Response to Reviewers: Accounting for motion in resting-state fMRI: What part of the spectrum are we characterizing in autism spectrum disorder?

Mary Beth Nebel, Daniel E. Lidstone, Liwei Wang, David Benkeser,
Stewart H. Mostofsky, Benjamin B. Risk

We thank the reviewers for their thoughtful comments on our submission, which we believe have resulted in a much clearer presentation of our approach for treating missingness due to motion quality control. Below, we address their concerns point by point. Reviewer comments are indicated in *italics*; our response follows in normal text, and excerpts from the manuscript are in calibri font. In the revised manuscript, modifications are highlighted in **purple text**. Please note that after considering the comment R1.6, we elected to redefine our study sample as complete cases. This decision had a minor impact on all analyses, although our overall message remains the same. The purple coloring in the manuscript includes these changes.

Reviewer 1

Reviewer #1: This is a very interesting manuscript on an important topic (how does restriction of resting-state MR images to low motion cases potentially bias resulting associations?). There is much to like in this paper. The approach is principled and laid out in detail. The application to ASD subjects and controls is apropos to the issues at hand. The critiques below indicate my take on moderate weaknesses in the current version in light of these notable strengths. The critiques below are in this spirit, and many involve more framing and discussion rather than changing the method. Note, questions/clarifications/critiques are ordered in terms of their appearance in the paper, not in terms of importance.

R1.1 *bottom of p3 / top of p4: "In studies with missing data, an estimate of an association may be biased if the data are not missing at random in the sense that the difference in the mean outcome between the groups of interest in the observed data differs from the difference if all data were observed" Is it the mean difference that matters, or rather the strength of the association that differs between observed and missing observations?*

We are concerned about the bias in the mean difference, and we think the use of the word association at this point in the introduction made the sentence unclear. Later in the paragraph, we describe the role of the association between functional connectivity and the covariates. We have rephrased:

In studies in which the outcome is missing for some participants, the difference in the mean between two groups calculated from observed outcomes may be biased if data are not randomly missing (Hernan and Robins, 2020).

R1.2 *Figure 1: E* should be defined in the legend. DAGs also need to be described.*

We have updated the legend of Figure 1 to clarify undefined notation and describe the graphs:

Scan exclusion may induce selection bias. A indicates diagnosis, where $A = 0$ (lighter shading) represents the typically developing group and $A = 1$ (darker shading) represents the autism spectrum disorder (ASD) group. W represents a covariate that reflects symptom severity; Δ indicates resting-state fMRI usability, where Δ=1 is usable and Δ=0 is unusable. Y is the functional connectivity between two brain regions, and $Y(1)$ is the possibly counter fact functional connectivity from a usable fMRI scan. $E^*()$ denotes an expectation with respect to the probability measure of $\{Y(1), A, W\}$. Additional details are in Section 2.3.1. (Left panel) Children with usable resting-state fMRI data (in purple) may systematically differ from all enrolled children (in green). The distribution of symptom severity W differs between children with usable and unusable fMRI data ($W \leftrightarrow \Delta$). W is related to functional connectivity ($W \leftrightarrow Y$). Additionally, there are associational effects between ASD and functional connectivity ($A \leftrightarrow Y$) and between ASD and the symptom severity covariate ($A \leftrightarrow W$). Under these conditions, naïve estimators of group-level functional connectivity based only on participants with usable data may be biased. (Right panel) We propose to address this bias using doubly robust targeted minimum loss based estimation (DRTMLE), which involves three steps. 1. Fit the propensity model. 2. Fit the outcome model, which predicts functional connectivity from the covariates for participants with usable rs-fMRI data. Then use this model to predict functional connectivity for both usable and unusable participants. 3. Apply the DRTMLE algorithm, which uses the inverse probability of usability from step 1 and predictions of functional connectivity for all subjects (usable and unusable) from step 2 to break the pathways $A \leftrightarrow \Delta$, $W \leftrightarrow \Delta$, and $\Delta \rightarrow Y$.

R1.3 *Major point: the paper frames the excluded data as a "causal inference bias". In fact, it's more properly couched as a "selection bias" or "Berksonian bias" (e.g., Modern Epidemiology 4th Ed. p.318). There is ample reason to change terminology on this point. First, the associations discussed are not causal (e.g., autism does not cause connectivity difference, more likely vice versa). Second, the potential outcome Y(1) is never compared to Y(0) (as would be the case in a causal estimand, e.g., Y(1) - Y(0)). Third, calling the models causal confuses with another source of bias discussed in the paper, i.e., that related to the impact of motion on connectivity estimates (which really is related to confounding and hence causal bias). This reviewer thus strongly recommends using "correction for selection bias" rather than "causal models" for the terminology throughout the paper.*

We have revised the Introduction to describe the bias as selection bias, and we have added a summary of how selection bias and confounding bias are related to the fourth paragraph:

> The graph in Figure 1 illustrates how excluding high-motion participants could obscure the relationship between a diagnosis of ASD (A) and functional connectivity (Y) by changing the joint distribution of diagnosis and a covariate related to symptom severity (W). Bias can arise from a lack of exchangeability, as children with usable data differ from those with unusable data. The term "'confounding bias" is sometimes used to describe bias that arises when a predictor and outcome share a common cause (e.g., $\Delta \rightarrow W$ and $\Delta \rightarrow Y$). The term "selection bias" or "Berksonian bias" may be used to describe bias from conditioning on common effects (e.g., $W \rightarrow \Delta$ and $Y \rightarrow \Delta$). These concepts often overlap (Lash et al., 2021). In both cases, the key source of bias is a violation of exchangeability (Hernan and Robins, 2020). In Figure 1, we describe the bias as selection bias, since it can be viewed as originating from a study that selects children that pass motion QC. If autistic children with usable rs-fMRI data are phenotypically more similar to typically developing children than those that were excluded, observed group differences may be reduced relative to group differences if we were able to collect usable rs-fMRI data from all participants.

We agree that ultimately we are not attempting to estimate a causal effect of autism versus typical development on functional connectivity, which would require a contrast between two potential outcomes. Rather, we use the notion of counterfactuals to clarify what our estimand of interest is: an associational measure of a difference between autistic children without an intellectual disability and TD children if, possibly counter fact, movement of all children was within an acceptable range.

We now use the phrase "associational" in the revised Figure 1:

> Additionally, there are associational effects between ASD and functional connectivity ($A \leftrightarrow Y$) and between ASD and the symptom severity covariates ($A \leftrightarrow W$).

At the beginning of Section 2.3.1, we write

> Our goal is to estimate the difference in average functional connectivity between autistic children without an intellectual disability and typically developing children. We will use a causally informed approach to correct this associational estimand for potential selection bias following exclusion due to failed motion QC.

Later in Section 2.3.1, we clarify the counterfactual we are defining:

> We use the potential outcomes notation and let $Y(1)$ denote the functional connectivity in this hypothetical world (Hernan and Robins, 2020). Our counterfactual $Y(1)$ is not functional connectivity if assigned a diagnosis of autism, but rather functional connectivity under an intervention that reduces motion to an acceptable level in all children.

We go on to explain how our estimand differs from an average treatment effect:

> This associational estimand differs from an average treatment effect (ATE) commonly considered in causal inference (Hernan and Robins, 2020), which integrates across the distribution of the covariates for the pooled population (autistic and typically developing

children) to contrast the counterfactual of being assigned a diagnosis of autism with being assigned typically developing.

In addition, we have carefully reworded the Discussion to reiterate that we are proposing an approach to treat our associational target parameter for selection bias:

> We further propose a statistical approach for addressing the data loss and possible selection bias following motion QC using DRTMLE.

In Section 4.5,

> The target parameter used in estimating an average treatment effect in causal inference marginalizes with respect to the distribution of variables pooled across treatments, which would address biases introduced by the first set of variables. Our deconfounded group difference is associational and addresses the biases introduced by the second set of variables.

Then in Section 4.7,

> Selection bias could impact analyses of rs-fMRI data collected from such developmental samples if the sample of included children that are able to lay motionless tend to be more mature than the full sample.

Regarding *Third, calling the models causal confuses with another source of bias discussed in the paper, i.e., that related to the impact of motion on connectivity estimates (which really is related to confounding and hence causal bias)*: We are careful to not use the phrase "causal model" in the paper, and our previous changes clarify that our estimand is associational.

Regarding *This reviewer thus strongly recommends using "correction for selection bias" rather than "causal models" for the terminology throughout the paper*: In addition to the aforementioned changes and use of selection bias in 2.3.1, the phrase "causal models" does not appear in the paper.

R1.4 *p6: "Children were ineligible to participate if their full scale intelligence quotient...was less than 80 and they scored below 65 on 1) the Verbal Comprehension Index and 2) the Perceptual Reasoning Index (WISC-IV) or the Visual Spatial Index and the Fluid Reasoning Index (WISC-V)..." In other words, there is already potentially a selection bias with respect to ASD persons even before eliminating scans due to head motion. This merits discussion: if the results of an analysis are biased, what population are they biased for? If the study sample is already a convenience sample, why is it important to have associations that reflect the sample before selecting on participants with low head motion?*

We apologize for the lack clarity. The target population for the original data-collecting studies was children without an intellectual disability. The goal of these studies was to estimate the association between functional connectivity and a diagnosis of autism separate from any association between functional connectivity and intellectual disability, and it is important to discuss our findings in the context of this goal.

We have revised the text in Section 2.1.1 Study Population to clarify that these restriction criteria were defined by the data-collecting studies.

> The following exclusion criteria were defined by the data-collecting studies.

We have also updated all descriptions of our study sample. In the Abstract:

> We use a study of autism spectrum disorder in children without an intellectual disability to illustrate the problem and the potential solution.

In the Introduction:

> In this study, we first describe our motivating dataset, an aggregation of phenotypic and rs-fMRI data from 173 autistic children without an intellectual disability and 372 typically developing children who participated in one of several neuroimaging studies at Kennedy Krieger Institute (KKI) between 2007 and 2020.

In the Methods:

> Participants included 173 autistic children without an intellectual disability (148 boys) and 372 typically developing children (258 boys).

Then in the Discussion, we include a broader discussion of this target population and its limitations (see also our response to R3.1):

> A goal of the original data-collecting studies was to estimate the association between functional connectivity and a diagnosis of ASD separate from any association between functional connectivity and intellectual disability. A recent cohort-based study of heritability and familial risk indicated that ASD without a co-occurring intellectual disability may have a greater genetic basis than ASD with an intellectual disability (Xie et al., 2020), suggesting that these ASD phenotypes may be etiologically distinct. Approximately 35% of autistic children have an intellectual disability (Maenner et al., 2021), and our findings may not generalize to this segment of the autism population (Reiter et al., 2019). Unless we as a field make a concerted effort to correct for the overrepresentation of cognitively abled children in autism research, we will never be able to answer this question. In addition, an estimated 70% of autistic children have at least one comorbid disorder (Simonoff et al., 2008). The original data-collecting studies allowed for two of the most common comorbid mental health disorders, namely ADHD and social anxiety disorder, but excluded all others. Because research informs the revision of diagnostic criteria and health and educational policy, we must improve recruitment and data collection from children we have historically excluded from rs-fMRI research of autism.

We also discuss the shortcomings of a convenience sample in this section:

> Although DRTMLE can address issues of missing data due to motion QC, it does not address possible biases in the sample of children with behavioral data that appear in the study, which may differ from the target population. The target population in the original data-collecting studies was children without an intellectual disability. Even

within this target population, our study is a convenience sample in that participants meeting the eligibility criteria were recruited using flyers and patient records, rather than randomly sampled. In fact, we had a five-to-one ratio of boys to girls in our ASD sample compared to the estimated four-to-one relative risk nationwide (Maenner et al., 2021). This convenience sampling approach is an important shortcoming in many fMRI studies. A recent study found a healthy volunteer bias in the UK Biobank sample (Fry et al., 2017). Bradley and Nichols (2022) used propensity score weighting methods to decrease this bias in a study of brain structure but noted that these methods require access to external individual-level population data, which is often unavailable. Extending the DRTMLE approach to accommodate situations in which the study sample deviates from the target population is an important area for future work.

R1.5 *p8: "Available phenotypic data varied according to the study in which participants enrolled." Was there a difference in ASD-brain associations across the two studies? This relates to the last point, what is the population that the attempt to create unbiased associations relevant for? This should be discussed.*

We have removed this inaccurate sentence from the manuscript. The data are from multiple studies, where we included typically developing children from studies of ADHD and Tourette Syndrome that occurred concurrently with the ASD studies. The version of the tools used to assess phenotypic characteristics changed over time (ADOS and ADOS-2, SRS and SRS-2, WISC-IV and WISC-V) as the assessments were updated between 2007 and 2020. The target population, 8-13 year-old children without an intellectual disability, was the same across time. Some children participated in more than one study. We always used the highest quality scan but filled in missing phenotypic data from another study when possible to limit missingness. Thus, it would be challenging to compare DRTMLE estimates of the deconfounded group difference across studies.

R1.6 *p10: "The study sample for our application of the deconfounded group difference is defined as the subset of participants with a complete set of demographic information (sex, socioeconomic status, and race) and the selected predictors along with nineteen children in which motor overflow is imputed as described in Section 2.3.2." Again, why is it important to make associations unbiased for this specific sample? Why not impute missing covariates?*

We reconsidered our imputation approach after receiving your helpful comment. Our initial hope was to reduce missingness in motor overflow, but we may have worsened the situation because we did not impute motor overflow when GAI was missing (since GAI was used in the prediction of motor overflow), and the children missing both motor overflow and GAI tended to have more severe autism (higher ADOS). We are also concerned that our imputation approach did not incorporate the uncertainty of the prediction process, which can cause biased variances. To gain insight into this issue, we reran our comparison of the naïve and deconfounded group differences using two versions of our dataset: complete predictor cases (no imputation for any predictors) and after expanding our original imputation approach to all numerical missing covariates.

The complete case dataset contains 137 autistic and 348 typically developing children from the original 173 autistic and 373 typically developing children. The imputed covariate data

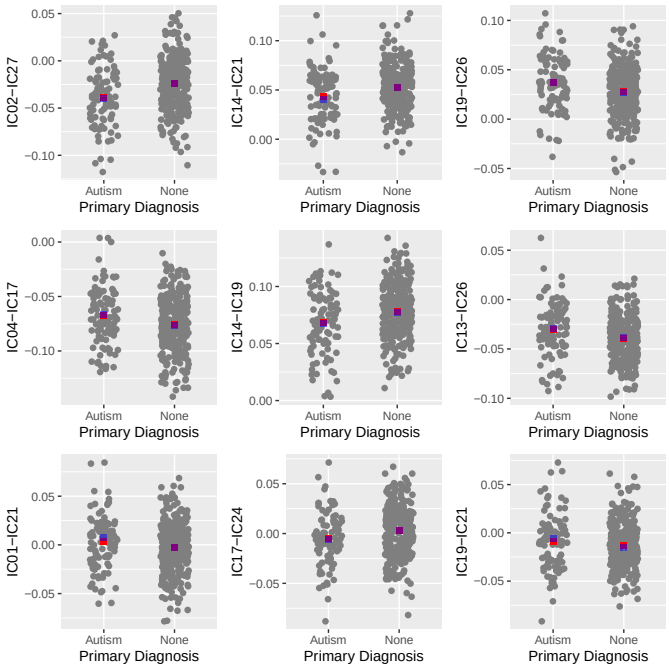
set contains 163 autistic and 370 typically developing children. When estimating naïve differences, we used the same children that passed the motion QC corresponding to the usable data analyzed in DRTMLE, such that the sample changed depending on the method of handling missing data. Using both complete cases and imputed covariate cases, DRTMLE has minor impacts on the group means (Fig. R.1). Using both complete cases and imputed covariate cases, DRTMLE indicates more edges showing a significant difference between autistic and typically developing children (Fig. R.2 and Fig. R.3). The similarity of the results suggests that our imputation approach did not introduce large biases. However, our focus for this paper is on the missingness of the outcome. For simplicity and to avoid distraction from this focus, we present the complete predictor cases dataset (no imputation for any predictors) in the paper. Our approach is a critical first step to address bias in functional connectivity studies. We also note that our Figure 3 improves the transparency of these issues with missingness in the covariates and outcomes, which are sometimes overlooked. All figures and results have been updated to reflect this change in the study sample. We discuss the possible biases due to missing covariates and challenges with imputation in the context of our application in Section 4.2:

> While our method represents an important first step towards addressing bias associated with motion QC, its effectiveness is limited by the missingness of the predictors. Our application assumes that the predictors are missing completely at random. In the case of GAI and motor overflow, missingness may have occurred because some children were unable to complete the cognitive and behavioral tests. The current study focuses on the possible bias due to missingness in the outcome. Imputation to address possible bias due to missingness in the covariates in the context of our application to functional connectivity deserves its own treatment. Multiple imputation involves generating multiple datasets to incorporate uncertainty in the imputation process and a variance adjustment for the between-imputation variance (Little and Rubin, 2019). Ideally, all relationships that will be investigated in the analysis should be included in the imputation process, including relationships with the outcome (Azur et al., 2011), as Moons et al. (2006) found imputation that did not include the outcome resulted in large biases. However, our application includes 153 outcomes (each edge) that contain a biased pattern of missingness (the bias we are correcting using DRTMLE), and it is unclear how to best handle this situation.

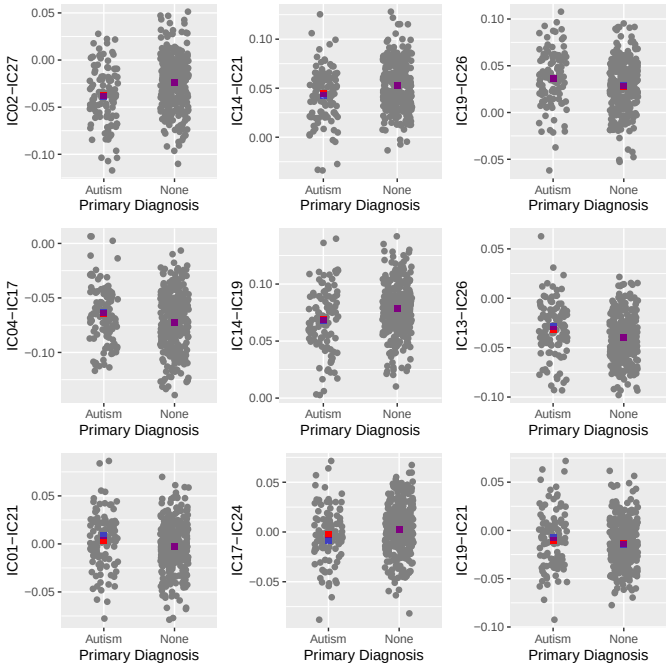
R1.7 *p15: "The adjusted residuals are the same data inputted to the deconfounded group difference and are calculated from the residuals of a linear model with mean FD, max FD, the number of frames with FD< 0.25 mm, sex, race, socioeconomic status, and diagnosis with the effect of diagnosis added back in as described in Section 2.3.2." I don't understand this very well, I think it needs more clarification.*

We added clarification and a citation to Section 2.3.2:

> For each edge, we fit a linear model with mean FD, max FD, number of frames with FD$<$ 0.25 mm, sex (reference: female), race (reference: African American), socioeconomic status, and primary diagnosis (reference: Autism) as predictors. We include sex, race, and socioeconomic status in this model because they differed between autistic and typically developing children (see Section 3.1.1). We then extracted the residuals and

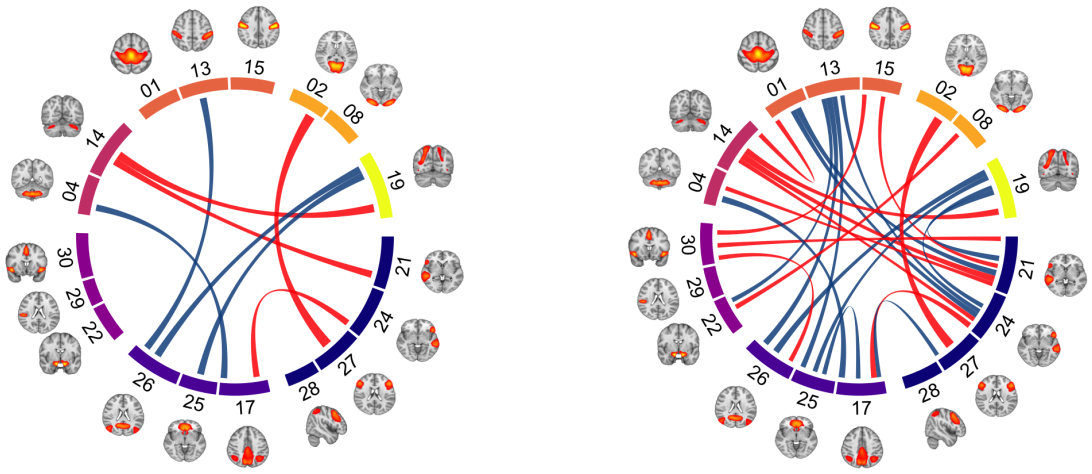


(a) no imputation



(b) imputation

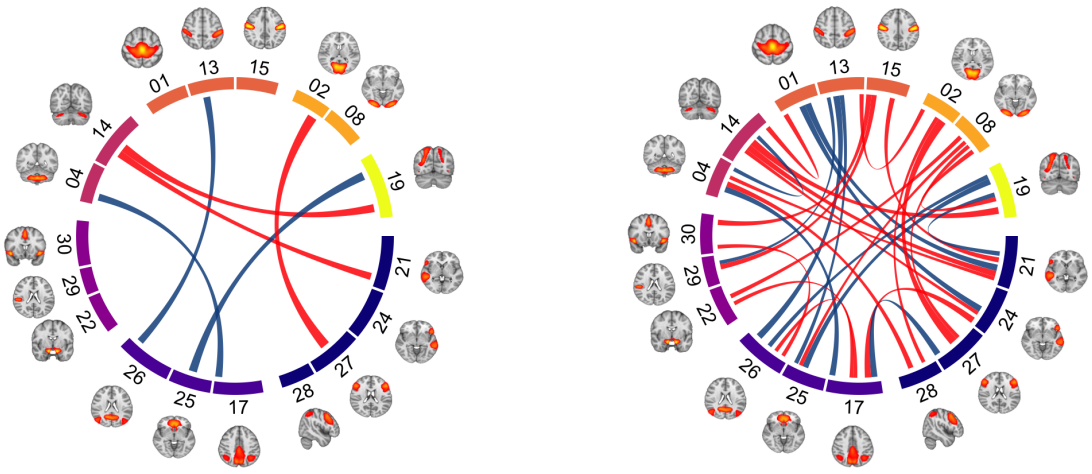
Figure R.1: **Plots of partial correlations, naïve means, and DRTMLE means for each diagnosis group for the nine components with smallest DRTMLE p-values**. Naïve means appear in red. DRTMLE means appear in blue, with opacity such that means close to each other result in purple. When applied in conjunction with the lenient motion QC, the DRTMLE means are similar to the naïve means using both a) complete cases and b) imputed covariate cases .



(a) Naïve Z-Statistic

(b) DRTMLE Z-Statistic

Figure R.2: **Comparison of edges showing a significant ASD-TD difference using the naïve approach and DRTMLE when limited to complete cases**. Z-statistics for ASD versus TD using a) the naïve test and b) using DRTMLE. Connections are thresholded using a false discovery rate (FDR) of 0.20. Blue lines indicate ASD>TD (4 in naïve, 13 in DRTMLE). Red lines indicate ASD<TD (3 in naïve, 12 in DRTMLE).



(a) Naïve Z-Statistic

(b) DRTMLE Z-Statistic

Figure R.3: **Comparison of edges showing a significant ASD-TD difference using the naïve approach and DRTMLE after covariate imputation**. Z-statistics for ASD versus TD using a) the naïve test and b) using DRTMLE. Connections are thresholded using a false discovery rate (FDR) of 0.20. Blue lines indicate ASD>TD (3 in naïve, 11 in DRTMLE). Red lines indicate ASD<TD (3 in naïve, 20 in DRTMLE). This approach does not incorporate uncertainty due to imputation.

added the estimated intercept and effect of primary diagnosis. This approach controls for mean effects that differ by group (ASD versus typically developing) that ideally would be equal, and adjusted residuals are described in the context of site harmonization in Fortin et al. (2018). Then the "naïve" approach is comparable to the approach used in Di Martino et al. (2014), who included diagnosis, sex, age, and mean FD in a linear model. See Section 4.5 for additional discussion.

We also discuss this in Section 4.5: Accounting for variables that should be balanced between diagnosis groups.

R1.8 *p16: "Our parameter of interest is the difference in functional connectivity between autistic and typically developing children: $\psi^* = E(Y(1)|A = 1) - E(Y(1)|A = 0)$..." Related to point 3. this is not a causal model as it involves only one potential outcome (Y(1)). That's fine, but it should be described as a selection model instead.*

As we described in our response to comment R1.3, we have revised the text to clarify that we are interested in an associational parameter that compares functional connectivity in autistic children without an intellectual disability to typically developing children, but we use the notion of counterfactuals related to motion control to define this association.

R1.9 *Figure 2: more details of the simulations would be useful, esp. to compare these results with the results of the real data analyses.*

We have added details in a new Section 2.3.2 Toy Example and tutorial:

> We simulate a dataset with bias and estimate the deconfounded group difference in a tutorial available at `https://github.com/mbnebel/DeconfoundedFMRI`. We generate a sample in which approximately 25% of the participants have ASD, which is similar to the real data (approximately 30% ASD). Then we generate a covariate representing ASD severity, denoted W_c, equal to zero for the typically developing children and generated from a log normal distribution in the ASD group (log μ=2, sd=0.4). We generate nine additional standard normal variables unrelated to diagnosis. Then for the propensity model, data usability is generated from a logistic regression model, $\text{logit}(E[\Delta = 1|W_c = w_c]) = 2 - 0.2 * w_c$. At the mean W_c in the simulated ASD group, the effect is $-0.2 * 7.4 = -1.5$. In the real data, there were non-linearities with steeper slopes at higher ADOS. The slope was approximately -0.077 at the mean ADOS=14.3, leading to $-0.077 * 14.3 = -1.1$ (see Section 3.1.2). The simulation design resulted in approximately 88% and 60% usable data in the typically developing and ASD groups, respectively, compared to 84% and 72% using the lenient criteria in the real data.
>
> Next, we defined the outcome model using a linear model in which the slope is -0.2 for W_c, 0 for the nine other covariates, and 0 and 1.4 for the typically developing and ASD intercepts, respectively. This simulation design resulted in the correlation between W_c and Y equal to -0.56, and the "true" functional connectivity, i.e., $E^*[Y(1)|A = a]$, equal to approximately -0.20 in the ASD and 0 in typically developing groups leading to a between groups Cohen's d=0.51. In our dataset, ADOS is weakly correlated with the partial correlations (min=-0.21, max=0.17 across 153 edges for children with usable data under lenient motion QC), and the largest naïve Cohen's d was 0.50 (the challenges

of calculating a Cohen's d with DRTMLE are discussed in Section 4.3). We note that a study on sleeping autistic toddlers found correlations between functional connectivity and ADOS as great as -0.78 in certain subgroups and some large effect sizes (>0.8) (Lombardo et al., 2019). We generate a random sample equal to 550, then estimate the deconfounded group difference, as depicted in Fig. 2. We will see that in the real data analysis using lenient motion QC, the naïve difference and deconfounded group difference are more similar than this toy example. However, the toy example illustrates the impact of selection bias under a plausible experimental setup.

R1.10 *p22: "The AUCs for predicting usability across the seeds ranged from 0.75 to 0.92, whereas the AUC was 0.68 using logistic regression and 0.69 using a logistic additive model, which indicates that the super learner often improves the accuracy of the propensity model." These are results and don't belong in the Methods section.*

As suggested, we have moved this paragraph to the beginning of Section 3.2 of the Results.

R1.11 *Figure 5: "included (yellow) and excluded (lavender)"....it may just be me, but the excluded look gray, not lavender.*

We have changed all figure captions to refer to this color as slate blue and have modified references to these figures in the Results accordingly.

R1.12 *For the stringent vs. lenient motion thresholding, it's disappointing that the stringent thresholding wasn't analyzable due to small sample. Leaving aside the power concern, can the authors speculate on the whether the stringency impacts the bias and the ability of the proposed method to correct it?*

We added a discussion of this in Section 4.3:

> Studies using stricter criteria may induce larger biases, in which case there may be larger differences between the DRTMLE means and the naïve means. We were unable to apply DRTMLE following strict motion QC because only 29 autistic children had usable data. Visual inspection of the densities in Figure 5 reveals that the sampling bias was larger in the strict versus lenient case for all phenotypes. We speculate that the stricter criteria would result in larger differences between DRTMLE and the naïve approach.

R1.13 *top of p30: Are the smaller standard errors of the analyses due to using a more efficient analysis method rather than due to bias correction per se? Do the real-data analyses seem to be correcting bias after all?*

We more closely examined the changes in means and standard errors associated with the 17 edges indicated by DRTMLE but not the naïve approach as showing an ASD-TD difference at FDR=0.20 (Figure R3). Only a subset of the additional edges indicated by DRTMLE showed a decrease in the standard errors relative to the naïve approach, but all showed an increase in the absolute mean difference between the ASD and typically developing groups. This suggests that both a change in the means and a decrease in standard errors contributes to the more extensive ASD-TD differences observed using DRTMLE. We have revised this section for clarity:

In this application, the deconfounded means were very similar to the naïve means (Web Supplement Figure S.4). Additionally, partial correlations were highly variable, with the range of partial correlations in the ASD and typically developing groups broadly overlapping. However, the small changes in the means also contributed to the more extensive ASD-TD differences in DRTMLE versus the naïve approach depicted in Figure 7. Of the additional 17 edges selected at FDR=0.20 by DRTMLE, eight of the 17 had a decrease in the standard error but all 17 had an increase in the absolute difference between the ASD and typically developing groups. Across all edges, the absolute difference between groups increased in 106/153 edges and the standard error decreased in 81/153. We discuss effect sizes in Section 4.3.

We revised Section 4.3 beginning with the heading,

> Similarities between naïve and deconfounded means, sample size limitations, inference, and effect size.

Then,

> When using the lenient criteria on the KKI dataset, the bias corrections using DRTMLE were small (Web Supplement Figure S.3). This may be due to weak relationships between the phenotypic variables and the partial correlations, which is consistent with recent ABCD study results suggesting that the largest, replicable associations between functional connectivity and behavioral phenotypes ranged from $r = 0.14 - 0.34$ (Marek et al., 2022). Additionally, our data processing and the use of partial correlations from group ICA may have decreased motion artifacts at the expense of attenuating associations. Our simulated example in Figure 2 provides an example where the relationship with autism severity is stronger, and we see larger bias.

Also, please see our response to comment R1.12 above.

R1.14 *Figure 7: It would be better to compare differences in the magnitude of effects rather than just significance. That would get more at bias rather than efficiency.*

To visualize the bias, we compare naïve means and DRTMLE means for each diagnosis group for the nine components with the smallest DRTMLE p-values in Web Supplement Figure S.4, which is also Fig. R.1 in this response. To clarify this point, we have modified the caption of Figure 7:

> **The DRTMLE deconfounded group difference revealed more extensive differences than the naïve approach**. Z-statistics for ASD versus TD using a) the naïve test and b) using DRTMLE. Connections are thresholded using a false discovery rate (FDR) of 0.20. Blue lines indicate ASD>TD (4 in naïve, 13 in DRTMLE). Red lines indicate ASD<TD (4 in naïve, 12 in DRTMLE). Brain regions contributing to each independent component are illustrated and components are grouped by functional assignment. Navy nodes: control. Blue violet: default mode. Purple: salience/ventral attention. Magenta: pontomedullary/cerebellar. Coral: somatomotor. Orange: visual. Yellow: dorsal attention. FDR=0.05 is plotted in Web Supplement Figure S.3. See

Web Supplement Figure S.4 for a visualization of the naïve and deconfounded means and individual-level partial correlations. These plots were generated using the circlize package in R (Gu et al., 2014) and the tutorial provided by Mowinckel (2018).

R1.15 *p32: "The estimate of mean functional connectivity should be representative of all children enrolled in the study..." If it's a convenience sample, why is that the case?*

We clarified this sentence:

> The estimate of mean functional connectivity should be representative of all children enrolled in the study, assuming the enrolled participants are a representative sample from the target population. However, we observed that participants with usable rs-fMRI data differed from participants with rs-fMRI data that would have been excluded using conventional approaches.

Please also see our previous response to comment 4.

R1.16 *The Discussion section could be substantially shortened by not repeating prior results.*

We removed the following from the second paragraph of the discussion: ~~Additionally, rs-fMRI exclusion probability changed with symptom severity and age.~~

We removed the following from the third paragraph of the discussion: ~~For instance, evidence of a relationship between SRS and functional connectivity was stronger among participants who passed strict motion QC than among participants who passed lenient motion QC (Figure 6)~~

We removed the first four sentences in Section 4.1. ~~Collectively, the findings suggest that group differences in functional connectivity are more robust using the DRTMLE approach as compared to the naïve approach. What evidence do we have to support the validity of these findings? First, the sign of average group effects remained consistent across methods, as did the direction of group differences (Web Supplement Table S.3. In this study, DRTMLE appears to have a larger effect on the variance than on the group means, and we discuss the implications of this and effect size in Section 4.2. Second,~~

Reviewer 3

Reviewer #3: This manuscript describes an examination of the effects of lenient and stringent motion correction methods for resting-state fMRI (rs-fMRI) data on the characteristics of a sample of children (N = 545, ages 8 to 13 years) with and without autism spectrum disorder (ASD). Additionally, the authors tested a machine learning approach (doubly robust targeted minimum loss based estimation; DRTMLE) for addressing non-random missing data due to excessive motion in rs-fMRI analyses. Findings indicated children with ASD who were older and had higher cognitive abilities and fewer social and motor control difficulties were less likely to be excluded based on motion correction methods. Additionally, the authors found that DRTMLE revealed more differences in functional connectivity measures during rs-fMRI between the ASD and typically developing (TD) groups than a naïve approach.

Overall, this manuscript has many strengths and highlights important implications for rs-fMRI studies of youth with neurodevelopmental disorders such as ASD and ADHD. The manuscript is

well-written and thorough. Additionally, the authors have also made all of their scripts publicly available, which is much appreciated. Despite these strengths, I believe there are a few ways the authors could further improve the manuscript prior to publication.

R3.1 *The authors do a nice job of explaining how sample characteristics differ based on motion correction approaches. However, as they note, some participants are excluded from rs-fMRI studies for other reasons (e.g., having intellectual disability, epilepsy, or depression). Notably, many of these conditions can co-occur with a diagnosis of ASD and thus result in even more bias in the ASD samples eventually included in analyses. It would be useful for the authors to mention how this issue may further limit the representativeness of ASD samples in rs-fMRI studies of youth, and if possible, to perhaps include the numbers of participants from the ASD and TD groups who did not meet the various inclusion criteria for the present study.*

We would like to clarify that the inclusion criteria explained in *Section 2.1.1 Study Population* were for the original data-collecting studies. Unfortunately, we cannot determine the number of participants who did not meet the various inclusion criteria for the ASD and typically developing groups because the data-collecting IRB protocols require that we expunge records of participants who are deemed ineligible during the initial phone screening.

We have added a discussion of the limitations of our study sample, which includes population estimates of some of the common comorbid conditions that were considered exclusionary.

> A goal of the original data-collecting studies was to estimate the association between functional connectivity and a diagnosis of ASD separate from any association between functional connectivity and intellectual disability. A recent cohort-based study of heritability and familial risk indicated that ASD without a co-occurring intellectual disability may have a greater genetic basis than ASD with an intellectual disability (Xie et al., 2020), suggesting that these ASD phenotypes may be etiologically distinct. Approximately 35% of autistic children have an intellectual disability (Maenner et al., 2021), and our findings may not generalize to this segment of the autism population (Reiter et al., 2019). Unless we as a field make a concerted effort to correct for the over-representation of cognitively abled children in autism research, we will never be able to answer this question. In addition, an estimated 70% of autistic children have at least one comorbid disorder (Simonoff et al., 2008). The original data-collecting studies allowed for two of the most common comorbid mental health disorders, namely ADHD and social anxiety disorder, but excluded all others. Because research informs the revision of diagnostic criteria and health and educational policy, we must improve recruitment and data collection from children we have historically excluded from rs-fMRI research of autism.

R3.2 *DRTMLE appears to be a promising approach for addressing the issue of missing data in rs-fMRI analyses. I believe it is important for the authors to clarify, though, that analysis approaches for addressing biased data do not solve the issue of having biased samples. If youth with ASD who are included in rs-fMRI studies are not representative of the full population of youth with ASD (e.g., those with more severe presentations), then we as a field also need methods for improving recruitment and data acquisition with those who are currently excluded.*

Please see our response to the previous comment and our comment below. In particular, we now state at the beginning of Section 4.2,

> Although DRTMLE can address issues of missing data due to motion QC, it does not address possible biases in the sample of children with behavioral data that appear in the study, which may differ from the target population.

R3.3 *The authors report extensive mock training sessions as one potential method for reducing missingness due to motion (section 4.5). This relates to my above point, and I think the manuscript would benefit from a discussion of the success rates of different mock scanning approaches (e.g., Horien et al., 2020) as well as a description of other motion-reduction methods for rs-fMRI (e.g., FIRMM or movie-watching; Greene et al., 2018).*

We have added a discussion of the merits of various experimental approaches to improving participant compliance in the scanner in the second paragraph of the new Discussion Section 4.2 The target population, other possible biases, and methods to address them:

> Scanning these underrepresented populations requires thoughtful experimental accommodations. Some approaches for improving the likelihood of collecting usable data from pediatric participants focus on scan preparation: having a caregiver model scan procedures, practicing in an MRI simulator (Nordahl et al., 2016; Horien et al., 2020; Simhal et al., 2021), or playing a virtual reality-based MRI game (Stunden et al., 2021; Pua et al., 2020). Others focus on modifications during the scan: passive movie-viewing to reduce boredom (Vanderwal et al., 2019), providing real-time feedback to participants using framewise integrated real-time MRI monitoring (FIRMM) software (Dosenbach et al., 2017; Greene et al., 2018), or using personalized incentive systems to reward compliance with MRI instructions (Pua et al., 2020). The efficacy of these strategies varies with age, and some strategies come at a cost. For instance, extensive MRI simulator practice has been used to scan small samples of autistic children with intellectual disabilities (Nordahl et al., 2016), but this places additional burdens on families by requiring multiple visits. In addition, some studies with autistic children have observed that head motion, while reduced after extensive training, is still associated with symptom severity (Simhal et al., 2021; Gabrielsen et al., 2018). Preliminary evidence suggests virtual reality-based games played at home may be as effective as MRI simulator training for typically developing children, (Stunden et al., 2021); however, it remains unclear if children with neurodevelopmental disorders would respond similarly. Investigating in-scanner strategies, Greene et al. (2018) found that passive movie-viewing and real-time feedback both reduced head motion, but movie-viewing changed functional connectivity. Finn and Bandettini (2021) found that changes in functional connectivity elicited by movie-viewing in adults aided in the prediction of behavioral traits, but similar work has not yet been conducted in children. As we develop methods for increasing participant compliance, our tolerance for what constitutes acceptable levels of head motion will likely change (for instance, Marek et al. (2019) excluded 40% of the ABCD study sample which was collected using FIRMM, while Marek et al. (2022) excluded 60%). We hope our proposed approach will be used in combination with some of these experimental strategies to advance studies including important subgroups of autistic children.

R3.4 *The authors may consider examining the relationships between the various demographic and clinical variables of interest and mean or max FD within each sample, as some of these variables have been shown to relate to motion during rs-fMRI (e.g., sex and age; Ekhtiari et al., 2019).*

As suggested, we have conducted additional analyses to examine the relationships between phenotype and mean FD, as well as sex and mean FD, with changes detailed below.

In Section 2.2.2,

> We also conducted a similar analysis using univariate GAMs assuming Gaussian errors to examine how the phenotypes are related to mean FD. We conducted separate analyses for the study sample (both usable and unusable cases), children passing the lenient criteria, and children passing the strict criteria, again using FDR correction for seven comparisons within each sample. We also examined whether mean FD differed by sex for these three samples using Mann-Whitney U-tests.

Then in Section 3.1.2,

> When we considered children with usable and unusable data, we observed similar patterns between mean FD and the covariates as we saw between the probability of exclusion and the covariates. Children with higher (more severe) ADOS scores, SRS scores, inattentive symptoms, hyperactive/impulsive symptoms, or poorer motor control moved more, while older children and children with higher GAI moved less (Web Supplement Figure R.4, black lines, all FDR-adjusted p$<$.005). We saw similar patterns following lenient motion QC except that the tendency for children with higher GAI to move more was not significant (FDR-adjusted p=.496). The strict motion QC appears to eliminate the relationship between motion and behavioral phenotypes (FDR-adjusted p>0.05). However, this may be due in part to restricting the distribution of phenotypes, as highlighted by the change in the distribution of ADOS. These biases are further examined in the next section. Before scan exclusion, there was a tendency for boys to move more than girls (p=0.03, uncorrected, r=0.10), but there was not a significant difference between boys and girls with usable data in the lenient (p=0.27) or strict samples (p=0.17).

Then we have added Figure R.4 to the Web Supplement.

R3.5 *Relatedly, could the authors please provide rationale for why sex and SES were not included as covariates in the GAMs exploring exclusion odds and covariates (section 2.2.2)?*

These models do not include sex, SES, and race to be consistent with the later definition of the propensity model. These variables were not included in the propensity models due to imbalance between diagnosis groups (Table 1). However, we recreated the figures using GAMs that control for sex, SES, and race, and found that the patterns were very similar (Fig. R.5). We added the following to Section 2.2.2 to clarify:

> We did not include SES, sex, or race in these models because these variables are not included in the propensity model in Section 2.3.3 due to imbalance between diagnosis groups. Results when controlling for these variables were highly similar (not shown.)

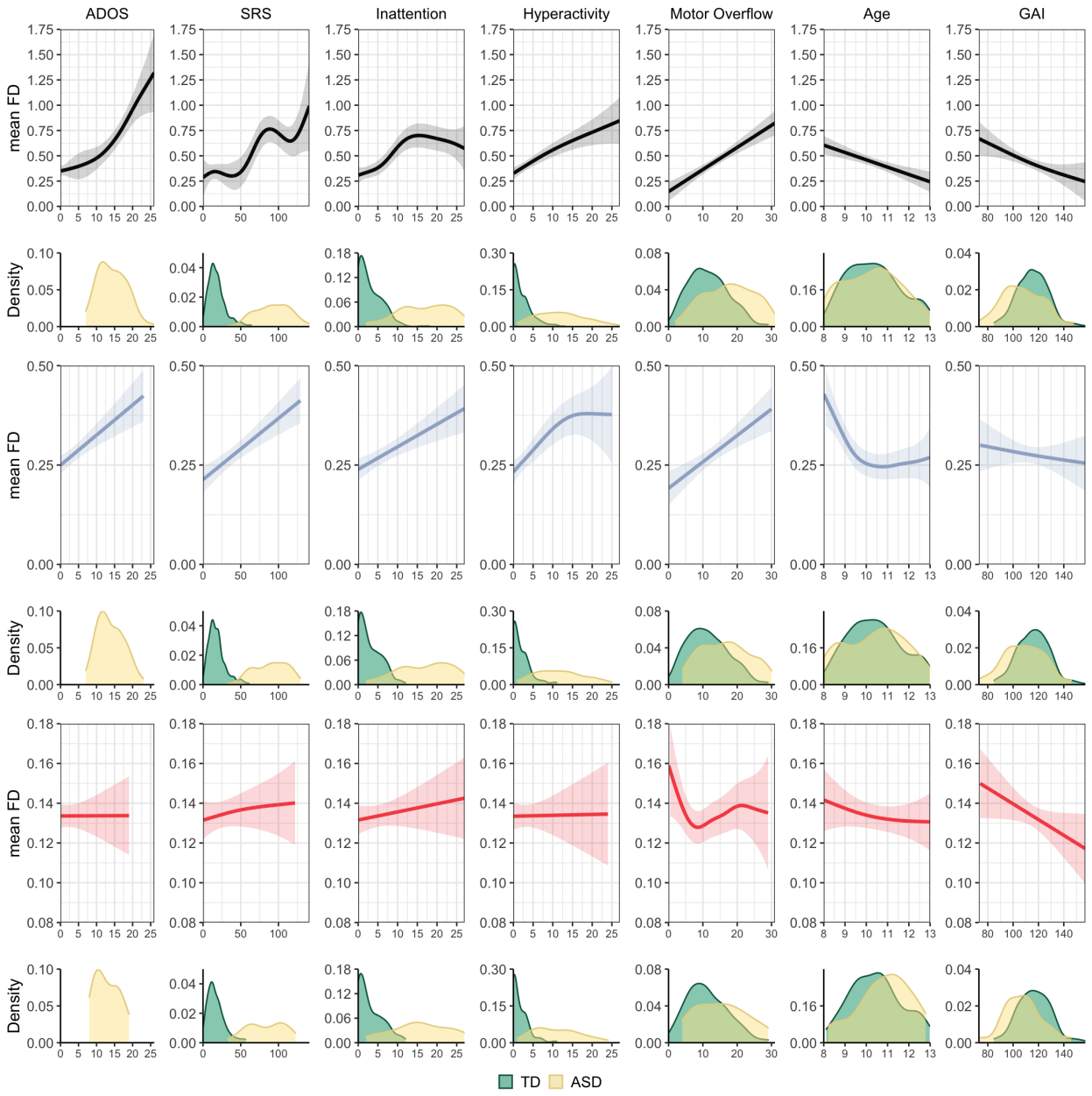
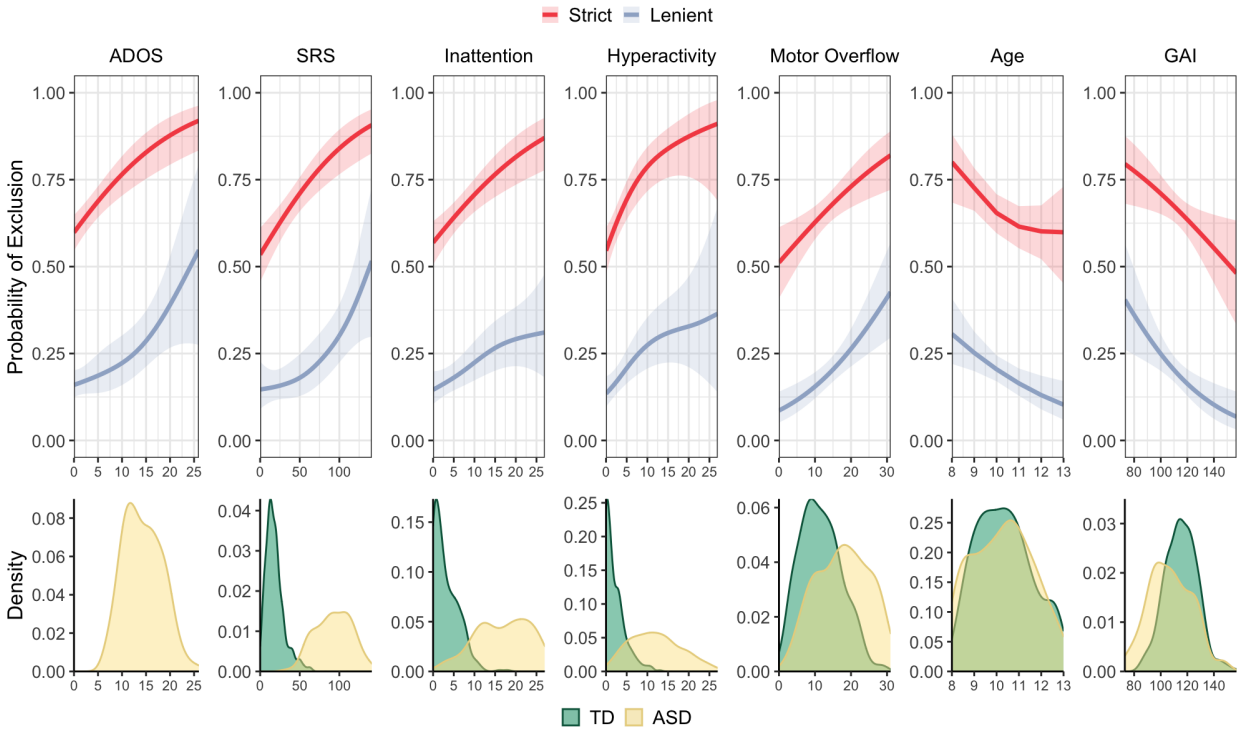
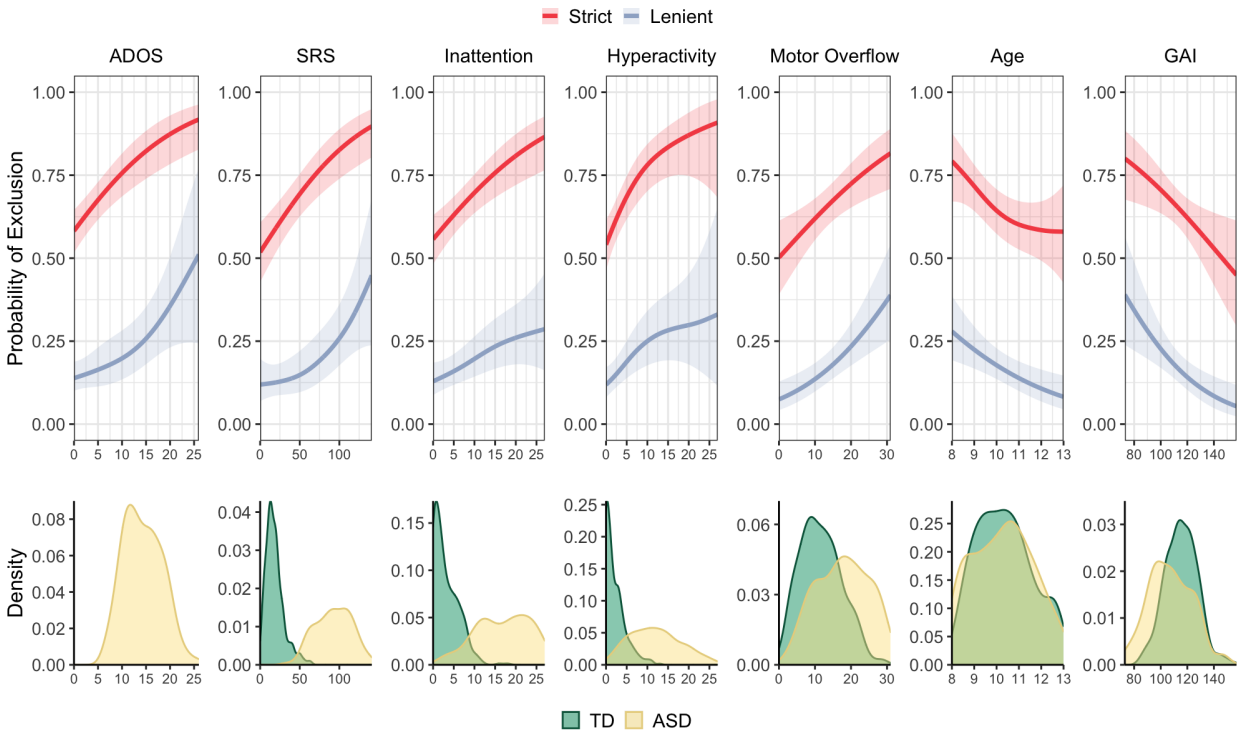


Figure R.4: **Univariate analysis of mean framewise displacement as a function of participant characteristics**. From left to right: Autism Diagnostic Observation Schedule (ADOS) total scores, social responsiveness scale (SRS) scores, inattentive symptoms, hyperactive/impulsive symptoms, total motor overflow, age, and general ability index (GAI) for children with usable and unusable rsfMRI data (top row, black lines), children with usable data under lenient motion QC (slate blue lines), and children with usable data under strict motion QC (red lines). Variable distributions for each diagnosis group are displayed across the bottom panel (TD=typically developing, green; ASD=autism spectrum disorder, yellow).



(a) no covariates



(b) sex, race, and SES as covariates

Figure R.5: **rs-fMRI exclusion probability changes with phenotype and age.**

A related discussion is in Section 4.5.

R3.6 *Please provide a measure of effect size in addition to the adjusted p-values for the Mann-Whitney tests (section 3.1.3).*

We have added effect sizes to Table S.2 and S.3, and we modified the text in Section 2.2.3:

> We then used one-sided Mann-Whitney U tests to test for differences between included and excluded participants for each measure stratified by diagnosis. We also calculated effect sizes as $Z/\sqrt{N}$.

We also modified the text in Section 3.1.3:

> For the lenient motion QC, median values for included and excluded participants, effect sizes, and FDR-adjusted p values for each measure and diagnosis group are summarized in Web Supplement Table S.1.
>
> Significant p-values were associated with small effect sizes ($r = 0.12$ to 0.25).
>
> Differences between included and excluded children also tended to occur using the strict criteria, although in general significance was reduced, owing in part to the reduced sample size in the included group but also to some reduced effect sizes (i.e, motor overflow).
>
> Median values for included and excluded participants, effect sizes, and FDR-adjusted p values for each measure and diagnosis group are summarized in Web Supplement Table S.2.

R3.7 *Could the authors please provide more information about why calculating effect sizes in DRTMLE is a problem (p. 35, line 720)?*

This is any interesting problem, and we have added the following to the discussion:

> Unfortunately, calculating effect sizes in DRTMLE is an open problem. We would need to define a new parameter of interest, the population pooled standard deviation, under the counterfactual that all data are usable and then define an estimator of this parameter. When calculating Cohen's d in a two sample t-test setting, the standard errors of each mean can be multiplied by $\sqrt{n_1}$ and $\sqrt{n_2}$ to recover each standard deviation, which can then be pooled. In DRTMLE, this would not result in an estimate of the standard deviations of each group under the counterfactual of all usable data, since the standard errors are derived from the influence function of ψ. This is an important avenue for future research.

R3.8 *To enhance clarity, it may be more appropriate to say resting-state fMRI rather than "fMRI" in the manuscript's title.*

We have made the suggested change to the title of the manuscript.

References

Azur, M. J., Stuart, E. A., Frangakis, C., and Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? International Journal of Methods in Psychiatric Research, 20(1):40–49.

Bradley, V. and Nichols, T. E. (2022). Addressing selection bias in the UK Biobank neurological imaging cohort. medRxiv, (2017):2022.01.13.22269266.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D., Gallagher, L., Kennedy, D., Keown, C. L., Keysers, C., Lainhart, J. E., Lord, C., Luna, B., Menon, V., Minshew, N. J., Monk, C., Mueller, S., Müller, R.-A., Nebel, M. B., Nigg, J. T., O'Hearn, K., Pelphrey, K. A., Peltier, S. J., Rudie, J. D., Sunaert, S., Thioux, M., Tyszka, J. M., Uddin, L. Q., Verhoeven, J. S., Wenderoth, N., Wiggins, J. L., Mostofsky, S. H., and Milham, M. P. (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry, 19(6):659–667.

Dosenbach, N. U., Koller, J. M., Earl, E. A., Miranda-Dominguez, O., Klein, R. L., Van, A. N., Snyder, A. Z., Nagel, B. J., Nigg, J. T., Nguyen, A. L., Wesevich, V., Greene, D. J., and Fair, D. A. (2017). Real-time motion analytics during brain MRI improve data quality and reduce costs. NeuroImage, 161:80–93.

Finn, E. S. and Bandettini, P. A. (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. NeuroImage, 235:117963.

Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., McInnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M., and Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. NeuroImage, 167:104–120.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., and Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. American Journal of Epidemiology, 186(9):1026–1034.

Gabrielsen, T. P., Anderson, J. S., Stephenson, K. G., Beck, J., King, J. B., Kellems, R., Top, D. N., Russell, N. C., Anderberg, E., Lundwall, R. A., Hansen, B., and South, M. (2018). Functional MRI connectivity of children with autism and low verbal and cognitive performance. Molecular Autism, 9(1):1–14.

Greene, D. J., Koller, J. M., Hampton, J. M., Wesevich, V., Van, A. N., Nguyen, A. L., Hoyt, C. R., McIntyre, L., Earl, E. A., Klein, R. L., Shimony, J. S., Petersen, S. E., Schlaggar, B. L., Fair, D. A., and Dosenbach, N. U. (2018). Behavioral interventions for reducing head motion during MRI scans in children. NeuroImage, 171:234–245.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize implements and enhances circular visualization in R. Bioinformatics, 30(19):2811–2812.

Hernan, M. A. and Robins, J. M. (2020). Causal Inference: What If. Chapman Hall/CRC, Boca Raton.

Horien, C., Fontenelle, S., Joseph, K., Powell, N., Nutor, C., Fortes, D., Butler, M., Powell, K., Macris, D., Lee, K., Greene, A. S., McPartland, J. C., Volkmar, F. R., Scheinost, D., Chawarska, K., and Constable, R. T. (2020). Low-motion fMRI data can be obtained in pediatric participants undergoing a 60-minute scan protocol. Scientific Reports 2020 10:1, 10(1):1–10.

Lash, T. L., VanderWeele, T. J., Haneuse, S., and Rothman, K. J. (2021). Modern Epidemiology, Fourth Edition. Wolters Kluwer.

Little, R. J. and Rubin, D. B. (2019). Statistical analysis with missing data, volume 793. John Wiley & Sons.

Lombardo, M. V., Eyler, L., Moore, A., Datko, M., Barnes, C. C., Cha, D., Courchesne, E., and Pierce, K. (2019). Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. eLife, 8:e47427.

Maenner, M. J., Shaw, K. A., Bakian, A. V., Bilder, D. A., Durkin, M. S., Esler, A., Furnier, S. M., Hallas, L., Hall-Lande, J., Hudson, A., Hughes, M. M., Patrick, M., Pierce, K., Poynter, J. N., Salinas, A., Shenouda, J., Vehorn, A., Warren, Z., Constantino, J. N., DiRienzo, M., Fitzgerald, R. T., Grzybowski, A., Spivey, M. H., Pettygrove, S., Zahorodny, W., Ali, A., Andrews, J. G., Baroud, T., Gutierrez, J., Hewitt, A., Lee, L.-C., Lopez, M., Mancilla, K. C., McArthur, D., Schwenk, Y. D., Washington, A., Williams, S., and Cogswell, M. E. (2021). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. MMWR. Surveillance Summaries, 70(11):1–16.

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., Moore, L. A., Conan, G. M., Uriarte, J., Snider, K., Lynch, B. J., Wilgenbusch, J. C., Pengo, T., Tam, A., Chen, J., Newbold, D. J., Zheng, A., Seider, N. A., Van, A. N., Metoki, A., Chauvin, R. J., Laumann, T. O., Greene, D. J., Petersen, S. E., Garavan, H., Thompson, W. K., Nichols, T. E., Yeo, B. T. T., Barch, D. M., Luna, B., Fair, D. A., and Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. Nature 2022, pages 1–7.

Marek, S., Tervo-Clemmens, B., Nielsen, A. N., Wheelock, M. D., Miller, R. L., Laumann, T. O., Earl, E., Foran, W. W., Cordova, M., Doyle, O., Perrone, A., Miranda-Dominguez, O., Feczko, E., Sturgeon, D., Graham, A., Hermosillo, R., Snider, K., Galassi, A., Nagel, B. J., Ewing, S. W., Eggebrecht, A. T., Garavan, H., Dale, A. M., Greene, D. J., Barch, D. M., Fair, D. A., Luna, B., and Dosenbach, N. U. (2019). Identifying reproducible individual differences in childhood functional brain networks: An ABCD study. Developmental Cognitive Neuroscience, 40:100706.

Moons, K. G., Donders, R. A., Stijnen, T., and Harrell, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. Journal of Clinical Epidemiology, 59(10):1092–1101.

Mowinckel, A. (2018). Circular Plots in R and Adding Images [blog post]. retrieved from https://drmowinckels.io/blog/2018-05-25-circluar-plots-in-r-and-adding-images on 31/10/2021.

Nordahl, C. W., Mello, M., Shen, A. M., Shen, M. D., Vismara, L. A., Li, D., Harrington, K., Tanase, C., Goodlin-Jones, B., Rogers, S., Abbeduto, L., and Amaral, D. G. (2016). Methods for acquiring MRI data in children with autism spectrum disorder and intellectual impairment without the use of sedation. Journal of Neurodevelopmental Disorders, 8(1):1–10.

Pua, E. P. K., Barton, S., Williams, K., Craig, J. M., and Seal, M. L. (2020). Individualised MRI training for paediatric neuroimaging: A child-focused approach. Developmental Cognitive Neuroscience, 41:100750.

Reiter, M. A., Mash, L. E., Linke, A. C., Fong, C. H., Fishman, I., and Müller, R. A. (2019). Distinct Patterns of Atypical Functional Connectivity in Lower-Functioning Autism. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4(3):251–259.

Simhal, A. K., Filho, J. O., Segura, P., Cloud, J., Petkova, E., Gallagher, R., Castellanos, F. X., Colcombe, S., Milham, M. P., and Di Martino, A. (2021). Predicting multiscan MRI outcomes in children with neurodevelopmental conditions following MRI simulator training. Developmental Cognitive Neuroscience, 52:101009.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., and Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. Journal of the American Academy of Child and Adolescent Psychiatry, 47(8):921–929.

Stunden, C., Stratton, K., Zakani, S., and Jacob, J. M. (2021). Comparing a Virtual Reality-Based Simulation App (VR-MRI) With a Standard Preparatory Manual and Child Life Program for Improving Success and Reducing Anxiety During Pediatric Medical Imaging: Randomized Clinical Trial. Journal of medical Internet research, 23(9).

Vanderwal, T., Eilbott, J., and Castellanos, F. X. (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging.

Xie, S., Karlsson, H., Dalman, C., Widman, L., Rai, D., Gardner, R. M., Magnusson, C., Sandin, S., Tabb, L. P., Newschaffer, C. J., and Lee, B. K. (2020). The Familial Risk of Autism Spectrum Disorder with and without Intellectual Disability. Autism Research, 13(12):2242–2250.

Xie, S., Karlsson, H., Dalman, C., Widman, L., Rai, D., Gardner, R. M., Magnusson, C., Sandin, S., Tabb, L. P., Newschaffer, C. J., and Lee, B. K. (2020). The Familial Risk of Autism Spectrum Disorder with and without Intellectual Disability. *Autism Research*, 13(12):2242–2250.

Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., Li, Q., Zuo, X. N., Castellanos, F. X., and Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, 76:183–201.

Zhou, D., Cornblath, E. J., Stiso, J., Teich, E. G., Dworkin, J. D., Blevins, A. S., and Bassett, D. S. (2020). Gender diversity statement and code notebook v1.0 [software]. *retrieved from https://github.com/dalejn/cleanBib on 30/11/2021.*