Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Liwei Wang Date

The impact of quality control exclusion criteria on functional connectivity in children with neurodevelopmental disorders

By

Liwei Wang
Master of Science in Public Health

Biostatistics and Bioinformatics

Benjamin Risk, Ph.D.
Advisor

David Benkeser, Ph.D.
Committee Member

The impact of quality control exclusion criteria on functional connectivity in children with neurodevelopmental disorders

By

Liwei Wang
B.E., Tianjin University, Tianjin, 2013
M.Sc., Emory University, GA, 2020

Advisor: Benjamin Risk, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics and Bioinformatics
2020

Abstract

The impact of quality control exclusion criteria on functional connectivity in children with neurodevelopmental disorders
By Liwei Wang

Background: Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD) are two serious neurodevelopmental diseases in the United States with increasing diagnosis rates. Resting-state functional magnetic resonance imaging (rsfMRI) may be a useful tool to characterize the neural underpinnings of neurodevelopmental disorders. Motion can create large artifacts in fMRI, and consequently, data with head motion is often excluded. These excluded images may contribute to selection bias, potentially mischaracterizing these neurodevelopmental disorders compared to typically developing individuals.

Objective: The purpose of this study was to improve estimates of functional connectivity in neurodevelopmental disorders by using detailed phenotypic information to account for non-random missingness in fMRI data in a large sample of elementary school-aged children, where missingness arises from quality control exclusion criteria.

Methods: We analyzed phenotypic information from 758 individuals and assessed the functional connectivity of a neurodevelopmental disease group and a typical developing group based on analyses of 473 children (8-13 years old), which included 119 subjects with ASD, 119 subjects with ADHD, and 235 typical developing subjects. We examined seed-based functional connectivity for a region located in the left posterior cingulate cortex (L-PCC). We applied doubly robust targeted minimum loss-based estimator (DRTMLE) to account for the possible confounding due to the patterns of missingness (driven by motion) being related to the severity of the neurodevelopmental disorders (captured by phenotypic information). We compared the standard analysis ignoring possible confounding with DRTMLE.

Results: First, the non-random missingness in SRS, PANESS, and GAI is triggered by the participants from multiple studies and some of them did not collect SRS, PANESS, and GAI. Second, we found ASD hyperconnectivity between the L-PCC and frontoparietal regions, and this hyperconnectivity was more extensive in DRTMLE than in the naive approach. We also found regions of ASD-related hypoconnectivity in the temporal lobe in DRTMLE that was not apparent in the naive approach. Differences between ADHD and TD generally resembled those in ASD versus TD but less extensive.

Conclusion: Head motion exclusion criteria may lead to biases in the sample of children included in analyses of neurodevelopmental disorders, where more severe pathology may be excluded due to motion. DRTMLE allows the incorporation of phenotypic information to weight data, which may have led to the larger differences

between ASD and typically developing observed in the DRTMLE approach versus the naive approach that ignores missingness. This suggests DRTMLE can be used to improve our understanding of neurodevelopmental disorders.

The impact of quality control exclusion criteria on functional connectivity in children with neurodevelopmental disorders

By

Liwei Wang
B.E., Tianjin University, Tianjin, 2013
M.Sc., Emory University, GA, 2020

Advisor: Benjamin Risk, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics and Bioinformatics
2020

Acknowledgments

Initially, thanks to my parents and family to support me to pursue the Master of Science degree. Without their encouragement, I could not have the chance to step to my graduation from the Department of Biostatistics and Bioinformatics.

I appreciate my thesis advisor Dr. Benjamin Risk of the Department of Biostatistics and Bioinformatics at Emory University. With his instruction and aid in the statistical aspect of the neuroimaging area, I overcame difficulties in the project that I am interested in and finish it as my thesis.

I appreciate Dr. Mary Beth Nebel of the School of Medicine at Johns Hopkins University as the practicum supervisor to guide my research in the last summer and give me lots of advice on the thesis. With her instruction, my scientific writing skills and background knowledge in neuroimaging have been improved a lot.

I am also grateful to Dr. David Benkeser of the Department of Biostatistics and Bioinformatics at Emory University as the reader of this thesis. Thanks for his statistical support on the causal inference area in our project. In addition, I am also deeply indebted to him for his very valuable comments on this thesis.

Finally, thanks to Jo Etzel and her tutorial on mapping the volume to surface by Connectome Workbench.

Contents

1 Introduction

Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD) are the most common neurodevelopmental diseases[13]. In the United States, ASD and ADHD are serious and growing neurodevelopmental diseases. The diagnostic criteria for the ASD include social deficit and restricted and repetitive behavior, interests and activities[1]. In 2014, 1.7% of children have been identified as ASD by Autism and Developmental Disabilities Monitoring (ADDM) Network[6]. ADHD is diagnosed as inattention, hyperactivity, and impulsivity[1]. Having based a 2016 parent survey, the estimated percentage of children diagnosed as ADHD is increased to 9.4% from 4.4%[9].

To characterize the mechanisms and causes of these neurodevelopmental diseases, many neuroimaging techniques are currently being investigated to analyze neural pathways. Functional magnetic resonance imaging (fMRI) is a common tool which detects the brain activities between brain regions by measuring the fluctuation of blood-oxygen-level-dependent (BOLD) signals[15]. Recently, resting-state fMRI (rsfMRI) has been used to detect differences in functional connectivity, which is defined as correlation between different brain regions across time[23]. However, head motion, which refers to subjects' head displacement during MRI scanning, is a significant challenge in fMRI[20]. Head motion can introduce spurious group differences in connectivity estimates derived from fMRI data[24]. Consequently, head motion exclusion criteria are applied to fMRI to remove images that may contain motion artifacts[7]. Several studies also have explored post-acquisition cleaning procedures for minimizing motion artifacts in functional connectivity estimates[7;18;22]. However, no technique can perfectly solve the problem of motion artifacts. aCompCor is a method used for reducing the distortion of head motion on the resting-state fMRI (rsfMRI) to improve the specificity of estimation of functional connectivity. However, a limitation of aCompCor is

that it is only applicable to low-motion data[18]. Global signal regression can reduce the impact of motion on functional connectivity, but there may still be issues with distance-dependent artifacts[18].

Neurodevelopmental diseases may be associated with excessive motion. Only recently, studies have found that head motion-related artifacts are associated with these traits of interest. Couvy-Duchesne et al.[8] found that head motion was moderately related to self-reported inattention and hyperactivity-impulsivity. As head motion criteria become increasingly stringent, data retention decreases while the potential to limit the generalizability of the study findings increases. While previous studies explored methods to alleviate the effect of artifacts related to head motion, they have not examined the possible impacts of selection bias on the resulting sample of children with neurodevelopmental disease. In these studies, bias may have been introduced prior to post-acquisition cleaning procedures, leading to possible mischaracterization of the functional connectivity among the children with neurodevelopmental disease.

Bias induced by missing data can sometimes be controlled using some form of covariate adjustment. Many methods have been developed to this end in the causal inference literature, often based on propensity scores. The propensity score is the probability of observing complete data on an individual given other relevant characteristics of the individual. Estimated propensity scores are sometimes used to match individuals to approximate a data set that might have been observed if there were no missingness[2]. Propensity scores can also be used for inverse probability weighting, which creates a "pseudo-population" in which missingness is independent of confounders by correcting for imbalance between features of individuals with complete vs. missing data[26]. However, IPWE can be unstable if the estimated propensity score is near zero. If some "types" of individuals (as defined by their covariate values) are only rarely observed to have complete data, then the inverse propensity weights can

become very large and cause inflated variance of the estimator. Another approach to covariate adjustment is using covariate-adjusted models of the outcome amongst individuals with the outcome measured. This outcome model is then re-weighted or standardized to the observed covariate distribution of all individuals in the study[21].

Other estimators combine a propensity score with an outcome model to yield a *doubly-robust estimator*. While the estimator is built based on two models, doubly-robust estimators only require either the propensity score or outcome regression to be consistently estimated in order to consistently estimate the association of interest. Accordingly, the stability and often the efficiency of the estimator is improved relative to other approaches. Here, we apply a Doubly-Robust Targeted Minimum Loss-based estimator (DRTMLE) to address the selection bias problem in our study[43].

In order to address the biases identified in the literature, the purpose of this study is to: 1) examine whether commonly used head motion exclusion criteria result in biased samples of children with ASD and ADHD, and 2) improve the estimates of differences in functional connectivity between these groups by adjusting for covariates, including social responsiveness score, the Physical and Neurological Exam for Subtle Signs (PANESS), and general ability index. We evaluated the relationship between the probability of being excluded due to head motion and each trait of interest, as well as evaluated how the traits of interest and motion related to these traits affects estimates of functional connectivity in children with neurodevelopmental diseases.

2 Data description

2.1 Study population

Our cohort is an aggregate of 758 children between 8- and 13-years old who participated in a neuroimaging study of either ASD or ADHD at Kennedy Krieger Institute

(KKI) between 2008 and 2019. Participants included 147 children with ASD (122 boys), 273 with ADHD (202 boys), and 338 typically developing (229 boys).

Table 1 summarizes demographic and phenotypic information for our cohort stratified by diagnosis. Diagnosis of ASD was confirmed using the Autism Diagnostic Observation Schedule-Generic (ADOS-G) or Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), depending on the date of enrollment, and the Autism Diagnostic Interview-Revised (ADI-R). To be included in the ADHD group, a participant had to 1) receive a t-score of 60 or higher on the inattentive or hyperactive subscales on the Conners' Parent or Teacher scales, or 2) received a score of 2 or 3 on at least 6/9 items on the Inattentive or Hyperactivity/Impulsivity scales of the ADHD Rating Scale-IV. Children were excluded from the TD group if they had a first-degree relative with ASD, if parent responses to either the Diagnostic Interview for Children and Adolescents-IV (DICA-IV) or for more recent participants, the Kiddie Schedule for Affective Disorders and Schizophrenia for School-Aged Children - Lifetime Version, revealed a history of a developmental or psychiatric disorder, with the exception of simple phobias, or if they scored above clinical cut-offs on the parent and teacher Conners' and ADHD Rating Scales. Available phenotypic data varied according to the study in which participants originally enrolled. Core ASD and ADHD symptomatology were quantified using parent responses to the Social Responsiveness Scale (SRS) questionnaire[5]and the DuPaul ADHD Rating Scale[17], respectively. The SRS askes a caregiver to rate a child's motivation to engage in social interactions, his/her ability to recognize emotional and interpersonal cues from other people, to interpret those cues correctly, and to respond to what he/she interprets appropriately. Higher SRS scores indicate more severe social deficits. The DuPaul ADHD Rating Scale askes a caregiver to rate the severity of inattention and hyperactivity/impulsivity symptoms over the last six months and yields two subdomain scores: inattention and hyperactivity/impulsivity. Higher DuPaul scores indicate more severe symptoms. In

addition to ASD and ADHD trait severity, basic motor control was examined using the Physical and Neurological Exam for Subtle Signs (PANESS) and intellectual ability using General Ability Index (GAI) derived from the Wechsler Intelligence Scale for Children[25]. PANESS assesses basic motor control by a laterality inventory and a detailed examination of subtle motor deficits. The motor deficits included overflow movements, involuntary movements, and dysrhythmia[11]. Higher PANESS scores indicated worse motor control. GAI is a measure of a child's broad intellectual ability that minimizes the influence of working memory and processing speed as compared to the full-scale intelligence quotient. Higher GAI scores indicated greater intellectual ability.

	TD (N=338)	ADHD (N=273)	ASD (N=147)	Total (N=758)	P value
Sex					0.0021[1]
F	109 (32.2%)	71 (26.0%)	25 (17.0%)	205 (27.0%)	
M	229 (67.8%)	202 (74.0%)	122 (83.0%)	553 (73.0%)	
Age (years)					0.6176[2]
Mean (SD)	10.34 (1.22)	10.32 (1.42)	10.41 (1.39)	10.35 (1.33)	
Range	8.02 - 12.97	8.02 - 13.00	8.01 - 12.99	8.01 - 13.00	
PANESS					<0.0001[2]
N-Miss	6	50	21	77	
Mean (SD)	11.60 (5.82)	16.97 (6.38)	17.39 (7.14)	14.43 (6.84)	
Range	0.00 - 29.00	1.00 - 30.00	2.00 - 31.00	0.00 - 31.00	
SRS					<0.0001[2]
N-Miss	95	148	12	255	
Mean (SD)	17.47 (12.47)	42.83 (18.16)	96.14 (26.27)	44.89 (37.58)	
Range	0.00 - 92.00	8.00 - 91.00	33.00 - 168.00	0.00 - 168.00	
GAI					<0.0001[2]
N-Miss	3	14	6	23	
Mean (SD)	115.47 (12.06)	109.54 (12.85)	106.21 (16.51)	111.60 (13.79)	
Range	83.00 - 157.00	79.00 - 146.00	70.00 - 146.00	70.00 - 157.00	
Inattention					<0.0001[2]
N-Miss	8	10	10	28	
Mean (SD)	2.93 (2.82)	19.02 (4.72)	16.85 (6.27)	11.34 (8.83)	
Range	0.00 - 16.00	3.00 - 27.00	2.00 - 27.00	0.00 - 27.00	
Hyperactivity/Impulsivity					<0.0001[2]
N-Miss	8	10	10	28	
Mean (SD)	2.13 (2.47)	14.25 (6.46)	12.01 (5.94)	8.35 (7.55)	
Range	0.00 - 13.00	0.00 - 27.00	1.00 - 27.00	0.00 - 27.00	

[1] Pearson Chi-Square Test

[2] Kruskal-Wallis rank sum test

Table 1: Summary of cohort deographic and phenotypic characteristics. For continuous variables, mean (standard deviation) and range are summarized; for binary and categorical variables, frequencies and percentages are summarized. Differences among each diagnostic group were tested using either the Kruskal-Wallis rank-sum test or Chi-square test.

2.2 Imaging Data

Each participant completed a mock scanning session to acclimate to the MRI environment and at least one rsfMRI scan on a Philips 3T Achieva (Philips Healthcare, Best, the Netherlands) using a single-shot, partially parallel (SENSE) gradient recalled echo planar sequence (TR/TE = 2500 ms/30ms, FA = 70, 3-mm axial slices

with no slice gap, 128-162 time points). Scans were visually inspected for artifacts; head motion was estimated using rigid body realignment (SPM12), and framewise displacement (FD) was calculated from the realignment parameters[19].

Resting-state fMRI data were processed using a comprehensive pipeline that utilized Statistical Parametric Mapping software (SPM12, Wellcome Trust Centre for Neuroimaging, Department of Cognitive Neurology, Cambridge UK) and custom Matlab (Mathworks, Inc.). Scans were slice-time adjusted, rigid-body realigned to adjust for head motion, and warped to the Montreal Neurological Institute stereotaxic space. Each rsfMRI scan was temporally detrended on a voxel-wise basis. The aCompCor strategy was used to estimate spatially coherent nuisance components from tissues not expected to demonstrate signals relevant to brain activity, as this method has been shown to selectively attenuate physiological variation in the fMRI signal and motion artifact. The aCompCor components were regressed from the data along with detrended realignment estimates and their first derivatives. A bandpass filter was applied (.01-.1Hz pass band) and the data were spatially smoothed (6-mm FWHM Gaussian kernel).

3 Methods

3.1 Estimates of functional connectivity accounting for selection bias

To investigate how accounting for demographic and phenotypic differences between included and excluded participants would impact functional connectivity differences between diagnostic groups, we calculated DRTMLE estimates of functional connectivity for each diagnostic group under the three motion exclusion conditions separately. Our simulation study also considers an IPWE. Consider using n_{dx} observations of

children with a particular diagnosis. For the ith child, $i = 1, \ldots, n_{dx}$, we observe (A_i, X_i, Y_i), where $A_i = 1$ if the child is included and 0 if excluded, X_i are covariates measured on each child (SRS, inattention, hyperactivitity/impulsivity, PANESS, General Ability Index, age, and sex), and Y_i is functional connectivity. The IPWE is computed separately for each diagnostic group as

$$\hat{\mu}_{dx}^{IPWE} = \frac{1}{n_{dx}} \sum_{i=1}^{n_{dx}} \frac{Y_i 1_{A_i=1}}{\hat{p}_{dx}(A_i = 1|X_i)} \ ,$$

where $1_{A_i=1}$ is an indicator variable equal to one if $A_i = 1$ and 0 otherwise, and $\hat{p}_{dx}(A_i = 1|X_i)$ refers to the estimated propensity score for child i, i.e., the probability of not excluding child i's functional connectivity due to movement. We use the subscript dx throughout to clarify this is estimated separately for each diagnostic group. The DRTMLE estimator is computed as

$$\hat{\mu}_{dx}^{drtmle} = \frac{1}{n_{dx}} \sum_{i=1}^{n_{dx}} \hat{Q}_{dx}(X_i, 1) \ ,$$

where $\hat{Q}$ is an estimate of the outcome regression and $\hat{Q}_{dx}(X_i, 1)$ is the estimated average functional connectivity for a child with covariates X_i and whose data are not excluded due to movement. In the DRTMLE procedure this estimate is designed to satisfy

$$\frac{1}{n_{dx}} \sum_{i=1}^{n_{dx}} \frac{1_{A_i=1}}{\hat{p}_{dx}(A_i = 1|X_i)} \left(Y_i - \hat{Q}_{dx}(X_i, 1) \right) = 0 \ ,$$

which endows it with double-robustness[3]. In the results that follow, results labeled as doubly robust were produced using DRTMLE.

In our analysis, we used an ensemble of machine learning methods to estimate the propensity score and outcome regression.

Two R packages were implemented for the estimation of functional connectivity: (1)

Super Learner[14] and (2) DRTMLE[10]. The super learner is an algorithm that uses cross-validation to build an ensemble of several estimators, where the estimators can include logistic regression and machine learning methods. Super learner has been shown to asymptotically provide a fit to the true regression that is as good as the best of the candidate estimators. To estimate propensity scores, we used polynomial multivariate regression splines (`earth` R package), the marginal probability of missingness, and a generalized additive model (`gam` R package). To estimate the outcome regression, we used polynominal multivariate regression splines, main terms linear regression, linear regression with all two way interactions, random forest (`randomForest` R package), and extreme gradient boosting (`xgboost` R package). These regressions were then used to estimate the mean functional connectivity in each diagnosis group using the `drtmle` R package.

3.2 Simulation description

The purpose of this simulation was to ensure that the causal inference method based on the included scans only could be implemented into our dataset to obtain an appropriate estimate of averaged functional connectivity among the general population. We wanted to observe that the estimation was stable and approximated the true averaged functional connectivity to guarantee that the causal inference method can alleviate the effect of selection bias on the functional connectivity among children with neurodevelopmental diseases and typical developing children.

For this simulation (N=10,000,000), we generated the following design. (1) The child's characteristics (w_i) were independent and identically distributed as N (0, 1); (2) Primary diagnosis (dx_i) as ASD was distributed as binomial and the probability to be diagnosed as ASD was related to characteristic of the children. (3) Both w_i and dx_i impacted the probability of included scans (p_{2i}), and the indicator variable

for included, $included_i$, followed a Bernoulli distribution with probability p_{2i}. (4) The functional connectivity was linearly related to the child's characteristics and the child in which diagnostic group, but was independent of the scan that included or excluded by the gross motion exclusion. Here functional connectivity was assumed as independent of gross motion exclusion, so the estimation of functional connectivity in a pseudo world with all the scans that were included by motion control criterion approximated the functional connectivity in the real overall population regardless of included or excluded by motion control criterion. For this simulation, two causal inference methods were applied for this simulation and compared to each other: (1) IPWE and (2) DRTMLE.

$dx_i \sim Bin(p_{1i})$.

$E[dx = 1|w_i] = p_{1i}$.

$log(\frac{p_{1i}}{1-p_{1i}}) = \beta_{11} + \beta_{12} w_i$ (1).

(We assumed $\beta_{11} = 0$ and $\beta_{21} = 8$).

$included_i \sim Bin(p_{2i})$.

$E[included = 1|w_i, dx_i] = p(included = 1|w_i, dx_i) = p_{2i}$.

$log(\frac{p_{2i}}{1-p_{2i}}) = \beta_{21} + \beta_{22} w_i + \beta_{23} I(dx_i = 1)$ (2).

(We assumed $\beta_{21} = 0$, $\beta_{22} = -2$, and $\beta_{23} = -0.5$)

$y_i = \beta_{31} + \beta_{32} w_i + \beta_{33} I(dx_i = 1) + \epsilon_i$ (3).

(We assumed $\beta_{21} = 0$, $\beta_{22} = 20$, and $\beta_{23} = 1$)

$\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

(Here σ was the standard error of the residual and we assumed $\sigma = 1$).

Initially, we obtained the group difference (δ_1) on functional connectivity between TD and ASD group over the whole dataset:

$\delta_1 = E[Y|dx = 1] - E[Y|dx = 0]$

By law of large number, when the size of dataset was very large, the sample mean of function connectivity could approximate to the expected value of functional connectivity conditionally on diagnosis.

$\hat{\delta}_1 = \frac{\sum_{i:dx_i=1} y_i}{n_1} - \frac{\sum_{i:dx=0} y_i}{n-n_1}$

$n_1 = \sum_n^{i=1} 1(dx_i = 1)$ for all the i that $dx_i = 1$;

Second, We obtained the group difference (δ_2) between TD and ASD group under the another scenario that all the excluded scans are removed.

$\delta_2 = E[Y|dx = 1 \ \& \ included = 1] - E[Y|dx = 0 \ \& \ included = 1]$

Same by law of large number,

$\hat{\delta}_2 = \frac{\sum_{i:dx_i=1 \& usable_i=1} y_i}{n_{21}} - \frac{\sum_{i:dx_i=0 \& usable_i=1} y_i}{n_2-n_{21}}$

$n_2 = \sum_{i=1}^{n} I(included_i = 1)$ for all the i that $usable_i = 1$;

$n_{21} = \sum_{i:included_i=1} I(dx_i = 1)$ for all the i that both $usable_i = 1$ and $dx_i = 1$

TD group ($\hat{\mu}_{usable=1|dx=0}$):

$\hat{\mu}_{included_i=1|dx=0} = \frac{1}{n-n_1} \sum_{i:dx_i=0} \frac{y_i I(included_i=1)}{p_{2i}}$

ASD group ($\hat{\mu}_{usable=1|dx=1}$):

$\hat{\mu}_{included=1|dx=1} = \frac{1}{n_1} \sum_{i:dx_i=1} \frac{y_i I(usable_i=1)}{p_{2i}}$

Then we obtained that the group difference between TD and ASD group by inverse probability weighted estimation.

$\delta_3 = \frac{1}{n_1} \sum_{i:dx=1} \frac{yiI(included_i=1)}{p_{2i}} - \frac{1}{n-n_1} \sum_{i:dx=0} \frac{yiI(included_i=1)}{p_{2i}}$

3.3 Impact of motion exclusion criteria on selection bias

In our study, we considered three levels of exclusion due to head motion. The lenient case was modeled after the head motion exclusion thresholds used in the paper describing the Autism Brain Imaging Data Exchange (ABIDE)[12]: scans were excluded if they had a mean FD greater than 0.77 mm, which was the sample mean for the ABIDE dataset. The moderate case was modeled after common head motion exclusion criteria prior to the discovery of the impact of micromovements on functional connectivity in 2012: scans were excluded if the participant moved more than the nominal size of a voxel between any two frames and had less than 5 minutes of data after removing frames above this threshold. In the strictest case, scans were excluded if mean FD exceeded .2 mm or FD for more than 16% of the frames exceeded .25 mm.[7].

Initially, Pearson's chi-squared test was used to assess whether the proportion of excluded data differed by diagnosis. As mentioned in the description of the data, the availability of phenotypic data varied according to the study in which participants

originally enrolled. Before assessing potential relationships between the probability of being excluded and any of the phenotypic variables, we first assessed patterns of missingness in the phenotypic data. We used Pearson's chi-squared test to compare the proportion of missing data for each phenotypic measure across diagnostic groups.

Next, we examined correlations among phenotypic characteristics to determine whether a single model containing all of the phenotypic characteristics or a separate model for each phenotypic characteristic was necessary. Both generalized linear regression (GLM) and generalized additive models (GAMs) were used to model the relationship between the log odds of exclusion and phenotypic measures, controlling for age and sex. If the effective degree of freedom (EDF) was approximately one, GLM was selected to assess the relationship between the log of odds of exclusion by gross head motion exclusion criteria and phenotypic characteristics controlling by sex and age for the reason that GLM is easier to interpret.

$included_i \sim Bin(p_i)$.

$$log(\tfrac{p_i}{1-p_i}) = \beta_0 + \beta_1 \mathbf{W} + \beta_2 age_{i\ centered} + \beta_3 sex_i \qquad \text{(1–GLM)}$$

$$log(\tfrac{p_i}{1-p_i}) = \beta_0 + \beta_1 f(\mathbf{W}) + \beta_2 age_{i\ centered} + \beta_3 sex_i \qquad \text{(2–GAM)}$$

3.4 Impact of motion exclusion on naive estimates of functional connectivity

We conducted a seed-based correlation analysis to investigate whether previous findings of atypical functional connectivity involving the default mode network (DMN) in ASD and ADHD were influenced by selection bias. Seed-based Correlation Analysis (SCA) is a common method to explore the functional connectivity within the brain

by calculating the connectivity map, which is defined based on a time series of the region of interest (ROI), the correlation of a time series of all the other voxels in the brain. The connectivity map shows the r-score to indicate how well the correlation between the specific seed and all of the other voxels.

Following the procedure used in Di Martino et al.[12], we focused on a hub of the DMN, the posterior cingulate cortex (PCC), and created a spherical 4-mm radius region-of-interest (ROI) centered at the following MNI coordinates: -8, -56, 26. For each scan, an average time-series was extracted for the PCC ROI and a pairwise Pearson correlation was calculated between the average PCC ROI time-series and the time-series of every other voxel in the brain. Pearson correlations were converted to z-scores representing functional connectivity with the PCC using Fisher's transform.

For the subset of participants in our cohort with complete cases (see Table 2), we calculated naive estimates of functional connectivity for the ASD group under the three motion exclusion conditions separately. We chose to focus on the ASD group because the ASD group lost the highest proportion of participants under all three exclusion conditions. Using only scans that passed a given motion control criteria, we used linear regression to estimate naive functional connectivity differences related to diagnosis while controlling for age, sex, and GAI.

	TD (N=235)	ADHD (N=119)	ASD (N=119)	Total (N=473)	P value
Sex					0.0789[1]
F	68 (28.9%)	39 (32.8%)	24 (20.2%)	131 (27.7%)	
M	167 (71.1%)	80 (67.2%)	95 (79.8%)	342 (72.3%)	
Age (years)					0.0902[2]
Mean (SD)	10.30 (1.15)	10.04 (1.27)	10.30 (1.37)	10.24 (1.24)	
Range	8.02 - 12.90	8.02 - 12.99	8.01 - 12.99	8.01 - 12.99	
SRS					< 0.0001[2]
Mean (SD)	17.45 (12.47)	43.02 (18.28)	96.01 (26.76)	43.65 (37.05)	
Range	0.00 - 92.00	11.00 - 91.00	33.00 - 168.00	0.00 - 168.00	
Inattention					< 0.0001[2]
Mean (SD)	2.83 (2.84)	19.44 (4.16)	16.82 (6.38)	10.52 (8.83)	
Range	0.00 - 11.00	10.00 - 27.00	2.00 - 27.00	0.00 - 27.00	
Hyperactivity/Impulsivity					< 0.0001[1]
Mean (SD)	2.08 (2.50)	13.65 (6.39)	11.94 (6.08)	7.47 (7.19)	
Range	0.00 - 13.00	0.00 - 27.00	1.00 - 27.00	0.00 - 27.00	
PANESS					< 0.0001[2]
Mean (SD)	9.69 (5.21)	14.42 (6.23)	15.42 (6.75)	12.32 (6.45)	
Range	0.00 - 27.00	1.00 - 27.00	2.00 - 29.00	0.00 - 29.00	
GAI					< 0.0001[2]
Mean (SD)	116.43 (12.53)	112.26 (13.36)	106.67 (16.32)	112.93 (14.32)	
Range	85.00 - 157.00	83.00 - 146.00	70.00 - 146.00	70.00 - 157.00	
Lenient Motion Control					< 0.0001[1]
Excluded	20 (8.5%)	25 (21.0%)	34 (28.6%)	79 (16.7%)	
Included	215 (91.5%)	94 (79.0%)	85 (71.4%)	394 (83.3%)	
Moderate Motion Control					< 0.0001[1]
Excluded	35 (14.9%)	35 (29.4%)	44 (37.0%)	114 (24.1%)	
Included	200 (85.1%)	84 (70.6%)	75 (63.0%)	359 (75.9%)	
Strict Motion Control					< 0.0001[1]
Excluded	127 (54.0%)	89 (74.8%)	94 (79.0%)	310 (65.5%)	
Included	108 (46.0%)	30 (25.2%)	25 (21.0%)	163 (34.5%)	

[1] Pearson Chi-Square Test

[2] Kruskal-Wallis rank sum test

Table 2: Participant characteristics by diagnostic group limited to participants with complete phenotypic records

4 Results

4.1 Simulations Results

As seen in Table 3, using one large dataset containing both included/excluded images, we observed that the averaged difference between the typical developing group and

ASD group using the only included functional connectivity data was very different from the averaged difference in functional connectivity between the typical developing group and ASD group using the included/excluded data. This difference represents the biases if analyses are conducted on only the included scans.

Seed	123	345	456	12345	54321
Group Difference	32.14	32.14	32.15	32.14	32.14
Group Difference Based on Included Scans	25.91	25.90	25.92	25.92	25.93
Estimated Group Difference by IPWE	32.14	31.99	32.24	32.08	31.99

Table 3: Simulation Results (N=10,000,000)

The IPWE of group difference of functional connectivity between TD and the ASD group based on included scans was found to be stably approximate to the average difference on functional connectivity from the included/excluded dataset. When sampling on the 300 observations from the dataset, the DRTMLE of group difference on functional connectivity between TD and ASD group was found to be more stable and closer to the group difference on functional connectivity regardless of whether the scans are included or excluded than was IPWE in Table 4. Depending on the our real sample size, doubly-robust estimator was more stable and closer to the true value than IPWE.

Seed	123	345	456	12345	54321
Group Difference	30.18	32.87	32.09	33.66	32.57
Group Difference Based on Included Scans	24.69	26.37	26.89	27.45	24.31
Estimated Group Difference by IPWE	23.98	25.58	38.03	42.90	19.98
Estimated Group Difference by DRTMLE	27.98	30.41	31.72	32.75	28.60

Table 4: Simulation Results (N=300)

4.2 Impact of motion exclusion criteria on selection bias

As seen in Table 5, of the 758 children in our cohort, 616 had at least one scan that passed the most lenient threshold for head motion (ABIDE); 555 had a least one scan that passed our moderate head motion criteria, and only 254 had at least one

scan that passed the strictest head motion exclusion criteria. We observed that the proportion of scans excluded differed by diagnosis using the lenient (χ^2=30.9, df =2, p< 10^{-06}), moderate (χ^2=24.7, df =2, p< 10^{-05}), and strict motion criteria (χ^2=40.9, df =2, p< 10^{-08}). Children with ASD or ADHD were more likely to be excluded than TD children (all post hoc p < .001 corrected). The proportion of excluded scans did not differ between the ASD and ADHD groups.

	TD (N=338)	ADHD (N=273)	ASD (N=147)	Total (N=758)	P value
Lenient Motion Control					< 0.0001[1]
Excluded	34 (10.1%)	67 (24.5%)	41 (27.9%)	142 (18.7%)	
Included	304 (89.9%)	206 (75.5%)	106 (72.1%)	616 (81.3%)	
Moderate Motion Control					< 0.0001[1]
Excluded	61 (18.0%)	88 (32.2%)	54 (36.7%)	203 (26.8%)	
Included	277 (82.0%)	185 (67.8%)	93 (63.3%)	555 (73.2%)	
Strict Motion Control					< 0.0001[1]
Excluded	184 (54.4%)	203 (74.4%)	117 (79.6%)	504 (66.5%)	
Included	154 (45.6%)	70 (25.6%)	30 (20.4%)	254 (33.5%)	

[1] Pearson Chi-Square Test

Table 5: Scan Inclusion/Exclusion By Diagnosis Group

As can be seen in Table 6, missingness in ADHD rating scales was not dependent on diagnostic group. However, we observed that missingness in SRS, PANESS and GAI were not independent of diagnostic group (p-value<.01). The highest proportion of children missing each of these measures was in the ADHD group. The current study is pooling participants from multiple studies with various scientific aims. The earliest of these studies did not collect SRS or PANESS data from children with ADHD, and likewise, the case-matched TD children. The current analysis assumes that the sample of TD and ADHD with SRS, PANESS, and GAI data is representative of the population of TD and ADHD, i.e., if we were to randomly sample TD children, then the distribution of SRS, PANESS, and GAI would be equal to the distribution in the current study, and similarly for ADHD.

	TD	ADHD	Autism	Total	p value
SRS					< 0.0001[1]
Missing Value (N)	95	148	12	255	
Inattention					0.8669[1]
Missing Value (N)	8	10	10	28	
Hyperactivity/Impulsivity					0.8669[1]
Missing Value (N)	8	10	10	28	
PANESS					< 0.0001[1]
Missing Value (N)	6	50	21	77	
GAI					0.0147[1]
Missing Value (N)	3	14	6	23	

[1] Pearson Chi-Square Test

Table 6: Missingness in Phenotypic Variables

Figure 1 illustrates that our phenotypic variables were all correlated, which would complicate the interpretation of effects estimated using a single model containing all of the phenotypic variables. To improve model interpretability, we created separate logistic regression models for each phenotypic variable while covarying for age and sex.

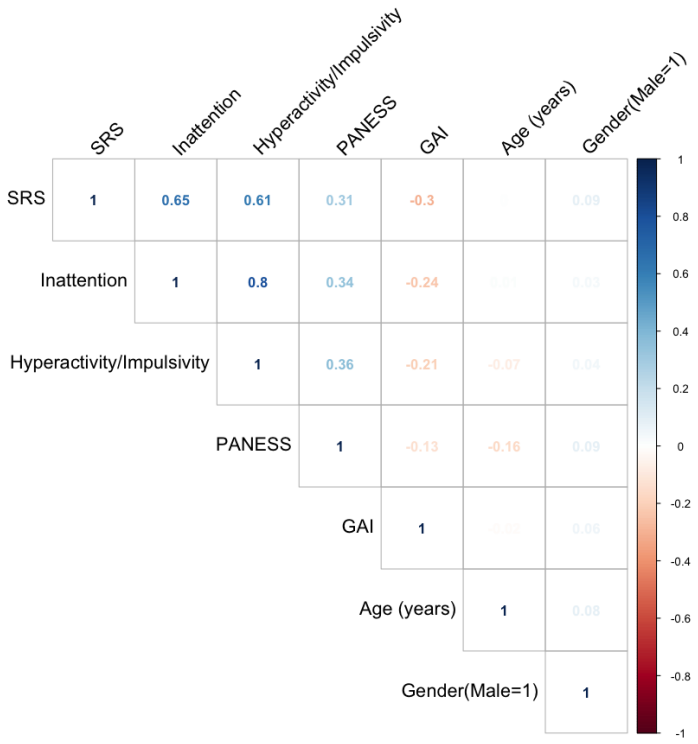
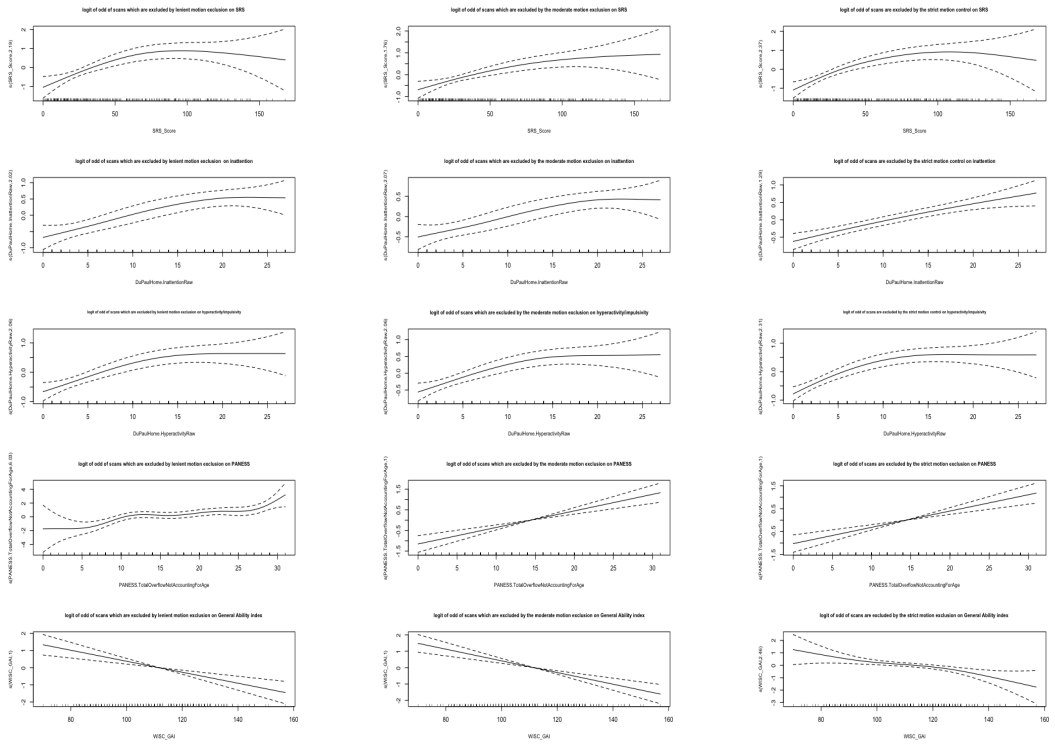


Figure 1: The Pearson correlation of children's phenotypic and demographic characteristics (from left to right): Social Responsiveness Scale (SRS), inattention, hyperactivity/impulsivity, PANESS, General Ability Index, age and sex (male as control)

As seen in Table 7 and Figure 2, we observed that the there are some estimated degree of freedom (EDF) that are equal to 1. That means that a linear relationship between log odds of exclusion using the motion criteria on phenotypic information controlling by age and sex. Because that GLM is easier to be explanation than GAM is, so when EDF=1, the selected model is GLM on relationship between log of odds of exclusion and the phenotypic characteristic in this study.

Phenotype	Lenient	Moderate	Strict
SRS	2.19	1.76	2.37
Inattention	2.02	2.07	1.29
Hyperactivity/Impulsivity	2.06	2.06	2.31
PANESS	6.03	1.00	1.00
General Index Ability	1.00	1.00	2.46

Table 7: Estimated degrees of freedom (EDF) of Social Responsiveness Scale (SRS), Inattention, hyperactivity/impulsivity, PANESS, General Ability Index using univariate generalized additive model controlling by age and sex



(a) Lenient motion exclusion

(b) moderate motion exclusion

(c) Strict motion exclusion

Figure 2: The relationship between the log odds of excluded scans of (from top to bottom) Social Responsiveness Scale (SRS), inattention, hyperactivity/impulsivity, PANESS, and General Ability Index using three levels head motion exclusion criteria using GAM . [a, Left]: Lenient motion exclusion, [b, Middle]: Moderate motion exclusion, [c, Right]: Strict motion exclusion

We observed that regardless of the level of head motion control used to assess the

scan motion quality, children whose scans were excluded, on average, had higher SRS (more severe social deficit), more overflow errors (worse motion control), were more inattentive, hyperactive, and impulsive, and had worse intellectual ability than children whose scans were included by motion control (Figure 3). Therefore, GLM or GAMs were used to model the relationship between the log odds of exclusion and each phenotypic measures, controlling for age and sex. All phenotypic measures were significantly related to the probability of exclusion using both levels of motion control (Figure 3). Children with more severe social deficits, inattentive symptoms, hyperactive/impulsive symptoms, or poorer motor control were more likely to be excluded, while children with higher GAI were less likely to be excluded (all $p < .001$).

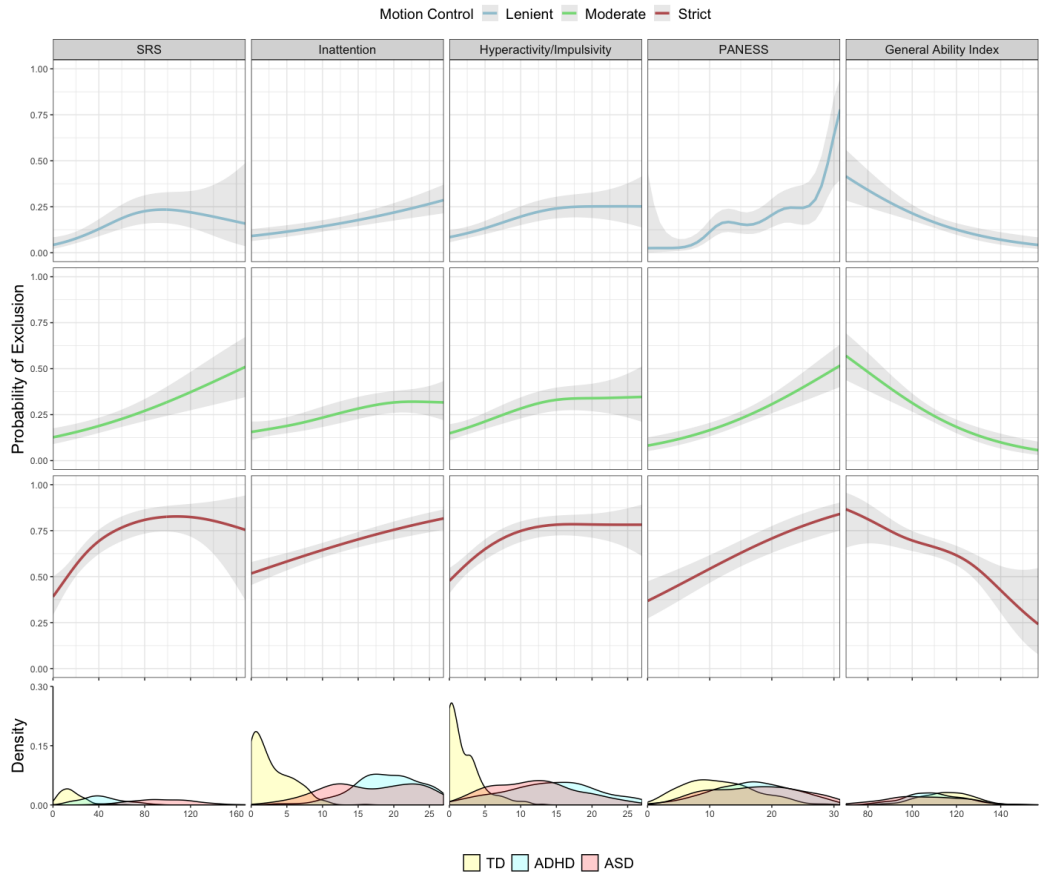
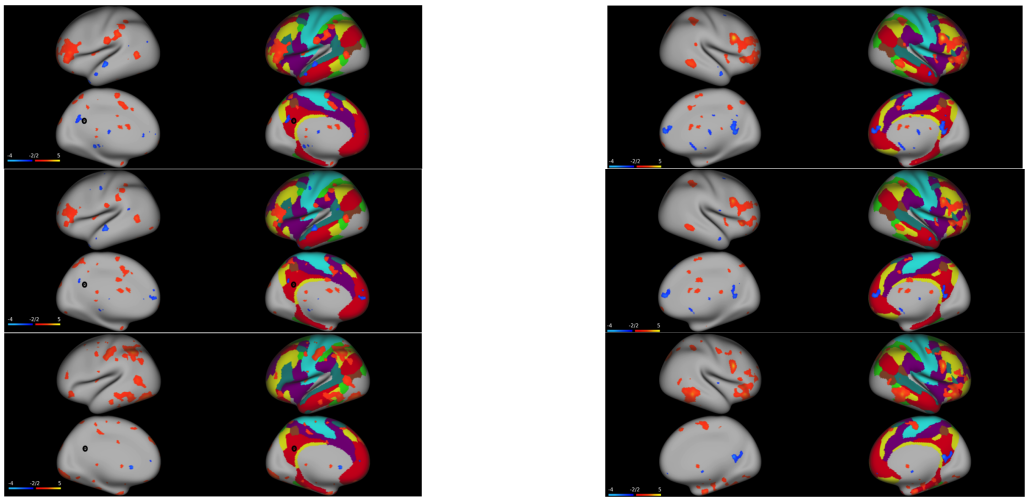


Figure 3: Exclusion probability as a function of (from left to right) Social Responsiveness Scale (SRS), inattention, hyperactivity/impulsivity, PANESS, and General Ability Index using the lenient (top row, blue lines), the moderate (second row, green lines) and strict motion exclusion (bottom row, red lines) , controlling for age and sex. Confidence intervals are shaded grey. Phenotypic distributions are displayed across the bottom panel and colored by diagnostic group (typical developing [TD]: yellow, Attention Deficit Hyperactivity Disorder [ADHD]: blue, Autism Spectrum Disorder [ASD]: red).

4.3 Impact of motion exclusion on naive estimates of functional connectivity

As previously mentioned, the availability of phenotypic data varied according to the study in which participants originally enrolled. When estimating DMN functional connectivity while accounting for selection bias, we had to limit our sample to participants with complete phenotypic records. Table 2 summarizes the subset of our

cohort with complete phenotypic records. We observed that for the children in the ASD group, the proportion of the excluded scans is always highest for all exclusion criteria. Among the three motion control criteria, most scans (79.0%) in the ASD group are excluded by the strict head motion exclusion criterion. Under this situation, limited observations could lead to improper estimation and low statistical power. In Figure 4, we obtained similar estimates ASD-related hypo- and hyper-connectivity under the lenient and moderate motion control. Nevertheless, when the number of included scans are very limited (N=25), the estimated ASD-related hypo- and hyper-connectivity by linear regression is very different from the lenient and moderate criteria.



(a) Left hemisphere

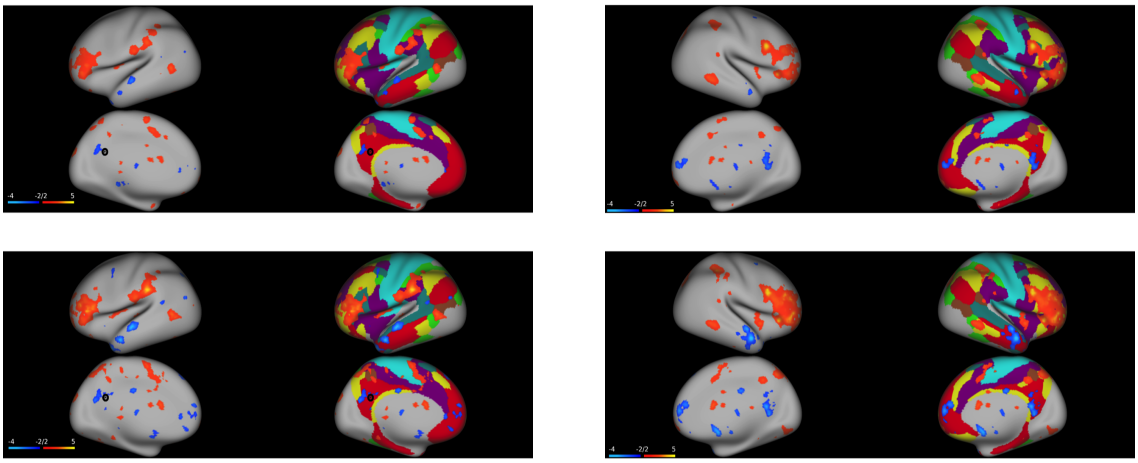
(b) Right hemisphere

Figure 4: Seed-based correlation analyses for a seed region located in left posterior cingulate cortex (L-PCC, x=-8, y=-56, z=26, (a) Left hemisphere's Z maps of the group differences between individuals with ASD and TD group across three level exclusion.[Top row]: Lenient motion control, [Second row]: Moderate motion control, [Bottom row]: Strict motion control. [Orange]: ASD > TD, [Blue]: ASD < TD (Z > 2, uncorrected), [Magenta]: Cingulo-Opercular, [Red]: Default, [Yellow]: Frontoparietal, [Light Blue]: Language, [Brown]: Posterior-Multimodal, [Aqua]: Somatomotor, [Green]: Dorsal-attention.

4.4 Estimates of functional connectivity accounting for selection bias

4.4.1 ASD versus TD

First, as seen in Figure 5 in the thresholded image ($Z > 2$), based on the lenient motion control criterion, we found little evidence of ASD-related hypoconnectivity (ASD $<$ TD). Interestingly, the DRTMLE estimates indicate hypoconnectivity within the default mode network, including the temporal lobe, right PCC, and right pre-frontal cortex (PFC). Regions of hyperconnectivity (ASD $>$ TD) with the PCC were found in frontoparietal, cingulo-opercular network, and dorsal-attention network. These regions are apparent in the naive estimates but more pronounced in the DRTMLE estimates. In addition, we compared linear regression estimation (top) and doubly-robust estimation (bottom) and then found that doubly-robust estimation on ASD-related hypo- and hyper-connectivity was more significant than estimation by linear regression on ASD-related hypo- and hyper-connectivity.

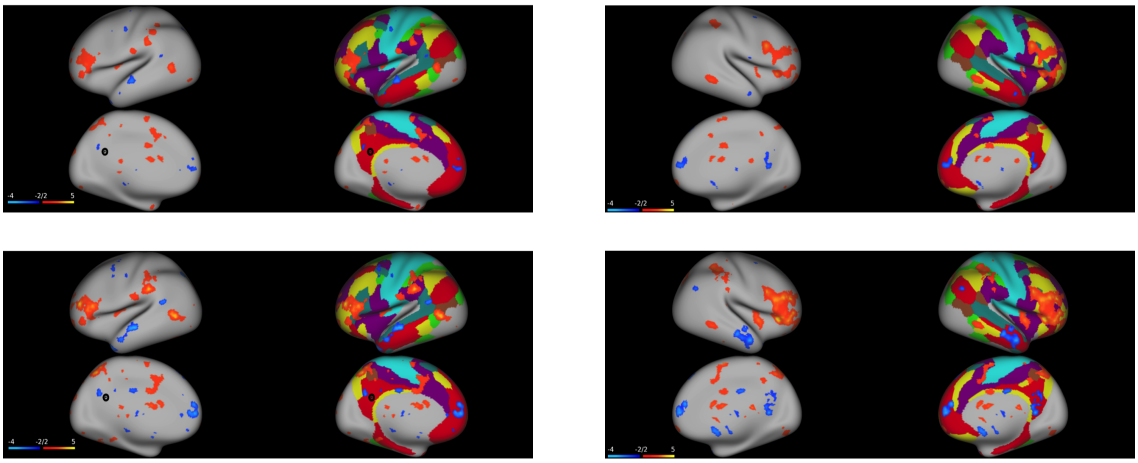


(a) Left hemisphere

(b) Right hemisphere

Figure 5: Seed-based correlation analyses based on the lenient head motion exclusion criterion referred Di Martino *et al* study for one core seed regions located in left posterior cingulate cortex (L-PCC,x=-8, y=-56, z=26), (a) Left hemisphere's Z maps of the group differences between individuals with ASD and TD group.(b) Right hemisphere's Z maps of the group differences between individuals with ASD and TD group.[Top]:linear regression estimation, [Bottom]:Doubly Robust estimation
L-PCC is depicted as black dots on the very inflate surface maps. [Orange]: ASD $>$ TD, [Blue]: ASD $<$ TD ($Z > 2$, uncorrected), [Magenta]:Cingulo-Opercular, [Red]: Default, [Yellow]: Frontoparietal, [Light Blue]: Language,[Brown]: Posterior-MultiModal, [Aqua]:Somatomotor, [Green]:Dorsal-attention

Second, as seen in Figure 6, on voxel level ($Z > 2$, uncorrected), based on the moderate head motion exclusion criterion, we found that similar ASD-related hypoconnectivity and hyperconnectivity located on same network regions shown as in Figure 5 by the moderate head motion exclusion criterion. We also found that DRTMLE on ASD-related hypo- and hyper-connectivity was more significant than estimation by linear regression on ASD-related hypo- and hyper-connectivity based on motion control criterion referred by Di Martino *et al*.

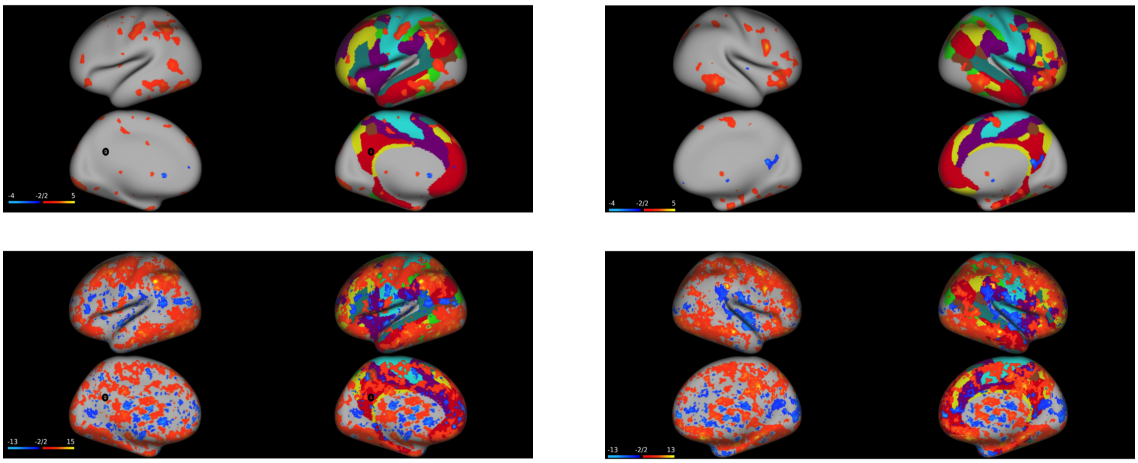


(a) Left hemisphere

(b) Right hemisphere

Figure 6: Seed-based correlation analyses based on the moderate motion control criterion for one core seed regions located in left posterior cingulate cortex (L-PCC,x=-8, y=-56, z=26), (a) Left hemisphere's Z maps of the group differences between individuals with ASD and TD group.(b) Right hemisphere's Z maps of the group differences between individuals with ASD and TD group.[Top]:linear regression estimation, [Bottom]:Doubly Robust estimation
L-PCC is depicted as black dots on the very inflate surface maps. [Orange]: ASD > TD, [Blue]: ASD < TD (Z > 2, uncorrected), [Magenta]:Cingulo-Opercular, [Red]: Default, [Yellow]: Frontoparietal, [Light Blue]: Language,[Brown]: Posterior-MultiModal, [Aqua]:Somatomotor. [Green]:Dorsal-attention

Finally, due to the sample size of included images by the strict head motion exclusion criterion is very limited (ASD: N=25), the Bottom Figure 7, on the voxel level (Z>2, uncorrected), the estimation of ASD-related hypo-connectivity and hyper-connectivity are distributed to the whole brain.



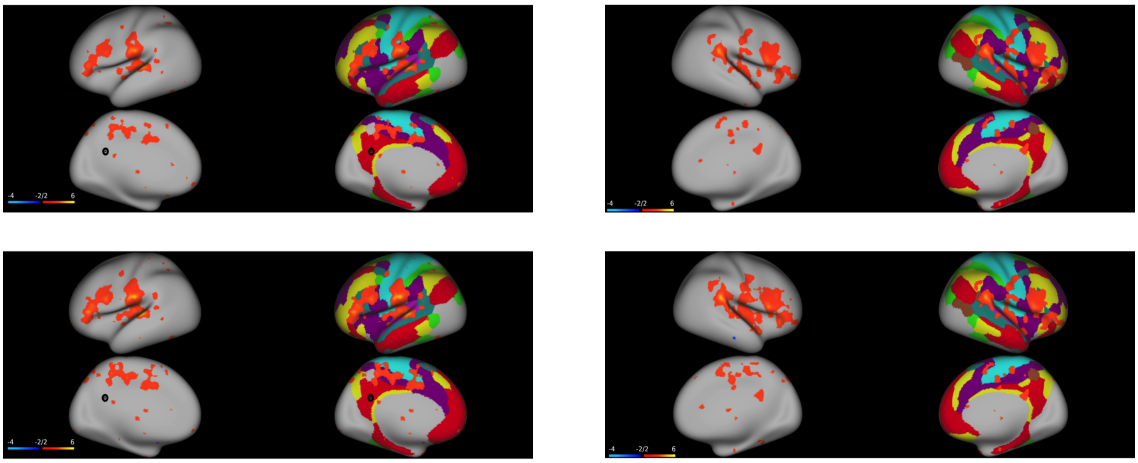
(a) Left hemisphere

(b) Right hemisphere

Figure 7: Seed-based correlation analyses based on the strict head motion exclusion criterion for one core seed regions located in left posterior cingulate cortex (L-PCC,x=-8, y=-56, z=26), (a) Left hemisphere's Z maps of the group differences between individuals with ASD and TD group.(b) Right hemisphere's Z maps of the group differences between individuals with ASD and TD group. [Top]:linear regression estimation, [Bottom]:Doubly Robust estimation
L-PCC is depicted as black dots on the very inflate surface maps. [Orange]: ASD > TD, [Blue]: ASD < TD (Z > 2, uncorrected), [Magenta]:Cingulo-Opercular, [Red]: Default, [Yellow]: Frontoparietal, [Light Blue]: Language,[Brown]: Posterior-MultiModal, [Aqua]:Somatomotor,[Green]:Dorsal-attention

4.4.2 ADHD versus TD

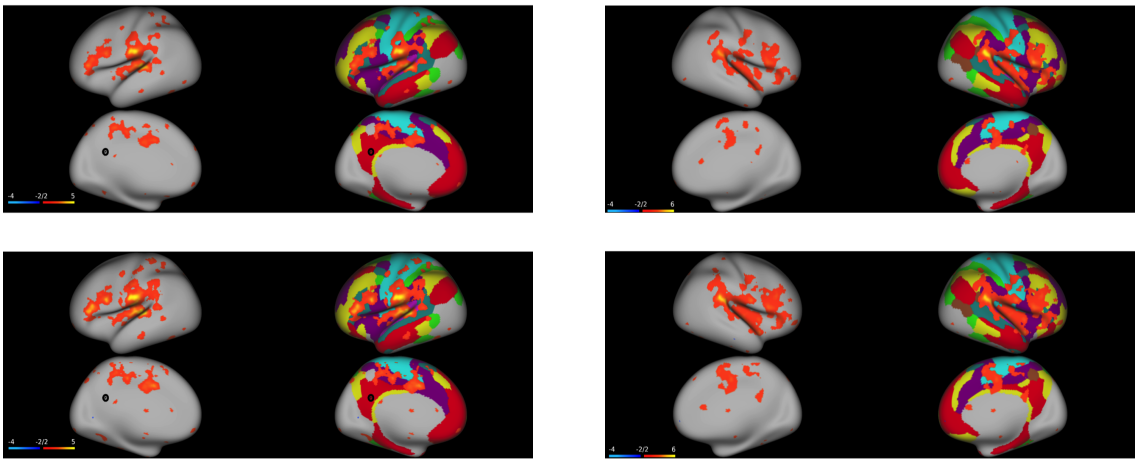
As seen in Figure 8, on the voxel level, we only found that ADHD-related hyper-connctivity (ASD > TD) (i.e Cingulo-Opercular network, Somatomotor, and Dorsal-Attention network, and parts of the brain network for language) for the core seed regions were located in the left posterior cingulate cortex (L-PCC) based on the motion control criterion of Di Martino *et al* study. In addition, we compared linear regression estimation (top) and doubly-robust estimation (bottom) and again we found that DRTMLE on ADHD-related hyperconnectivity was more significant than estimation by linear regression on ADHD-related hyperconnectivity mainly on the region of Language and Cingulo-Opercular network.



(a) Left hemisphere (b) Right hemisphere

Figure 8: Seed-based correlation analyses based on the lenient head motion exclusion criterion for one core seed regions located in left posterior cingulate cortex (L-PCC,x=-8, y=-56, z=26), (a) Left hemisphere's Z maps of the group differences between individuals with ADHD and TD group.(b) Right hemisphere's Z maps of the group differences between individuals with ADHD and TD group. [Top]:linear regression estimation, [Bottom]:Doubly Robust estimation
L-PCC is depicted as black dots on the very inflate surface maps. [Orange]: ADHD > TD, [Blue]: ADHD<TD (Z > 2, uncorrected), [Magenta]:Cingulo-Opercular, [Red]: Default, [Yellow]: Frontoparietal, [Light Blue]: Language,[Brown]: Posterior-MultiModal, [Aqua]:Somatomotor, [Green]:Dorsal-attention

Second, as seen in Figure 9, on voxel level(Z>2, uncorrected), we also found that similar ADHD-related hyperconnectivity located on same network regions shown as in Figure 8 by the lenient motion exclusion criterion. We did not found the significant difference between DRTMLE on ADHD-related hyperconnectivity and the estimation by linear regression on ADHD-related hypo- and hyper-connectivity based on the motion control criterion referred by Di Martino *et al.*study

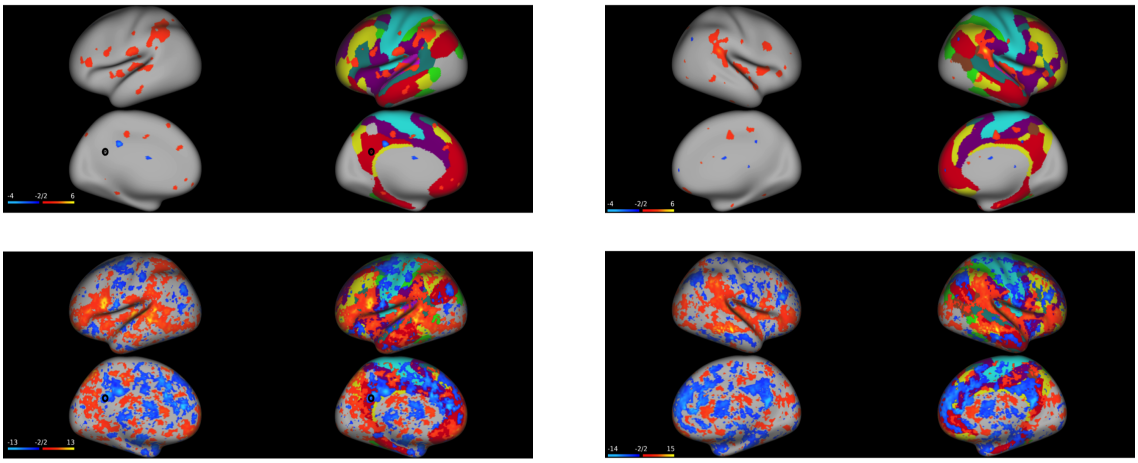


(a) Left hemisphere

(b) Right hemisphere

Figure 9: Seed-based correlation analyses based on the moderate head motion exclusion criterion for one core seed regions located in left posterior cingulate cortex (L-PCC,x=-8, y=-56, z=26), (a) Left hemisphere's Z maps of the group differences between individuals with ADHD and TD group.(b) Right hemisphere's Z maps of the group differences between individuals with ADHD and TD group. [Top]:linear regression estimation, [Bottom]:Doubly Robust estimation
L-PCC is depicted as black dots on the very inflate surface maps. [Orange]: ADHD $>$ TD, [Blue]: ADHD $<$ TD ($Z > 2$, uncorrected), [Magenta]:Cingulo-Opercular, [Red]: Default, [Yellow]: Frontoparietal, [Light Blue]: Language,[Brown]: Posterior-MultiModal, [Aqua]:Somatomotor, [Green]:Dorsal-attention

Finally, same as ASD versus TD, the number of included image by the strict head motion exclusion criterion is also limited (ADHD: N=30). As seen in Figure 10, Z map of ADHD-related hypo- and hyper-connectivity are all distributed to the whole brain.



(a) Left hemisphere

(b) Right hemisphere

Figure 10: Seed-based correlation analyses based on the strict head motion exclusion criterion for one core seed regions located in left posterior cingulate cortex (L-PCC, x=-8, y=-56, z=26), (a) Left hemisphere's Z maps of the group differences between individuals with ADHD and TD group.(b) Right hemisphere's Z maps of the group differences between individuals with ADHD and TD group.[Top]:linear regression estimation, [Bottom]:Doubly Robust estimation
L-PCC is depicted as black dots on the very inflate surface maps. [Orange]: ADHD $>$ TD, [Blue]: ADHD $<$ TD ($Z > 2$, uncorrected), [Magenta]:Cingulo-Opercular, [Red]: Default, [Yellow]: Frontoparietal, [Light Blue]: Language,[Brown]: Posterior-MultiModal, [Aqua]:Somatomotor,[Green]:Dorsal-attention.

5 Discussion

The purpose of this study was to find whether selection bias introduced by motion control criterion leads to mischaracterization of diagnostic group differences on functional connectivity. Using Doubly-Robust Nonparametric Estimation and Inference, we found evidence of a selection bias due to motion control criteria and that this can be minimized using DRTMLE when there is sufficient sample size of included images.

In our study, we found that at the voxel-level, the Z-map of ASD-related hypo-connectivity (ASD $<$ TD) for the core seed regions on the L-PCC mainly located in Default network and ASD-related hyperconnectivity(ASD $>$ TD) for the core seed

regions on the L-PCC mainly located in Cingulo-Opercular, and Frontoparietal network. This finding contrasts with Di Martino *et al* study where they found a more significant Z-map. The difference in these results could be attributed to target populations analyzed and the handling of covariate analysis. In our study, our target population was different from the population in Di Martino *et al* study. In our study only 8-13 year old school-aged children were enrolled, however, in Di Martino *et al.* study, a few sites enrolled participants spanning childhood to adults (7-64 years old). Macleod *et al* indicated that the neurological diseases may present later during adolescent development (15 or 16 years old) rather than be evident among younger children[16]. Therefore, the different age groups analyzed could explain why the difference on functional connectivity between ASD and TD group in our study (mean age: 10 years old) is not as significant as the difference on functional connectivity between ASD and TD in Di Martino's study(mean age: 13 years old). Secondly, the difference between our study and Di Martino *et al.* study might be attributed to the fact that we excluded the mean framewise displacement as a covariate for the group-level analysis, whereas they included this covariate in their analysis. Specifically, we did not account for the mean framewise displacement in our moderate motion control and strict motion control. In contrast, Di Martino *et al* used a more moderate criterion for motion control (mean frame-wise displacement (FD) <2 s.d. above the sample mean, 0.77mm). They also used the mean framewise displacement as a criterion for functional connectivity for selecting high quality images. In addition, compared with the estimation of functional connectivity using general linear regression, we found that doubly-robust estimation is better to detect the group difference on functional connectivity. This is because the doubly robust method uses all of the observation to do the estimation, however, linear regression only uses the observations whose imaging data is included by motion control criterion.

There are several strengths to this study. To our knowledge this is the first analysis

which has implemented the Doubly-Robust Nonparametric Estimation and Inference in the imaging field. This method also increases the power to detect the group difference between ASD and TD, and between ADHD and TD when the sample size is suitable. Notably, we expanded our analysis relative to Di Martino *et al* and included the ADHD and TD groups and found that signficant ADHD-related hyperconnectivity for the core seed regions on the L-PCC mainly located on Cingulo-Opercular, Lanuage, and Somatomotor network.

There are several limitations to this study. One limitation is that doubly robust estimation is not a good choice when the number of images included by the motion control criterion is relatively low (<30). The estimated variance of the DRTMLE may be anti-conservative in such small samples, which leads to a large Z statistic in many voxels. Therefore, the DRTMLE estimation may not be suitable for settings where strict motion control is required. Thus, in the future, the method that can be used with small sample sizes will improve analysis on functional connectivity based on the included scans by strict motion control. Another major limitation of this study is that we do not correct for multiple comparisons. An important avenue for future research is to correct for multiple comparisons when using DRTMLE.

In conclusion, the findings from this study can be used to assist with understanding neurodevelopmental disorders by using causal inference to address selection bias in imaging data introduced by head motion exclusion criteria.

References

[1] American Psychiatric Association. *American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Publishing, 2013. ISBN 9780890425541. doi: 10.1176/appi.books.9780890425596.744053.

[2] Peter C. Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 2011. ISSN 00273171. doi: 10.1080/00273171.2011.568786.

[3] D. Benkeser, M. Carone, M. J.Van Der Laan, and P. B. Gilbert. Doubly robust nonparametric inference on the average treatment effect. *Biometrika*, 104(4): 863–880, 12 2017. doi: 10.1093/biomet/asx053.

[4] David Benkeser. drtmle: Doubly-Robust Inference in R. Technical report, Emory University.

[5] Sven Bölte, Fritz Poustka, and John N. Constantino. Assessing autistic traits: Cross-cultural validation of the social responsiveness scale (SRS), 2008.

[6] Deborah L. Christensen, Jon Baio, Kim Van Naarden Braun, Deborah Bilder, Jane Charles, John N. Constantino, Julie Daniels, Maureen S. Durkin, Robert T. Fitzgerald, Margaret Kurzius-Spencer, Li Ching Lee, Sydney Pettygrove, Cordelia Robinson, Eldon Schulz, Chris Wells, Martha S. Wingate, Walter Zahorodny, and Marshalyn Yeargin-Allsopp. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance Summaries*, 2016. ISSN 15458636. doi: 10.15585/mmwr.ss6503a1.

[7] Rastko Ciric, Daniel H. Wolf, Jonathan D. Power, David R. Roalf, Graham L. Baum, Kosha Ruparel, Russell T. Shinohara, Mark A. Elliott, Simon B. Eick-

hoff, Christos Davatzikos, Ruben C. Gur, Raquel E. Gur, Danielle S. Bassett, and Theodore D. Satterthwaite. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, 154:174–187, 7 2017. ISSN 10959572. doi: 10.1016/j.neuroimage.2017.03.020.

[8] Baptiste Couvy-Duchesne, Jane L. Ebejer, Nathan A. Gillespie, David L. Duffy, Ian B. Hickie, Paul M. Thompson, Nicholas G. Martin, Greig I. De Zubicaray, Katie L. McMahon, Sarah E. Medland, and Margaret J. Wright. Head motion and Inattention/Hyperactivity share common genetic influences: Implications for fMRI studies of ADHD. *PLOS ONE*, 11(1), 1 2016. doi: 10.1371/journal.pone.0146271.

[9] Melissa L. Danielson, Rebecca H. Bitsko, Reem M. Ghandour, Joseph R. Holbrook, Michael D. Kogan, and Stephen J. Blumberg. Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016. *Journal of Clinical Child and Adolescent Psychology*, 2018. ISSN 15374416. doi: 10.1080/15374416.2017.1417860.

[10] David Benkeser. Doubly-Robust Nonparametric Estimation and Inference. Technical report, 2020. URL `https://github.com/benkeser/drtmle`.

[11] M.B. Denckla. Revised Neurological Examination for Subtle Signs (1985). *Psychopharmacology Bulletin*, 21(4):773–779, 1985.

[12] A. Di Martino, C. G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, B. Deen, S. Delmonte, I. Dinstein, B. Ertl-Wagner, D. A. Fair, L. Gallagher, D. P. Kennedy, C. L. Keown, C. Keysers, J. E. Lainhart, C. Lord, B. Luna, V. Menon, N. J. Minshew, C. S. Monk, S. Mueller, R. A. Müller, M. B. Nebel, J. T. Nigg, K. O'Hearn,

K. A. Pelphrey, S. J. Peltier, J. D. Rudie, S. Sunaert, M. Thioux, J. M. Tyszka, L. Q. Uddin, J. S. Verhoeven, N. Wenderoth, J. L. Wiggins, S. H. Mostofsky, and M. P. Milham. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19 (6):659–667, 2014. ISSN 14765578. doi: 10.1038/mp.2013.78.

[13] Ellen Doernberg and Eric Hollander. Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11, 8 2016. ISSN 10928529.

[14] Eric Polley, Erin LeDell, Chris Kennedy, Sam Lendle, and Mark van der Laan. Super Learner Prediction. Technical report, 2019. URL `https://github.com/ecpolley/SuperLearner`.

[15] Scott A Huettel, Allen W Song, and Gregory Mccarthy. *Functional Magnetic Resonance Imaging, Second Edition*. Sinauer Associates, 2008. ISBN 9780878932863.

[16] S. Macleod and R. E. Appleton. Neurological disorders presenting mainly in adolescence, 2 2007.

[17] Danielle Pappas. Test Reviews: ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation, 2006. ISSN 07342829.

[18] Linden Parkes, Ben Fulcher, Murat Yücel, and Alex Fornito. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage*, 171:415–436, 5 2018. ISSN 10959572. doi: 10.1016/j.neuroimage.2017.12.073.

[19] Jonathan D. Power, Kelly A. Barnes, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.10.018.

[20] Jonathan D. Power, Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, 2014. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.08.048.

[21] James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9-12):1393–1512, 1986. doi: 10.1016/0270-0255(86)90088-6.

[22] Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F. Beckmann, Matthew F. Glasser, Ludovica Griffanti, and Stephen M. Smith. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449–468, 4 2014. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.11.046.

[23] Larry R. Squire. *Encyclopedia of Neuroscience*. Elsevier, 2010. ISBN 9780080450469. doi: 10.1016/C2009-1-03742-3.

[24] Koene R.A. van Dijk, Mert R. Sabuncu, and Randy L. Buckner. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.07.044.

[25] Kelli Vaughn-Blount, Steuart T. Watson, Amber L. Kokol, Renee Grizzle, Russell N. Carney, Shannon S. Rich, Sonia LeClere, Maria Stylianou, Xenia Anastassiou-Hadjicharalambous, Mary Joann Lang, David Morrison, Robert Walrath, Daniel Patanella, Chandra Ebanks, Anne Fierro Vanderlaan, Jackie Laney, Seongwon Yun, Shelia M. Kennison, David Ritchie, Barry Nierenberg, Caroline McKnight, Rachel Kitson, Beth Anne Newman, Marie C. McGrath, Anne Fierro Vanderlaan, Heather N. Arduengo, Angela D. Mitchell, Anna

Mazur-Mosiewicz, Raymond S. Dean, Raymond S. Dean, Jennifer R. Vought, Raymond S. Dean, Marci S. DeCaro, and Denise E. Maricle. Wechsler Intelligence Scale for Children, Fourth Edition. In *Encyclopedia of Child Behavior and Development*. Pearson Education, 2011. doi: 10.1007/978-0-387-79061-9{_}3066.

[26] Wolfgang von der Linden, Volker Dose, and Udo von Toussaint. *Bayesian probability theory*. Cambridge University Press, 2010. ISBN 9781139565608. doi: 10.1017/CBO9781139565608.