function autism spectrum disorders. Front Neurosci. 2019;13:806.
37. Pan W, Kim J, Zhang Y, Shen X, Wei P. A powerful and adaptive association test for rare variants. Genetics. 2014;197(4):1081–95.
38. Al-Momani M, Hussein AA, Ahmed SE. Penalty and related estimation strategies in the spatial error model. Statistica Neerlandica. 2017;71(1):4–30.
39. Noble S, Mejia M, Zalesky A, Scheinost D. Leveling up: improving power in fMRI by moving beyond cluster-level inference. BioRxiv. 2021.
40. Chen H, Nomi JS, Uddin LQ, Duan X, Chen H. Intrinsic functional connectivity variance and state-specific under-connectivity in autism. Hum Brain Mapp. 2017;38(11):5740–55.
41. Chen S, Kang J, Wang G. An empirical Bayes normalization method for connectivity metrics in resting state fMRI. Front Neurosci. 2015;9:316.
42. Jung M, Kosaka H, Saito DN, Ishitobi M, Morita T, Inohara K, et al. Default mode network in young male adults with autism spectrum disorder: relationship with autism spectrum traits. Mol Autism. 2014;5(1):1–11.

43. Monk CS, Peltier SJ, Wiggins JL, Weng S-J, Carrasco M, Risi S, et al. Abnormalities of intrinsic functional connectivity in autism spectrum disorders. Neuroimage. 2009;47(2):764–72.
44. Murdaugh DL, Shinkareva SV, Deshpande HR, Wang J, Pennick MR, Kana RK. Differential deactivation during mentalizing and classification of autism based on default mode network connectivity. PLoS One. 2012;7(11):e50064.
45. Padmanabhan A, Lynch CJ, Schaer M, Menon V. The default mode network in autism. Biol Psychiatry Cogn Neurosci Neuroimaging. 2017;2(6):476–86.

Publisher's Note