Beyond Massive Univariate Tests: Covariance Regression Reveals Complex Patterns of Functional Connectivity Related to Attention-Deficit/Hyperactivity Disorder, Age, Sex, and Response Control

Supplementary Information

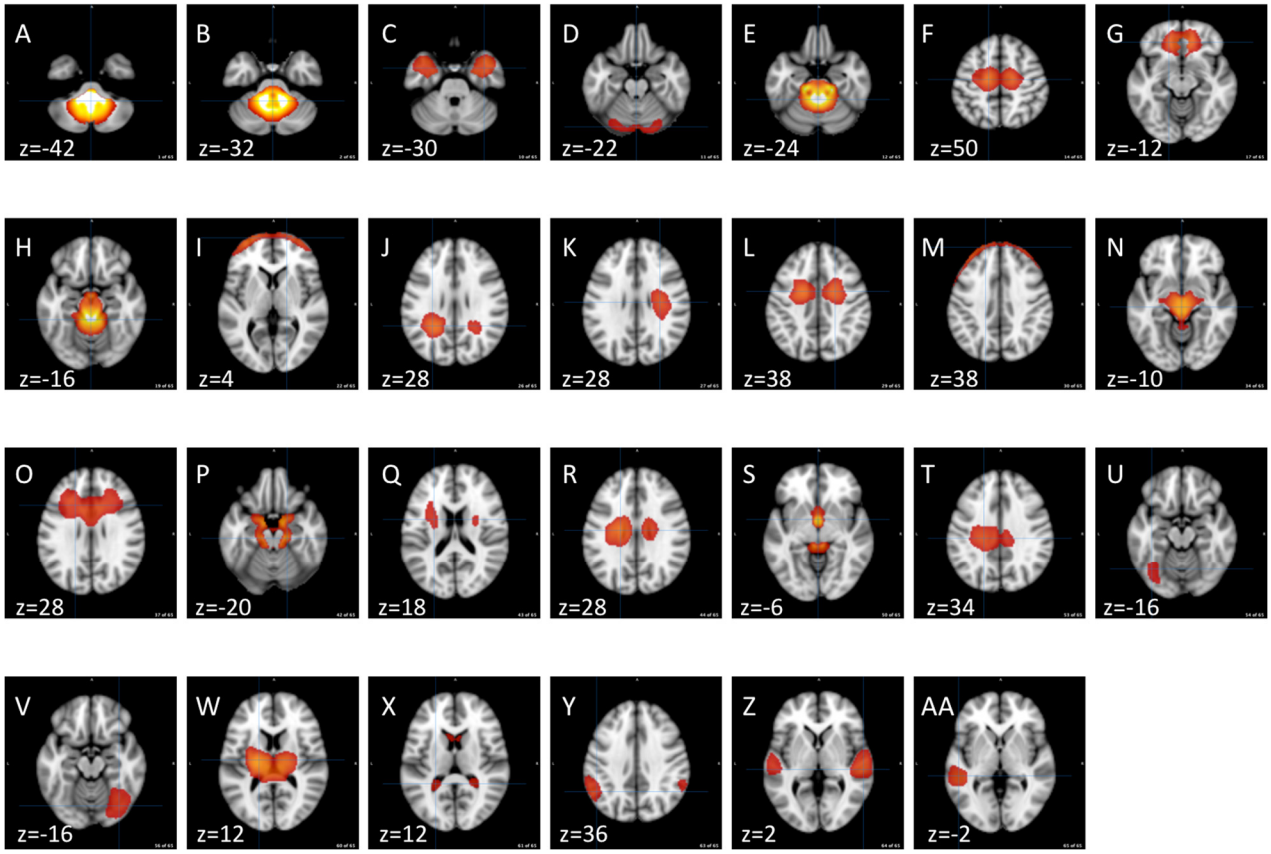
Diagnostic procedures. Intellectual ability was assessed using the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV)[1] or Fifth Edition (WISC-V).[2] Participants with general ability index (GAI) scores, a measure of intellectual reasoning ability that does not factor in working memory and processing speed scores, below 80 were excluded. To screen for reading disorders, children were administered the Word Reading subtest from the Wechsler Individual Achievement Test, Second Edition (WIAT-II)[3] or Third Edition (WIAT-III)[4] and were excluded for standard scores below 85. Diagnostic status was established through administration of either the Diagnostic Interview for Children and Adolescents, Fourth Edition (DICA-IV)[5] or the Kiddie Schedule for Affective Disorders and Schizophrenia for School Aged Children Present Lifetime version (KSADS-PL).[6] Children meeting criteria for diagnosis of conduct, mood, generalized anxiety, separation anxiety or obsessive–compulsive disorders on either interview were excluded whereas a comorbid diagnosis of oppositional defiant disorder (ODD) was permitted. Parents and teachers (when available) also completed the Conners Parent and Teacher Rating Scales-Revised Long Version or the Conners-3 (CPRS and CTRS)[7,8] and the ADHD Rating Scale-IV (ADHD-RS), home and school versions.[9] A diagnosis of ADHD was confirmed by a child neurologist or psychologist based on the diagnostic interview, which considered information provided by the parent about functioning at school, in addition to onset, course, duration, and frequency of symptoms, and parent/teacher rating scales (i.e., T-scores ≥ 65 or ≥ 6 symptoms endorsed on at least one rating scale). Inclusion in

the TD group required scores below clinical cutoffs (i.e., T-scores ≤ 60 and ≤ 4 symptoms endorsed on all parent/teacher rating scales.

Resting-state fMRI data acquisition. Resting state fMRI (rs-fMRI) was acquired on a 3.0 T Philips scanner using a single-shot, partially parallel, gradient-recalled echo planar sequence with sensitivity encoding and an ascending slice order (repetition time [TR]/echo time [TE] = 2500/30ms, flip angle = 75^{o}, sensitivity encoding acceleration factor of 2, 47 3-mm axial slices with no slice gap, in-plane resolution of 3.05×3.15 mm [84×81 voxels], duration = 5 min 20 sec – 6 min 30 s). Participants were instructed to relax, fixate on a cross-hair, and remain as still as possible.

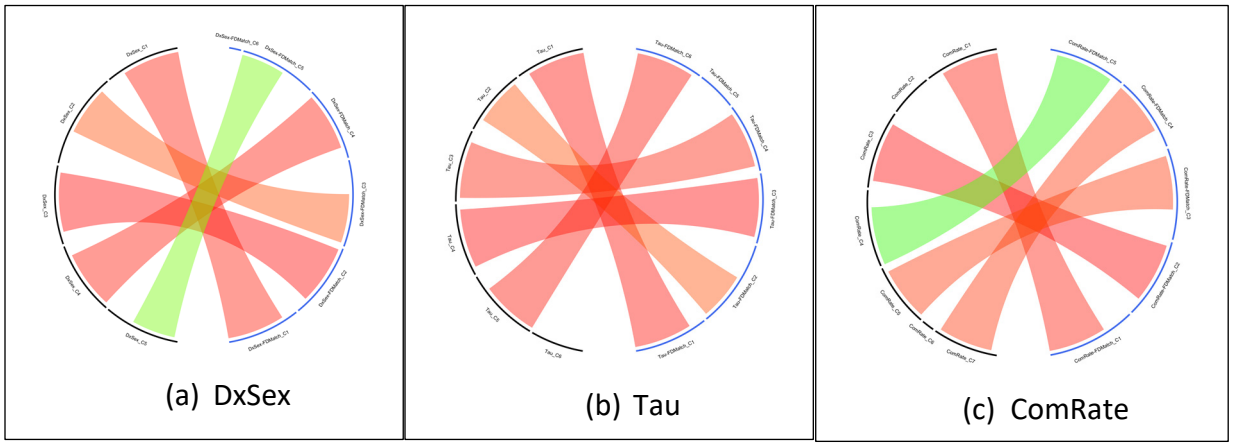
Preprocessing of fMRI data. Functional data were preprocessed using SPM12 (Wellcome Trust Centre for Neuroimaging, London, United Kingdom) and custom MATLAB (The Mathworks, Inc., Natick, Massachusetts) code. rs-fMRI scans were slice-time adjusted using the slice acquired in the middle of the TR as a reference, and rigid body realignment parameters were estimated to adjust for motion. The volume collected in the middle of the scan was spatially normalized using the Montreal Neurological Institute (MNI) EPI template.[12] The estimated rigid body and nonlinear spatial transformations were applied to the functional data together, producing 2-mm isotropic voxels in MNI space. Linear trends were removed, and the data were spatially smoothed using a Gaussian filter (6-mm full width at half maximum kernel). Mean framewise displacement (FD) was calculated using the realignment estimates.[10]

Supplementary Figure S1. Spatial maps of the 27 components classified as noise. Z coordinate in MNI space of each axial slice is indicated. Although most voxels contributing to component Y appear to be within grey matter, it was labeled as a noise component because it was not reliably identified across the 100 iterations of group ICA that we ran.



Secondary analyses on motion-matched subgroups. We applied the CAP regression on FD matched samples (TD n=29 girls, 66 boy; ADHD n=25 girls, 79 boys), where the matching was conducted for each diagnosis*sex group, and compared the results with the unmatched FD sample. **Supplementary Figure S2** presents the similarity between the components identified among the FD-matched subgroups and the identified components using all participants. As shown in the figures, the components are almost identical and the effects are also consistent, suggesting that the pattern of findings holds when diagnostic groups do not differ in motion during the scan.

Supplementary Figure S2. Similarity between the components identified using all samples and the components identified using FD matched samples.



ICA with Backward Reconstruction. We used an information theoretic approach to dimension estimation,[11] estimating the optimal number of components to represent each participant's data separately. We then chose the number of independent components (ICs) for the group to be the maximum dimension estimate across participants (65) to ensure that the group decomposition would represent all networks present in the full dataset. Prior to ICA, each participant's preprocessed data were variance normalized on a voxelwise basis and reduced to 100 principle components (PCs) using principal component analysis (PCA). Participant-specific PCs were temporally concatenated and a second PCA was used to reduce the aggregate data set to the maximum dimension estimated, 65 (defined above) using multi-power iteration.[12] ICA was repeated on the group-level PCs 100 times using the Infomax algorithm[13] and the ICASSO toolbox[14] with randomized initial conditions in GIFT to ensure stable ICs. Participant-specific spatial maps (SMs) and timecourses (TCs) were generated from the aggregate IC decomposition using a method based on PCA compression and projection.[15] The SMs represent the spatial topography of each component within the brain while the TCs represent the intrinsic level of engagement of each component over time.

Previous research has established that participant specific spatial maps that are back-reconstructed from the group components are consistent with components derived from single-

subject ICA.[15] It is still possible that subtle variations in the voxels that contribute to each IC exist between groups, so we tested the possibility that network topography differed across groups. First, we converted the participant-specific spatial maps of the 38 components we classified as representing resting state networks (signal components) to z values so that image intensities reflected the degree to which the component was present in each participant's data. These participant-specific signal spatial maps were combined in a second-level random effects analysis using a two-sample *t* test in SPM12. Voxels that contributed unequally to the components across groups were identified using a liberal voxelwise *p* value of .01 and a cluster-level *p* value of .05 corrected for multiple comparisons.[16] We found no significant group differences in network topography for any of the 38 signal components using this approach.

Covariate Assisted Principal Regression on RSN Components. We implemented covariate assisted principal (CAP) regression on the signal components identified by group ICA described above.[17] For subject i, let A_{it} denote the K IC time courses at time t for $t = 1, \ldots, T_i$. The CAP approach assumes that there exist orthogonal linear projections $\boldsymbol{\gamma}_r$ for $r = 1, \ldots, R$ ($R \leq K$), such that in the projection space, the data variation satisfies the following log-linear model:

$$\log\left\{\mathrm{Var}(\boldsymbol{\gamma}_r^\top A_{it})\right\} = \log\left\{\mathrm{Var}(Z_{itr})\right\} = \beta_{0r} + \mathbf{x}_i^\top \boldsymbol{\beta}_{1r},$$

where $Z_{itr} := \boldsymbol{\gamma}_r^\top A_{it}$; $\mathbf{x}_i$ is the vector of covariates of interest; β_{0r} and $\boldsymbol{\beta}_{1r}$ are model coefficients; $\boldsymbol{\gamma}_r^\top \boldsymbol{\gamma}_s = 1$ if $r = s$ and 0 otherwise. The projections and model coefficients can be estimated by maximizing the likelihood function assuming A_{it} is normally distributed with mean zero and covariance matrix $\boldsymbol{\Sigma}_i$ (for $t = 1, \ldots, T_i$ and $i = 1, \ldots, n$). The number of projections, R, is determined based on the level of deviation from diagonality.[18] To draw inference about the model coefficients, 95% confidence intervals were acquired from 500 bootstrap samples fixing the estimated linear projection. The method was implemented using the R package cap available on CRAN.

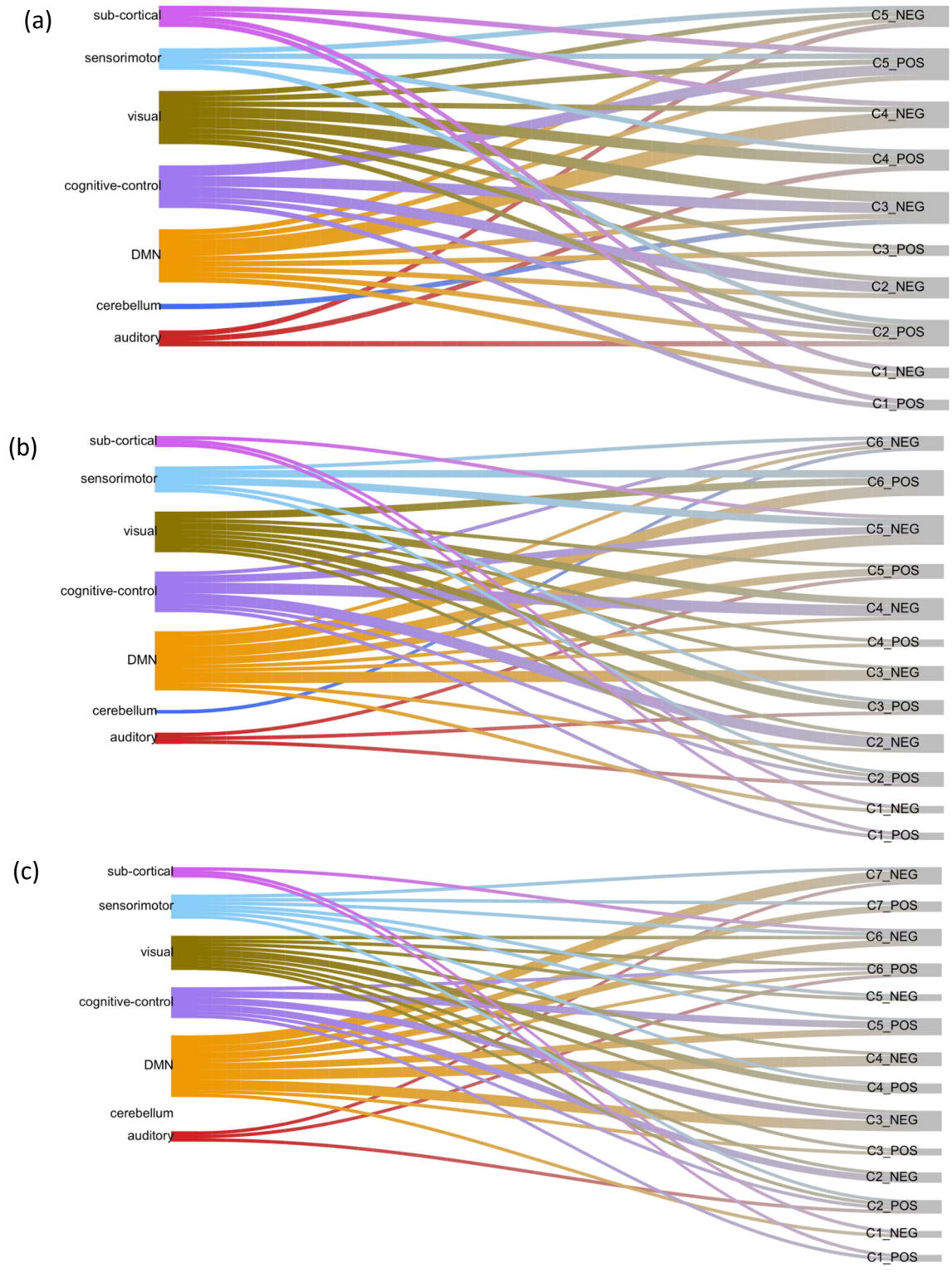
CAP components were reconstructed in voxel space to form brain maps representing orthogonal groups of signal ICs associated with the variables of interest. Let $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_R)$ denote the orthonormal projection matrix estimated by CAP regression. Each row of $\mathbf{\Omega} = \mathbf{\Gamma}^\top \mathbf{S}$ then represents the newly constructed CAP brain map, where $\mathbf{S}$ is the spatial maps obtained from the group ICA.

The CAP method identifies a linear projection of the covariance matrices such that between-subject variability in FC is most strongly associated with the covariates of interest. Let $\boldsymbol{\gamma}_r = (\gamma_{r1}, \ldots, \gamma_{rK})^\top$ and $\mathbf{\Sigma}_i = (\sigma_{ijk})_{j,k}$ for $i = 1, \ldots, n$, then

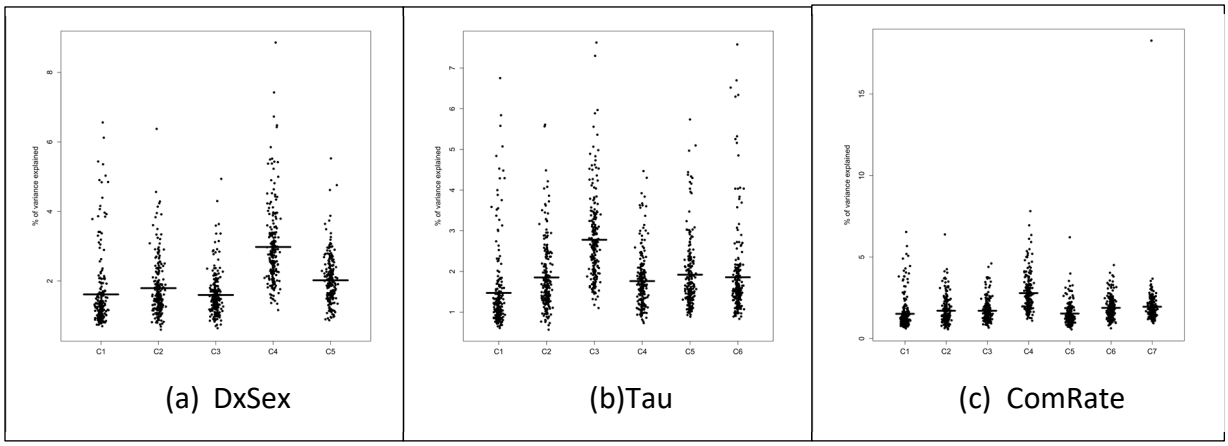
$$\mathrm{Var}(\boldsymbol{\gamma}_r^\top \mathbf{A}_{it}) = \boldsymbol{\gamma}_r^\top \mathbf{\Sigma}_i \boldsymbol{\gamma}_r = \sum_{j=1}^{K} \gamma_{rj}^2 \sigma_{ijj} + \sum_{j \neq k} \gamma_{rj}\gamma_{rk}\sigma_{ijk} = \exp\left\{\beta_{0r} + \mathbf{x}_i^\top \boldsymbol{\beta}_{1r}\right\}.$$

Assuming the IC time courses are standardized to have identical variance (for example, $\sigma_{ijj} = 1$ for any i and j), the CAP regression models the association between FC and the covariates. This association depends on both the sign of the β coefficient and the sign of the loading products. For a positive β estimate, FC between two ICs with the same loading sign (then the product is positive) is positively associated with the corresponding covariate; while FC for two ICs with the opposite signs (then the product is negative) is negatively associated with the covariate.

Supplementary Figure S3. River plot of IC loadings of the components identified by the CAP method for the (a) DxSex (b) Tau and (c) ComRate models. The IC indexes are color coded by the functional module.



Supplementary Figure S4. Percentage of variation explained by the identified components for each subject. From the figure, component C4 of the DxSex model, component C3 of the Tau model, and component C4 in the ComRate model explains the largest percentage of data variation.



Supplementary Table S1. Coefficients and 95% confidence intervals obtained from 500 bootstrap samples in the three models.

(a) The Dx-Sex model

Effect	DxSex				
	C1	C2	C3	C4	C5
GAI	-0.003 (-0.010, 0.004)	-0.005 (-0.009, -0.001)	-0.001 (-0.004, 0.002)	-0.000 (-0.003, 0.003)	-0.001 (-0.004, 0.002)
Age	0.145 (0.085, 0.205)	-0.022 (-0.062, 0.018)	0.122 (0.091, 0.153)	0.037 (0.006, 0.068)	-0.092 (-0.118, -0.065)
Dx*Sex	0.201 (-0.123, 0.541)	-0.035 (-0.245, 0.160)	-0.157 (-0.318, -0.002)	-0.177 (-0.371, 0.034)	-0.012 (-0.183, 0.149)
ADHD-TD (Boy)	0.118 (-0.066, 0.302)	0.118 (-0.034, 0.270)	-0.155 (-0.271, -0.039)	-0.137 (-0.236, -0.038)	0.071 (-0.026, 0.168)
ADHD-TD (Girl)	-0.083 (-0.350, 0.183)	0.153 (0.022, 0.285)	0.002 (-0.118, 0.122)	0.040 (-0.133, 0.213)	0.083 (-0.054, 0.219)
Boy-Girl (ADHD)	0.151 (-0.058, 0.360)	0.346 (0.216, 0.475)	-0.021 (-0.130, 0.088)	-0.434 (-0.562, -0.306)	0.224 (0.110, 0.338)
Boy-Girl (TD)	-0.050 (-0.301, 0.201)	0.381 (0.232, 0.530)	0.136 (0.016, 0.255)	-0.257 (-0.415, -0.099)	0.236 (0.112, 0.360)
Behavior (ADHD Boy)					
Behavior (TD Boy)					
Behavior (ADHD Girl)					
Behavior (TD Girl)					

(b) The GNG Tau model

Effect	Tau					
	C1	C2	C3	C4	C5	C6
GAI	-0.004 (-0.010, 0.003)	-0.004 (-0.008, -0.000)	0.000 (-0.002, 0.003)	-0.000 (-0.004, 0.003)	-0.003 (-0.007, 0.000)	0.000 (-0.005, 0.006)
Age	0.161 (0.100, 0.222)	0.023 (-0.015, 0.061)	0.015 (-0.021, 0.052)	0.126 (0.095, 0.157)	-0.039 (-0.081, 0.003)	0.035 (-0.020, 0.090)
Dx*Sex	-0.047 (-2.279, 2.210)	-0.039 (-1.534, 1.906)	1.143 (-0.290, 3.093)	0.328 (-0.914, 1.410)	-1.223 (-2.791, 0.047)	0.827 (-0.937, 3.609)
Dx*Beh	-0.062 (-0.506, 0.300)	0.106 (-0.142, 0.426)	0.075 (-0.201, 0.450)	-0.053 (-0.278, 0.147)	0.303 (0.002, 0.545)	-0.110 (-0.405, 0.349)
Sex*Beh	0.307 (-0.118, 0.755)	-0.029 (-0.458, 0.590)	-0.131 (-0.407, 0.167)	0.177 (-0.033, 0.403)	-0.284 (-0.505, -0.044)	0.495 (0.162, 0.872)
Dx*Sex*Beh	0.035 (-0.459, 0.502)	-0.029 (-0.458, 0.308)	-0.253 (-0.664, 0.059)	-0.105 (-0.355, 0.169)	0.314 (0.025, 0.660)	-0.205 (-0.792, 0.179)
ADHD-TD (Boy)	0.177 (-1.469, 1.824)	-0.314 (-1.599, 0.971)	0.751 (-0.072, 1.575)	0.590 (-0.444, 1.624)	-2.599 (-3.543, -1.654)	1.518 (-0.326, 3.362)
ADHD-TD (Girl)	0.225 (-1.661, 2.110)	-0.275 (-1.627, 1.077)	-0.391 (-1.960, 1.178)	0.262 (-0.690, 1.213)	-1.375 (-2.563, -0.187)	0.690 (-1.136, 2.517)
Boy-Girl (ADHD)	-1.437 (-3.888, 1.013)	-0.739 (-2.266, 0.788)	1.415 (0.026, 2.805)	-0.355 (-1.476, 0.765)	-0.052 (-1.457, 1.353)	-1.140 (-3.138, 0.858)
Boy-Girl (TD)	-1.390 (-3.348, 0.568)	-0.700 (-2.118, 0.718)	0.273 (-0.960, 1.505)	-0.683 (-1.663, 0.297)	1.171 (0.108, 2.234)	-1.967 (-3.636, -0.298)
Behavior (ADHD Boy)	0.187 (-0.070, 0.443)	0.267 (0.110, 0.423)	-0.206 (-0.317, -0.095)	-0.060 (-0.195, 0.075)	0.268 (0.100, 0.436)	0.007 (-0.260, 0.274)
Behavior (TD Boy)	0.213 (-0.082, 0.509)	0.189 (-0.057, 0.436)	-0.027 (-0.193, 0.138)	0.098 (-0.107, 0.302)	-0.350 (-0.482, -0.217)	0.322 (-0.012, 0.657)
Behavior (ADHD Girl)	-0.156 (-0.626, 0.315)	0.053 (-0.223, 0.329)	0.179 (-0.105, 0.462)	-0.132 (-0.350, 0.087)	0.238 (-0.074, 0.549)	-0.283 (-0.646, 0.080)
Behavior (TD Girl)	-0.094 (-0.082, 0.509)	-0.053 (-0.300, 0.194)	0.103 (-0.141, 0.348)	-0.079 (-0.241, 0.083)	-0.065 (-0.274, 0.143)	-0.173 (-0.421, 0.075)

(c) The GNG Commission Error Rate (CR) model

Effect	Com Rate						
	C1	C2	C3	C4	C5	C6	C7
GAI	-0.003 (-0.010, 0.003)	-0.004 (-0.008, -0.000)	-0.000 (-0.003, 0.003)	-0.001 (-0.004, 0.002)	-0.001 (-0.004, 0.003)	0.007 (0.004, 0.010)	-0.005 (-0.008, -0.002)
Age	0.151 (0.094, 0.209)	-0.010 (-0.055, 0.035)	0.136 (0.104, 0.168)	0.031 (-0.002, 0.065)	-0.026 (-0.059, 0.008)	0.013 (-0.020, 0.045)	-0.048 (-0.086, -0.009)
Dx*Sex	-0.066 (-0.795, 0.703)	-0.502 (-1.007, -0.042)	-0.087 (-0.439, 0.279)	-0.208 (-0.623, 0.219)	-0.110 (-0.625, 0.351)	0.007 (-0.408, 0.496)	0.670 (-0.077, 1.573)
Dx*Beh	-0.229 (-1.489, 1.124)	-0.538 (-1.274, 0.210)	-0.136 (-0.791, 0.557)	0.099 (-0.678, 0.971)	-0.037 (-1.028, 0.785)	-0.293 (-1.174, 0.539)	1.031 (-0.285, 2.662)
Sex*Beh	0.097 (-1.146, 1.413)	-0.832 (-1.620, -0.110)	0.618 (-0.081, 1.309)	-0.465 (-1.175, 0.353)	-0.917 (-1.507, -0.340)	0.819 (0.201, 1.505)	1.861 (0.587, 3.479)
Dx*Sex*Beh	0.643 (-0.917, 2.262)	1.119 (0.078, 2.207)	-0.029 (-0.977, 0.898)	0.166 (-0.707, 1.022)	0.089 (-0.930, 1.220)	0.076 (-1.002, 1.098)	-1.278 (-2.972, 0.167)
ADHD-TD (Boy)	-0.083 (-0.535, 0.369)	-0.164 (-0.514, 0.185)	-0.052 (-0.290, 0.186)	-0.204 (-0.436, 0.028)	0.161 (-0.081, 0.404)	-0.098 (-0.367, 0.171)	0.090 (-0.144, 0.325)
ADHD-TD (Girl)	-0.017 (-0.638, 0.603)	0.338 (-0.010, 0.686)	0.035 (-0.228, 0.298)	0.004 (-0.383, 0.391)	0.272 (-0.119, 0.663)	-0.106 (-0.459, 0.248)	-0.580 (-1.474, 0.314)
Boy-Girl (ADHD)	-0.140 (-0.640, 0.360)	0.206 (-0.165, 0.578)	-0.234 (-0.477, 0.010)	-0.276 (-0.546, -0.006)	0.330 (-0.071, 0.731)	-0.384 (-0.730, -0.039)	-0.070 (-0.303, 0.162)
Boy-Girl (TD)	-0.074 (-0.663, 0.515)	0.709 (0.368, 1.050)	-0.147 (-0.414, 0.120)	-0.068 (-0.426, 0.290)	0.440 (0.222, 0.659)	-0.392 (-0.654, -0.130)	-0.740 (-1.604, 0.123)
Behavior (ADHD Boy)	0.392 (-0.363, 1.147)	0.271 (-0.217, 0.760)	0.238 (-0.093, 0.570)	-0.314 (-0.567, -0.060)	0.294 (-0.100, 0.689)	0.080 (-0.231, 0.392)	0.099 (-0.219, 0.417)
Behavior (TD Boy)	-0.022 (-0.701, 0.658)	-0.310 (-0.838, 0.218)	0.403 (-0.103, 0.910)	-0.579 (-1.014, -0.144)	0.243 (-0.118, 0.604)	0.298 (-0.186, 0.783)	0.346 (-0.004, 0.696)
Behavior (ADHD Girl)	-0.348 (-1.074, 0.377)	-0.015 (-0.561, 0.530)	-0.351 (-0.851, 0.150)	-0.014 (-0.530, 0.501)	1.123 (0.350, 1.897)	-0.814 (-1.463, -0.165)	-0.484 (-0.855, -0.113)
Behavior (TD Girl)	-0.119 (-1.286, 1.048)	0.522 (-0.005, 1.049)	-0.214 (-0.729, 0.300)	-0.113 (-0.756, 0.529)	1.160 (0.692, 1.629)	-0.520 (-1.044, 0.003)	-1.515 (-3.133, 0.103)

Supplementary References

1. Wechsler DL. *Wechsler Intelligence Scale for Children - Fourth Edition (WISC-IV)*. The Psychological Corporation; 2003.

2. Wechsler DL. *Wechsler Intelligence Scale for Children - Fifth Edition (WISC-V)*. The Psychological Corporation; 2014.

3. Wechsler DL. *Wechsler Individual Achievement Test - Second Edition (WIAT-II)*. The Psychological Corporation; 2002.

4. Wechsler DL. *Wechsler Individual Achievement Test - Third Edition (WIAT-III)*. Pearson Education PsychCorp; 2009.

5. Reich W. Diagnostic interview for children and adolescents (DICA). *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000;39(1):59-66. doi:10.1097/00004583-200001000-00017

6. Kaufman J Birmaher, B, Axelson, D, Perepletchikova, F, Brent, D,& Ryan, N. Kiddie Schedule for Affective Disorders and Schizophrenia for School-Aged Children – Lifetime Version (Kiddie-SADS-PL 2013 Working Draft). Published online 2013.

7. Conners CK. *Conners' Rating Scales- Revised*. Multi-Health Systems, Inc.; 2002.

8. Conners CK. Conners 3. Published online 2008.

9. DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. *ADHD Rating Scale—IV*. Guilford Press; 1998.

10. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*. 2012;59(3):2142-2154. doi:10.1016/j.neuroimage.2011.10.018

11. Li YO, Adali T, Calhoun VD. Estimating the number of independent components for functional magnetic resonance imaging data. *Human brain mapping*. 2007;28(11):1251-1266. doi:10.1002/hbm.20359

12. Rachakonda S, Silva RF, Liu J, Calhoun VD. Memory Efficient PCA Methods for Large Group ICA. *Front Neurosci*. 2016;10. doi:10.3389/fnins.2016.00017

13. Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*. 1995;7(6):1129-1159.

14. Himberg J, Hyvarinen A, Esposito F. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage*. 2004;22(3):1214-1222. doi:10.1016/j.neuroimage.2004.03.027

15. Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. *Human brain mapping*. 2001;14(3):140-151.

16. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp*. 1996;4(1):58-73. doi:10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O

17. Zhao Y, Wang B, Mostofsky SH, Caffo BS, Luo X. Covariate Assisted Principal regression for covariance matrix outcomes. :17.

18. Flury BN, Gautschi W. An Algorithm for Simultaneous Orthogonal Transformation of Several Positive Definite Symmetric Matrices to Nearly Diagonal Form. *SIAM J Sci and Stat Comput*. 1986;7(1):169-184. doi:10.1137/0907013