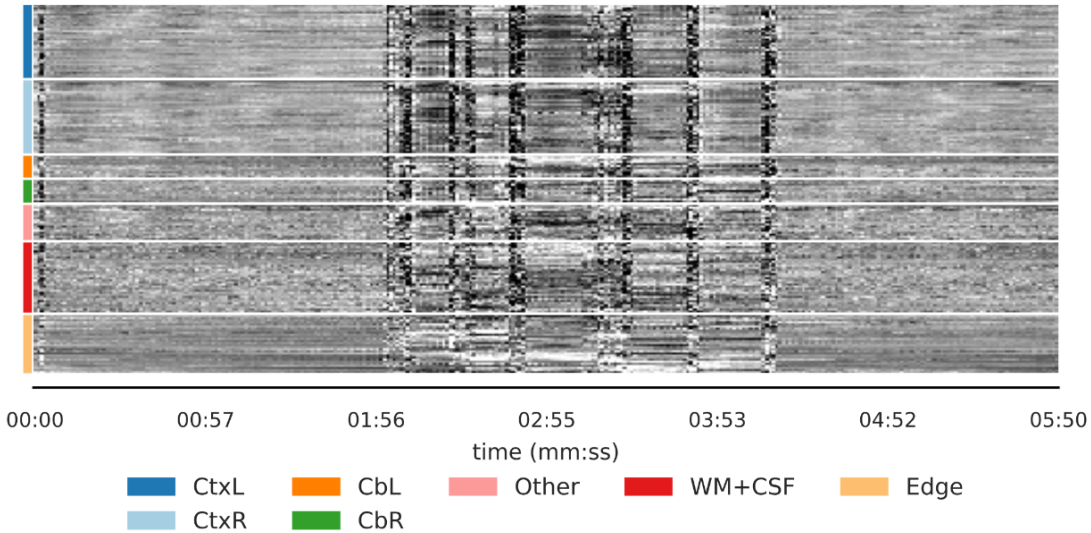
Improving estimates of functional connectivity in children with autism

Benjamin Risk

Department of Biostatistics and Bioinformatics, Emory University

thebrisklab.org

Problem I: Motion Artifacts



Rigid body motion correction

↓

Nuisance signals regression
(Ciric et al. 2017)

↓

Remove scans with excessive motion
(Satterthwaite et al. 2013, Power 2017, Parkes et al. 2018)

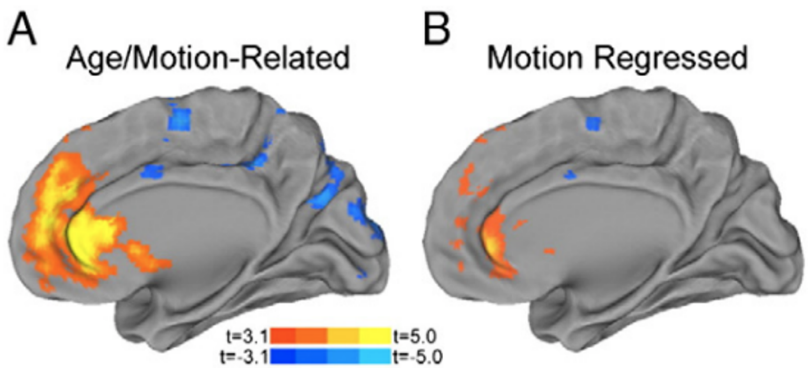
- Existing motion control using regression is inadequate because motion patterns are complex (Power et al. 2014).
- Current best practices remove time points with framewise displacement >0.2 mm, then remove participants if < 5 minutes of data remain after this scrubbing.

Challenges in pediatric neuroimaging

Motion challenges:

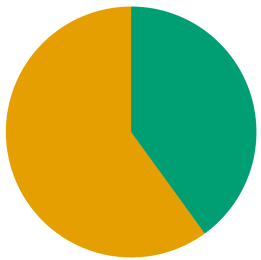
- Younger children move more (Satterthwaite et al. 2012)
- Confounding with neurodevelopmental trajectories.

"It really, really, really sucks. My favorite result of the last five years is an artifact." -*Steven Petersen, Professor of cognitive neuroscience at Washington University in St. Louis.*



Problem II: Motion QC creates selection bias

- Motion control leads to drastic reductions in sample size.
- ABCD study **removed 60 – 75%** of children due to excessive motion (Marek et al. 2022, Nielsen et al. 2019).
- In ABCD, this creates selection bias, disproportionately selecting for: higher SES, White participants, older, females, higher neurocognitive skills, fewer neurodevelopmental problems (Cosgrove et al. 2022).
- Unethical?



Connectivity and autism

- Connectivity theory of autism (Deen and Pelphrey 2012):
  - Underconnectivity in long-range connections (Just et al. 2012).
  - Overconnectivity in local connections (Keown et al. 2013).
- Disruptions in the default mode network (Yerys et al. 2015), including anterior-to-posterior parts (Di Martino et al. 2013).
- Some similarities with the patterns produced by motion artifacts.

Default mode decreased connectivity

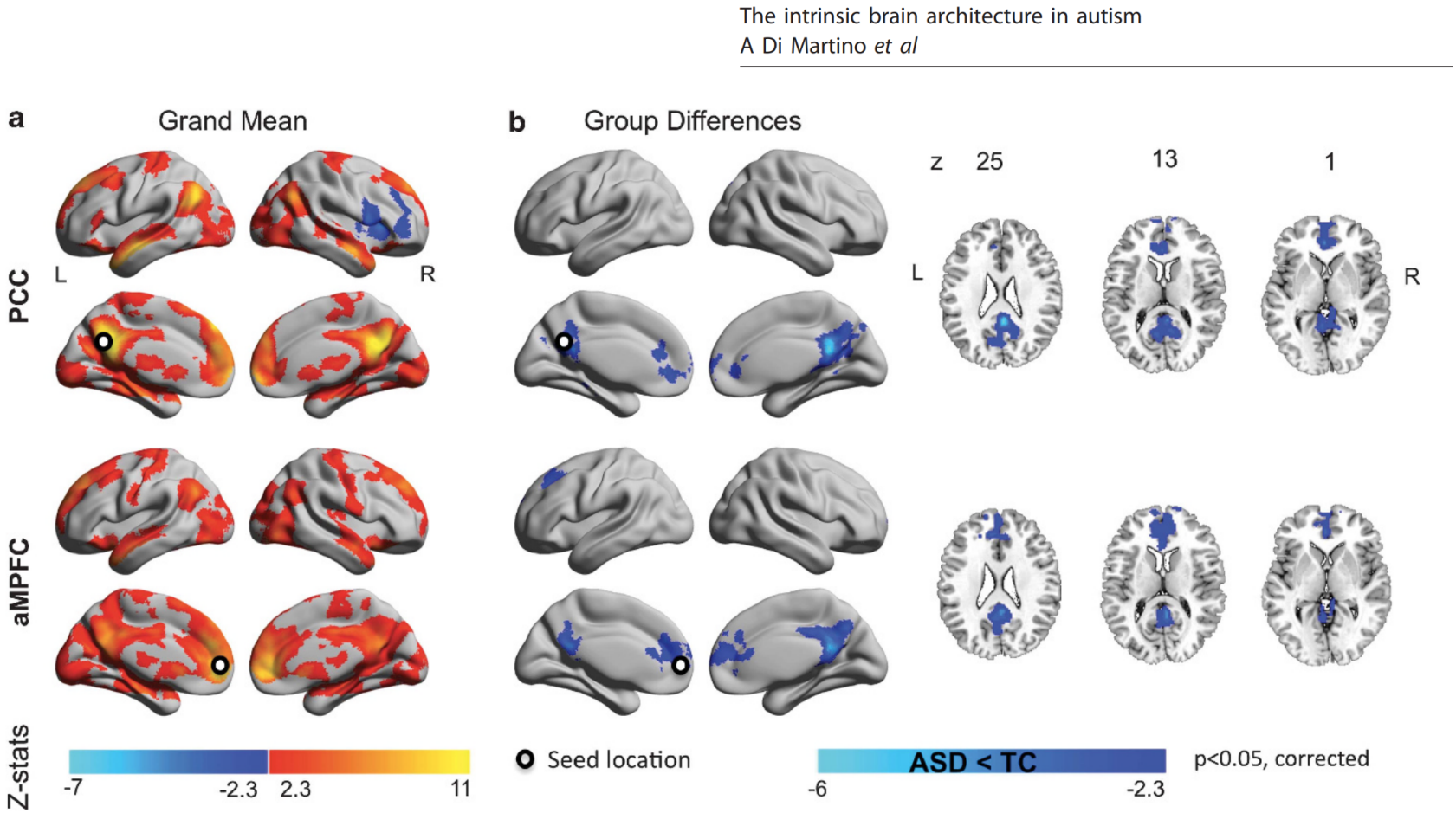


Figure: Di Martino et al. (2014)

Motion is higher in autistic children: subset of ABIDE

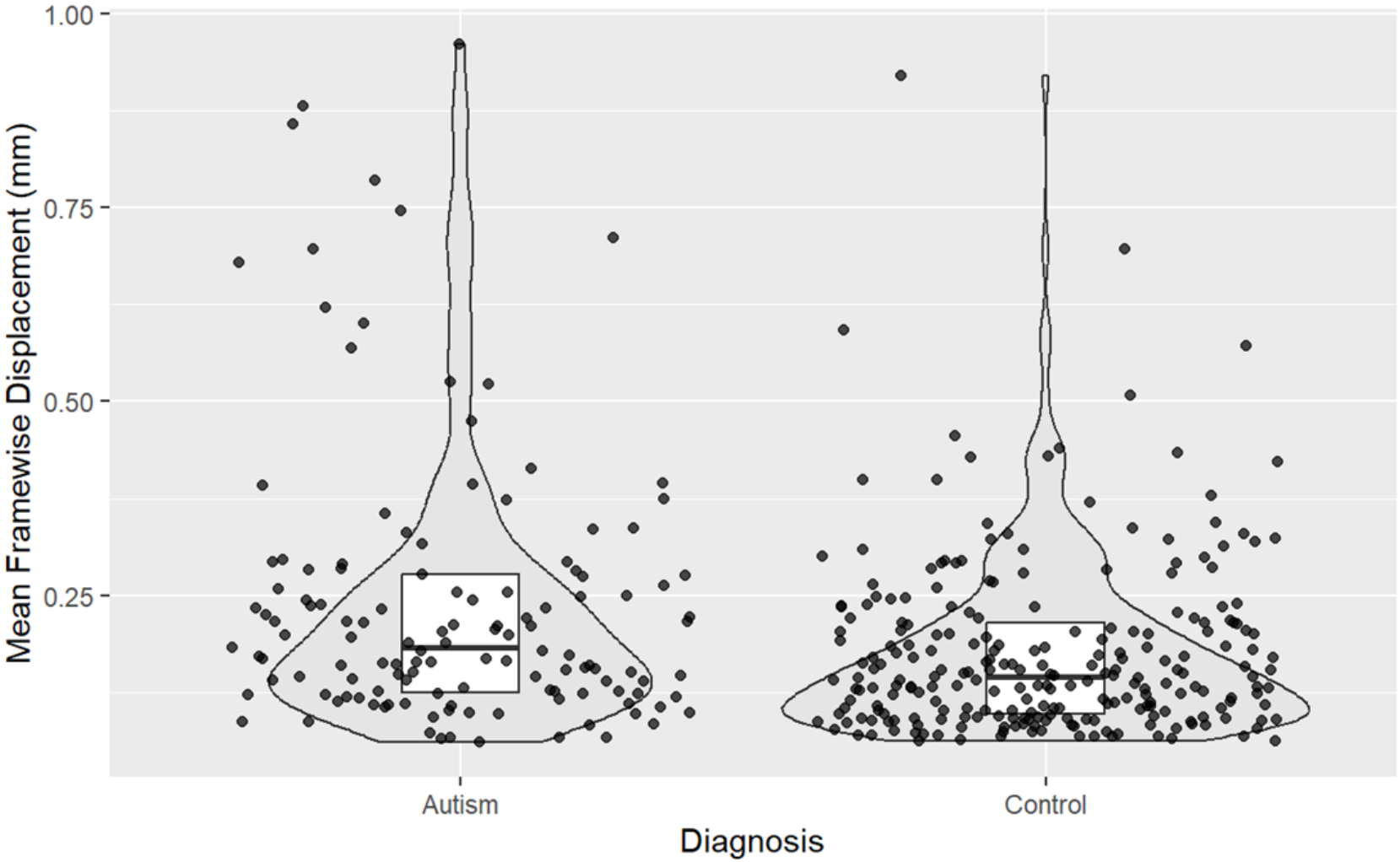


Figure: 8-13 year-olds from ABIDE.

Motion is higher in autistic children: BCAS dataset

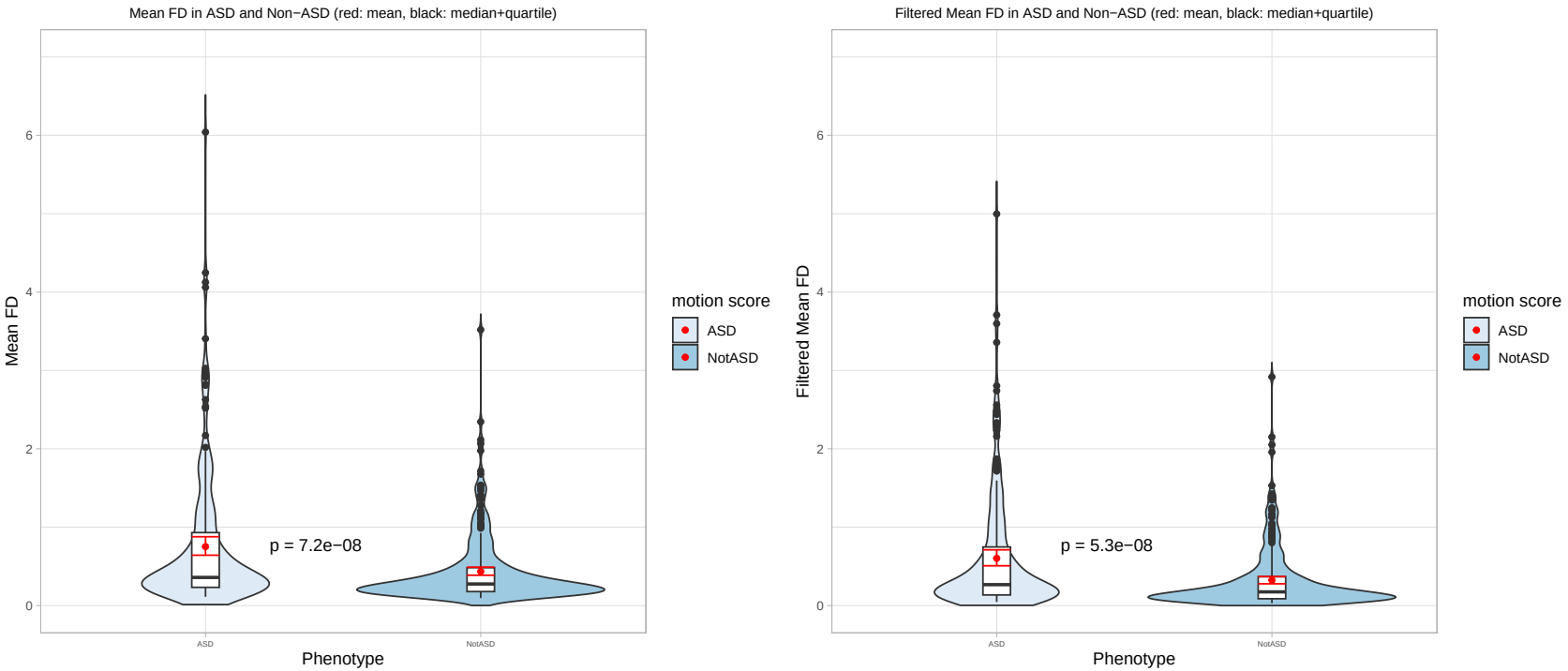


Figure: 8-13 year-olds from an ongoing study at Emory.

Participant removal creates selection bias in ASD

M.B. Nebel, D.E. Lidstone, L. Wang et al. NeuroImage 257 (2022) 119296

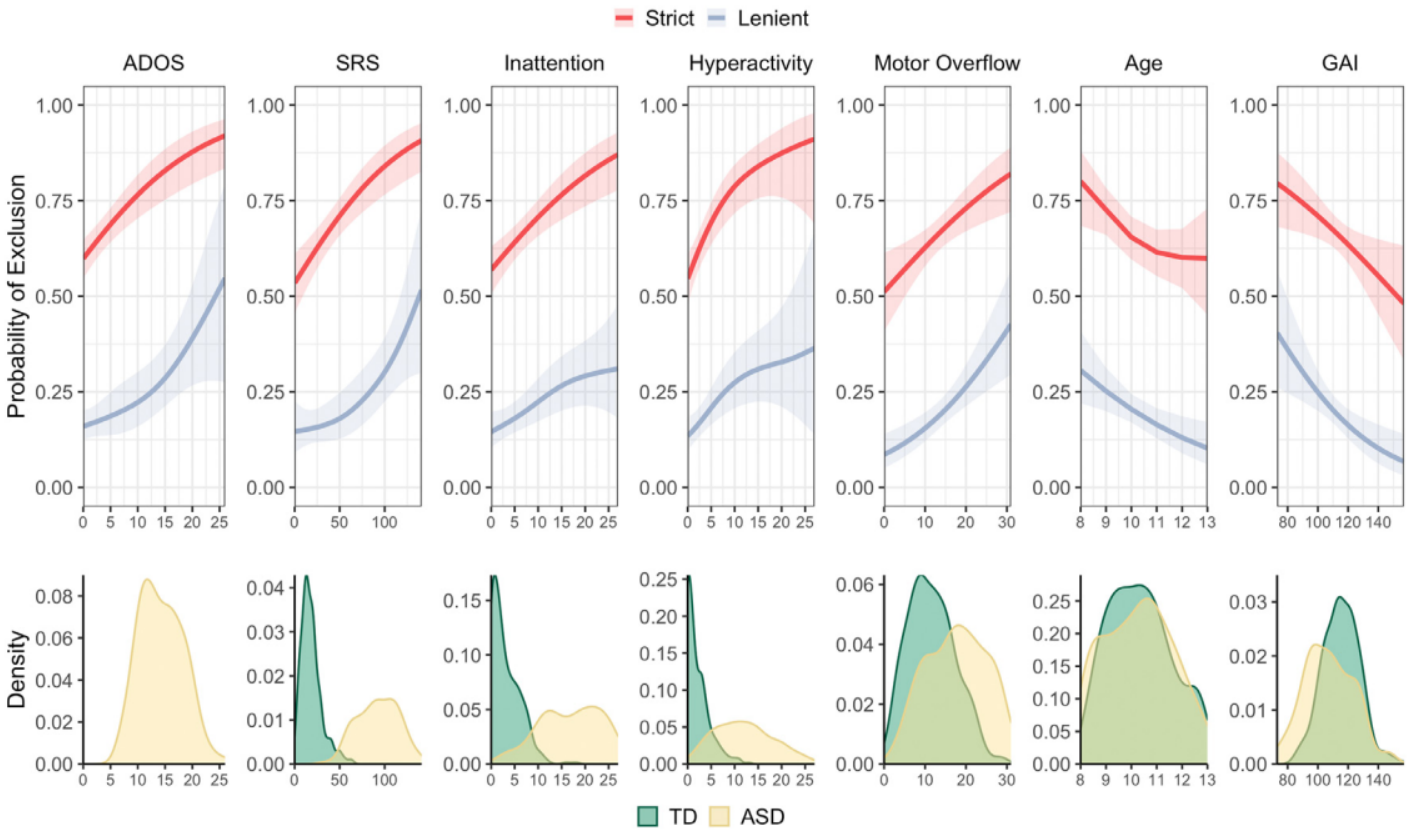
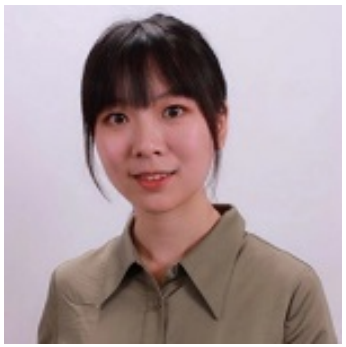


Fig. 4. rs-fMRI exclusion probability changes with phenotype and age. Univariate analysis of rs-fMRI exclusion probability as a function of participant characteristics. From left to right: Autism Diagnostic Observation Schedule (ADOS) total scores, social responsiveness scale (SRS) scores, inattentive symptoms, hyperactive/impulsive symptoms, total motor overflow, age, and general ability index (GAI) using the lenient (slate blue lines, all FDR-adjusted $p < 0.01$), and strict (red lines) motion quality control (all FDR-adjusted $p < 0.03$). Variable distributions for each diagnosis group (included and excluded scans) are displayed across the bottom panel (TD=typically developing, green; ASD=autism spectrum disorder, yellow).

Motion Control (MoCo) in Functional Connectivity Studies in Children with Autism Spectrum Disorder

Jialu Ran[1], Sarah Shultz[2], Benjamin Risk[1] and David Benkeser[1]

[1] *Department of Biostatistics and Bioinformatics,* [2] *Department of Pediatrics, Emory University*

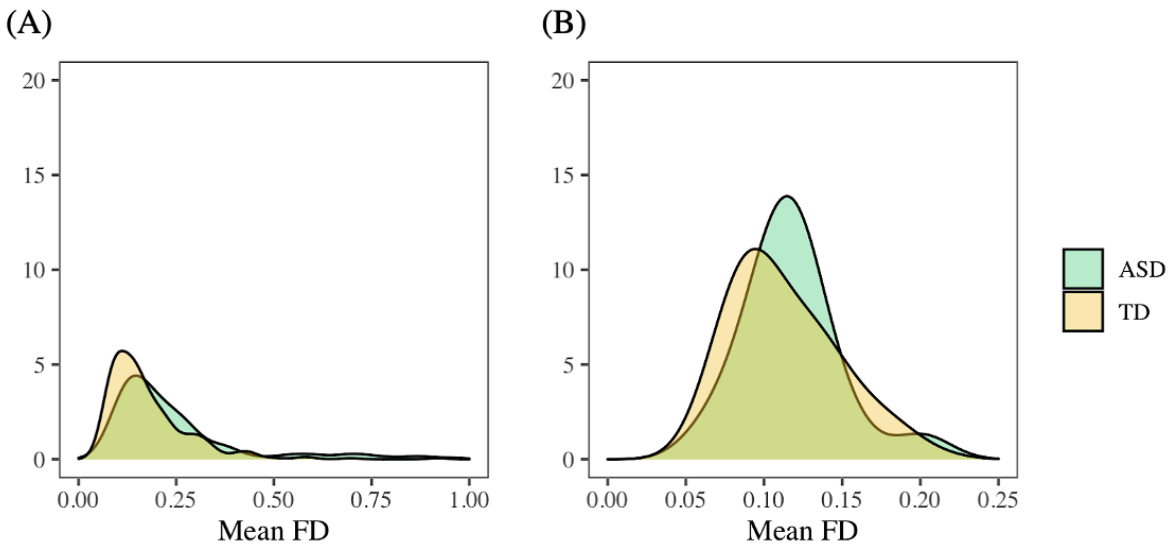


Stochastic Intervention

What would the autistic brain look like under low motion?

We use the framework of stochastic intervention (Díaz et al. 2021).

Replace mean framewise displacement M by tolerable value $M \sim P_{M|\Delta=1,0,X}$



Motion-Controlled Estimand (MoCo)

Define Z: variables that should differ between autistic and non-ASD children, e.g., autism severity.

Motion-Controlled Estimand (MoCo): the difference between the autististic and non-ASD brain when all children have a tolerable amount of motion.

$$\int \big[\{\mu_{Y|A,M,X,Z}(1,m,x,z)\, p_{Z|A,X}(z|1,x) - \mu_{Y|A,M,X,Z}(0,m,x,z)\, p_{Z|A,X}(z|0,x)\} \; p_{M|\Delta=1,A,X}(m|0,x)\, p_X(x)\big]\, dz\, dm\, dx\,.$$

Intuition: estimating what brain looks like under the counterfactual of low motion.

Estimation: one-step estimator

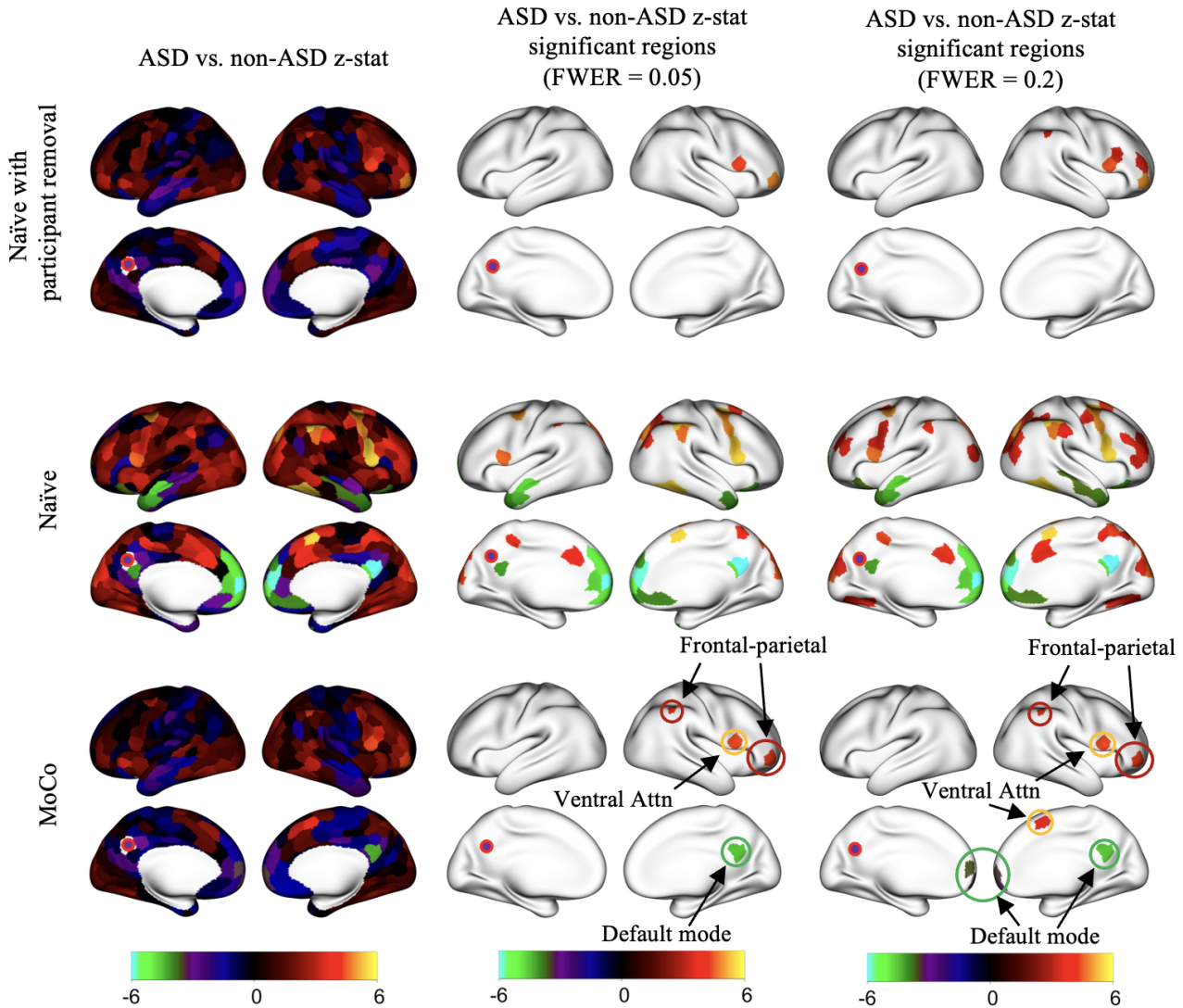
- We define a one-step estimator.
- Allows for statistical inference with machine-learning.
- Machine-learning of nuisance regressions:
  - Motion density estimation using highly adaptive lasso.
  - SuperLearner to learn conditional means and propensities: multivariate adaptive regression splines, LASSO, ridge regression, generalized additive models, generalized linear models, random forest, and xgboost.
- Multiple robustness.
- Simultaneous confidence bands via the efficient influence function (family-wise error rate control).

ABIDE Data

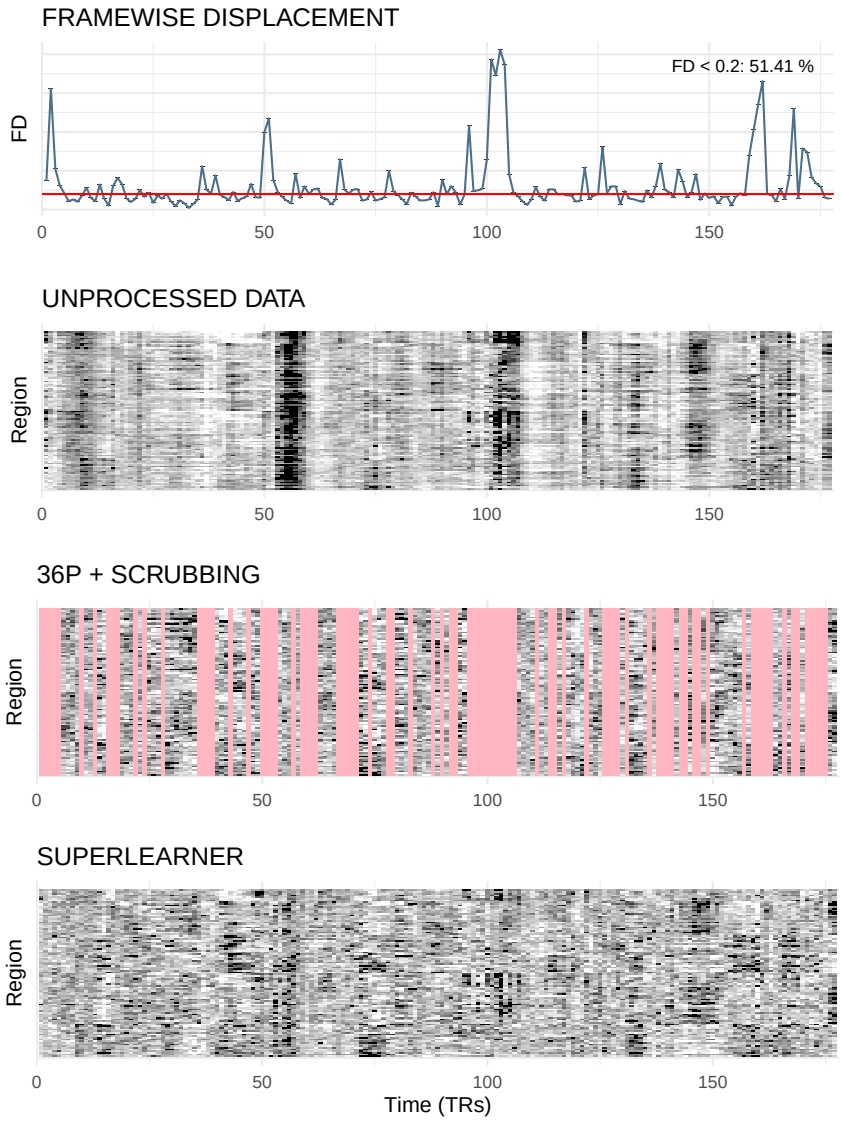
School-age children from Autism Brain Imaging Data Exchange (ABIDEI and ABIDEII) Dataset (Di Martino et al. 2014; 2017)

- variables:
  - A_i: 245 TD ($A = 0$), 132 ASD ($A = 1$) [377 8-13 yo children].
  - X_i: age, sex, handedness.
  - Z_i: autism diagnostic observation schedule, IQ, medication status.
  - M_i: mean frame-wise displacement (FD).
  - $\Delta_i = 1$: > 5 minutes of data free from ≥ 0.2 framewise displacement (Power et al. 2014) [126 TD (51%), 34 ASD (26%)].
- Y_{ij}: correlation between seed region in DMN and region j, $j = 1, \ldots, 400$ (Schaeffer 400 atlas).
- SuperLearner for nuisance regressions: random forest, xgboost, multivariate adaptive regression splines, LASSO, ridge regression, generalized additive models, generalized linear models (with and without interactions, and with and without forward stepwise covariate selection)
- Highly adaptive lasso for conditional density estimation (Hejazi et al. 2022).
- Cross-fitting with 5-fold cross-validation.

ABIDE Inference



Future directions: time series correction



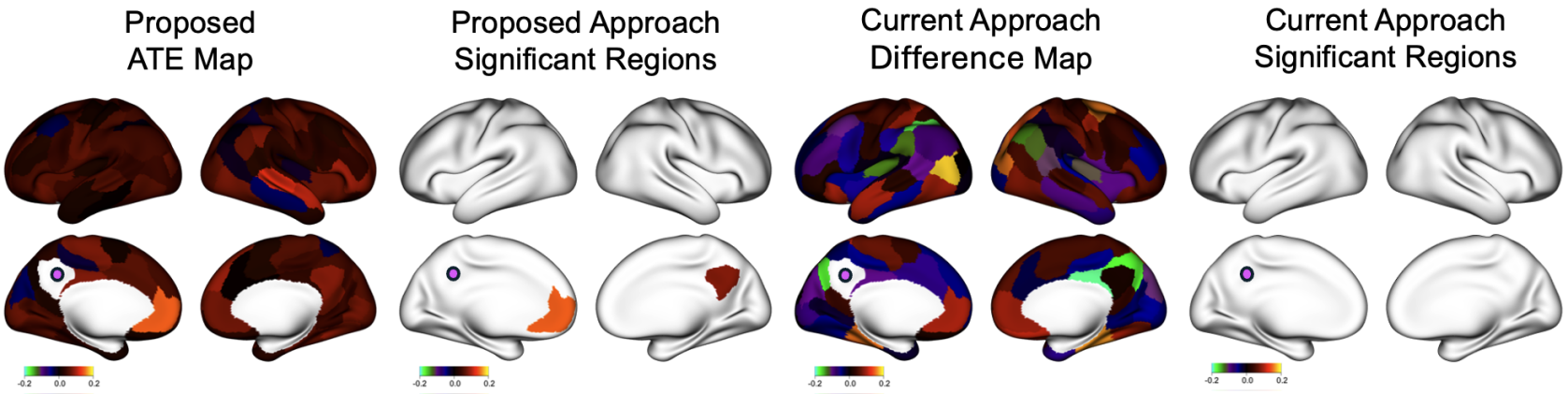


Figure: We have conducted a preliminary analysis using out-of-the-box super learner to remove motion artifacts in the time series of rs-fMRI in a study of autistic children, which demonstrates feasibility. Based on a seed region in the posterior default mode network (DMN) (fuchsia point), connectivity to the anterior DMN increased (FWER $p<0.10$, approximate Cohen's $D=0.85$). This effect was smaller and not significant when using 36p+scrubbing+removal.

Future directions: Brain Connectivity Across the Autism Spectrum

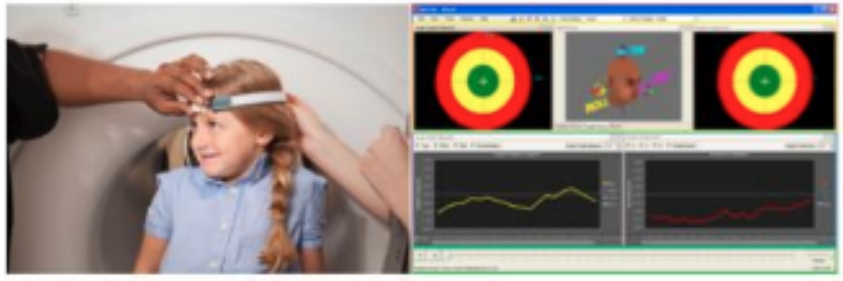
Aim: Characterize functional connectivity across the autism spectrum in school-age children (8-13 years old) through mock training combined with improved statistical methodology.

Previous studies have selection bias in *who* is in the study.
Exclusion criteria:

- Partly driven by site-specific limitations in feasibility.
- Limited training in a mock scanner.
- Exclude non-verbal children.
- Exclude IQ < 80.
- Not stated: exclude children that can't complete a scan.

Brain Connectivity Across the Autism Spectrum at Emory

- Multiple mock training sessions using MoTrak for motion feedback.
- FIRMM to monitor motion during MRI scan.
- Experienced staff: registered behavior technician, licensed masters in social work, licensed associate professional counsel, Applied Behavior Analysis training.
- Beneficial to families.
- brainconnectivitystudy.org



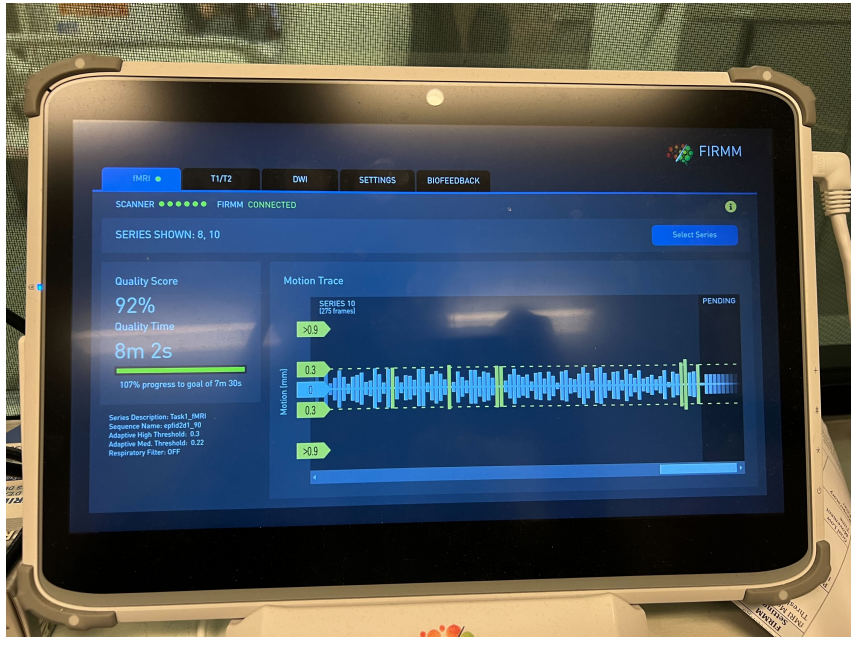


Figure: Top: MoTrak trains children to move less. FIRMM provides real-time motion feedback.

Reducing selection bias: study participants

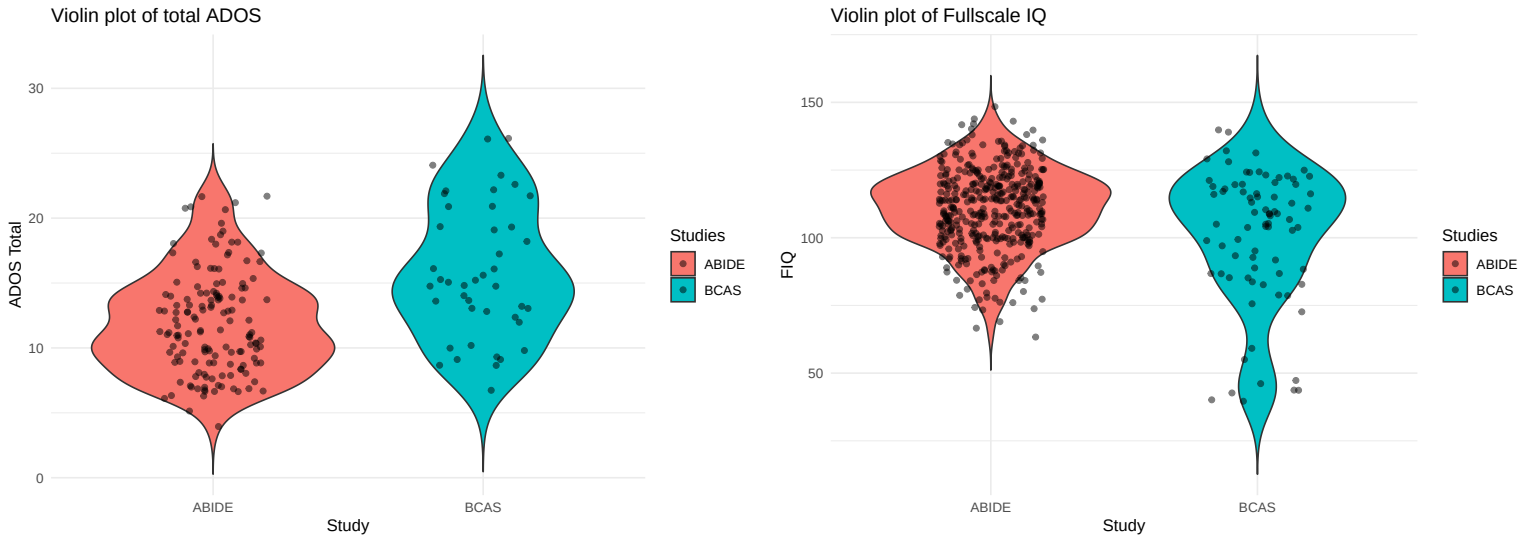


Figure: Comparison of the autism diagnostic observation schedule and full-scale IQ in the ABIDE study versus our study.

- Study sample is more representative of autistic children.
- Improved statistical methodology (MoCo and its extensions) will allow us make better use of the data.

Autism Language Preferences

Figure: Survey designed by Jamie Kortanek.

Marcus Autism Team

Summary

- Statistical inference with machine-learning to estimate what the brain looks like under low motion.
- The decrease in long-range brain connectivity is partially due to motion artifacts.
- MoCo works with imperfectly cleaned correlations to perform additional motion correction.
- Uses ensemble of machine learning methods (random forests, multivariate adaptive regression splines, lasso, xgboost, GAMs) to estimate nuisance regressions.
- Reduces selection bias and improves statistical power.

Thank you!

Thank you!

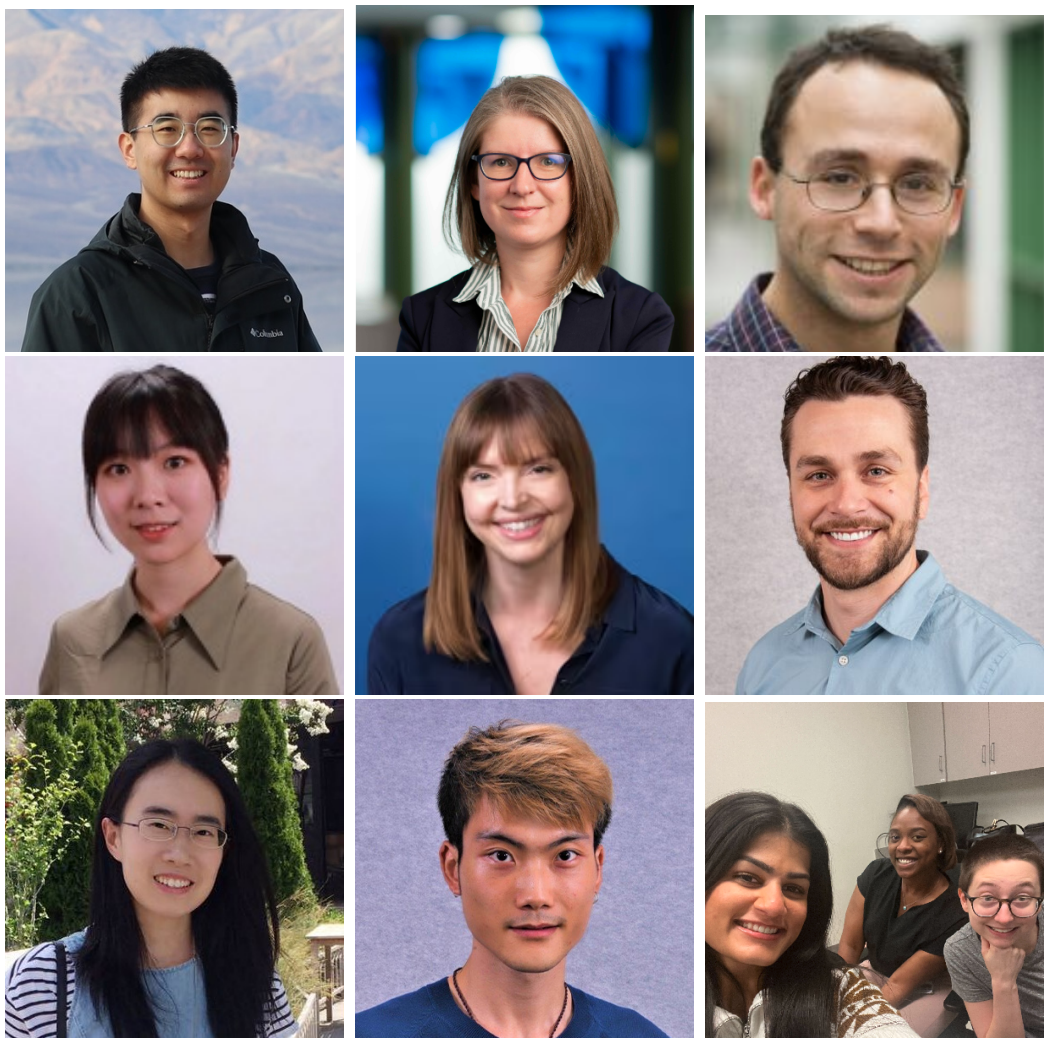


Figure: Zihang Wang (Emory), Irina Gaynanova (UM), Aleksandr (Sasha) Aravkin (UW), Jialu Ran (Emory), Sarah Shultz, David Benkeser, Xiyan Tan, Xucheng (Fred) Huang, Hely Patel, Jamie Kortanek, Ashante Thompson.

Acknowledgments

Jialu Ran, David Benkeser, Sarah Shultz, Kate Revill (FERN Imaging Center at Emory University), Jamie Kortanek, Ashante Thompson, Jennifer Hamel, Hely Patel, Jewel Okoronkwo, Sydney Olson, Sasha Greenspan, Auscia Williams, Malerie McDowell, Cheryl Klaiman, Jose Paredes, Lei Zhou, Mary Beth Nebel, Deqiang Qiu, Jinyu Wang, and Vishwadeep Ahluwalia.

This work supported by National Institute of Mental Health of the National Institutes of Health under award number R01 MH129855. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References I

K. T. Cosgrove, T. J. McDermott, E. J. White, M. W. Mosconi, W. K. Thompson, M. P. Paulus, C. Cardenas-Iniguez, and R. L. Aupperle. Limits to the generalizability of resting-state functional magnetic resonance imaging studies of youth: An examination of abcd study® baseline data. *Brain imaging and behavior*, 16(4):1919–1925, 2022.

B. Deen and K. Pelphrey. Perspective: brain scans need a rethink. *Nature*, 491(7422):S20–S20, 2012.

A. Di Martino, E. Al, and M. P. Milham. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry 2014 19:6*, 19(6):659–667, 6 2013. ISSN 1476-5578. doi: 10.1038/mp.2013.78.

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, et al. 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.

A. Di Martino, D. O'connor, B. Chen, K. Alaerts, J. S. Anderson, M. Assaf, J. H. Balsters, L. Baxter, A. Beggiato, S. Bernaerts, et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. *Scientific data*, 4 (1):1–15, 2017.

I. Díaz, N. S. Hejazi, K. E. Rudolph, and M. J. van Der Laan. Nonparametric efficient causal mediation with intermediate confounders. *Biometrika*, 108(3):627–641, 2021.

N. S. Hejazi, M. J. van der Laan, and D. C. Benkeser. haldensify: Highly adaptive lasso conditional density estimation in R. *Journal of Open Source Software*, 2022. doi: 10.21105/joss.04522. URL https://doi.org/10.21105/joss.04522.

M. A. Just, T. A. Keller, V. L. Malave, R. K. Kana, and S. Varma. Autism as a neural systems disorder: a theory of frontal-posterior underconnectivity. *Neuroscience & Biobehavioral Reviews*, 36(4):1292–1313, 2012.

C. L. Keown, P. Shih, A. Nair, N. Peterson, M. E. Mulvey, and R.-A. Müller. Local functional overconnectivity in posterior brain regions is associated with symptom severity in autism spectrum disorders. *Cell reports*, 5(3):567–572, 2013.

References II

S. Marek, B. Tervo-Clemmens, F. J. Calabro, D. F. Montez, B. P. Kay, A. S. Hatoum, M. R. Donohue, W. Foran, R. L. Miller, T. J. Hendrickson, S. M. Malone, S. Kandala, E. Feczko, O. Miranda-Dominguez, A. M. Graham, E. A. Earl, A. J. Perrone, M. Cordova, O. Doyle, L. A. Moore, G. M. Conan, J. Uriarte, K. Snider, B. J. Lynch, J. C. Wilgenbusch, T. Pengo, A. Tam, J. Chen, D. J. Newbold, A. Zheng, N. A. Seider, A. N. Van, A. Metoki, R. J. Chauvin, T. O. Laumann, D. J. Greene, S. E. Petersen, H. Garavan, W. K. Thompson, T. E. Nichols, B. T. Yeo, D. M. Barch, B. Luna, D. A. Fair, and N. U. Dosenbach. Reproducible brain-wide association studies require thousands of individuals. *Nature 2022 603:7902*, 603(7902): 654–660, 3 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04492-9. URL https://www.nature.com/articles/s41586-022-04492-9.

A. N. Nielsen, D. J. Greene, C. Gratton, N. U. Dosenbach, S. E. Petersen, and B. L. Schlaggar. Evaluating the prediction of brain maturity from functional connectivity after motion artifact denoising. *Cerebral Cortex*, 29(6):2455–2469, 2019.

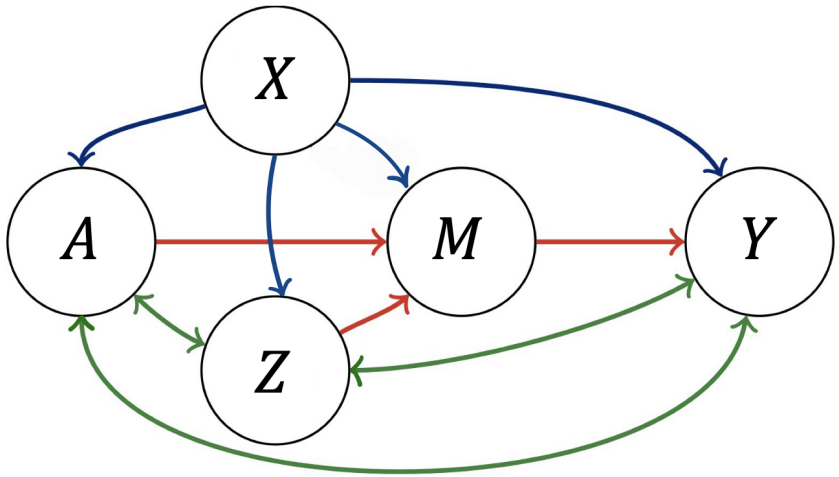
J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen. Methods to detect, characterize, and remove motion artifact in resting state fmri. *Neuroimage*, 84:320–341, 2014.

T. D. Satterthwaite, D. H. Wolf, J. Loughead, K. Ruparel, M. A. Elliott, H. Hakonarson, R. C. Gur, and R. E. Gur. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage*, 60(1):623–632, 2012.

B. E. Yerys, E. M. Gordon, D. N. Abrams, T. D. Satterthwaite, R. Weinblatt, K. F. Jankowski, J. Strang, L. Kenworthy, W. D. Gaillard, and C. J. Vaidya. Default mode network segregation and social deficits in autism spectrum disorder: Evidence from non-medicated children. *NeuroImage: Clinical*, 9:223–232, 2015.

Causal mediation approach: a controversial graph

Our approach treats motion as a mediator:



- $Y \in \mathbb{R}$: functional connectivity between two locations in the brain
- $A \in \{0, 1\}$: non-ASD (0), ASD (1)
- $M \in \mathbb{R}$: a motion variable (mean framewise displacement)
- X: demographic variables (age, sex and handedness)
- Z: variables related to autism symptomatology (autism diagnostic score, IQ, medication status)

Non-parametric model

Initially, consider flexibly modeling motion with non-parametric model:

$$Y_i = \mu_{Y|A,M,X}(A_i, M_i, X_i) + \epsilon_i \ ,$$

where $\mu_{Y|A,M,X}$ is an abitrary function of (A, M, X).

- The effect of A_i is no longer straightforward:

$$\mu_{Y|A,M,X}(1, m, x) - \mu_{Y|A,M,X}(0, m, x)$$

  is a multivariate function of m and x, hard to interpret.
- Non-parametric inference is challenging and slow, not $\sqrt{n}$
- Can't set $m = 0$: all children move at least a little bit.

Marginal effects

This motivates us to consider marginal effects, equivalent to the Average Treatment Effect (ATE):

$$\int \left\{\mu_{Y|A,M,X}(1,m,x) - \mu_{Y|A,M,X}(0,m,x)\right\} p_{M,X}(m,x)\,dm\,dx.$$

This standardizes mean FD to the same level in both groups.

- But includes motion artifacts.
- Neuronal signal may be washed out by motion artifacts.

MoCo Absolute Continuity Assumption

$$\int [\{\mu_{Y|A,M,X,Z}(1,m,x,z)\, p_{Z|A,X}(z|1,x) - \mu_{Y|A,M,X,Z}(0,m,x,z)\, p_{Z|A,X}(z|0,x)\} p_{M|\Delta=1,A,X}(m|0,x)\, p_X(x)]\, dz\, dm\, dx\,.$$

Assumption (Common support (positivity))

Let $P(B)$ be the probability measure of $\{Y, A, M, X, Z\}$ on some set B. Let $P^*(B)$ be the measure corresponding to the joint density $p_{Y|A,M,X,Z}(y|a,m,x,z)\, p_{Z|A,X}(z|a,x)\, p_{M|\Delta=1,A,X}(m|0,x)\, p_x(X)$.

We assume $P^* \ll P$.

$P^*(B) > 0 \implies P(B) > 0$ ensures that it is possible to observe functional connectivity Y in the ASD group across combinations of tolerable motion levels M and more severe symptomatology Z.

Identification

Theorem (Identifiability)

Under the following assumptions:

(A1) *No missing confounders:* $E_C\{Y(m) \mid A=a, X, Z\} = E_C\{Y(m) \mid A=a, M=m, X, Z\}$;

(A2) *Positivity:*

(A2.1) *for every x such that $p_X(x) > 0$, we also have $p_{a|X}(x) > 0$ for $a = 0, 1$;*

(A2.2) *for every (x, z, m) such that $p_X(x)\, p_{Z|a,X}(z \mid x)\, p_{M|\Delta=1,0,X}(m \mid x) > 0$, we also have that $p_{M|a,X,Z}(m \mid x, z) > 0$ for $a = 0, 1$.*

(A3) *Causal Consistency: for any child with observed motion value $M = m$, the observed functional connectivity measurement Y is equal to the counterfactual functional connectivity measurement $Y(m)$.*

The counterfactual $\theta_{C,a}$ is identified by θ_a, where

$$\theta_{\mathbf{1}} = \iiint \mu_{Y|\mathbf{1},M,X,Z}(m,x,z)\; p_{Z|\mathbf{1},X}(z \mid x)\; p_{M|\Delta=1,\mathbf{0},X}(m \mid x)\; p_X(x) dz\, dm\, dx$$

$$\theta_{\mathbf{0}} = \iiint \mu_{Y|\mathbf{0},M,X,Z}(m,x,z)\; p_{Z|\mathbf{0},X}(z \mid x)\; p_{M|\Delta=1,\mathbf{0},X}(m \mid x)\; p_X(x) dz\, dm\, dx$$

Theorem (Efficient Influence Function)

Define

$$\pi_a(x) = P(A = a|X = x),$$
$$\bar{\pi}_0(x) = P(A = 0|X = x)P(\Delta = 1 \mid A = 0, X = x),$$
$$r_a(m, x, z) = \frac{p_{M|\Delta=1,0,X}(m \mid x)}{p_{M|a,X,Z}(m \mid x, z)},$$
$$\eta_{\mu|A,Z,X}(a, z, x) = \int \mu_{Y|A,M,X,Z}(a, m, x, z)\, p_{M|\Delta=1,A,X}(m \mid 0, x)\, dm\,,$$
$$\xi_{a,\eta|X}(x) = \iint \mu_{Y|A,M,X,Z}(a, m, x, z)\, p_{M|\Delta=1,A,X}(m \mid 0, x)\, p_{Z|A,X}(z \mid a, x)\, dm\, dz\,.$$

In a nonparametric model, the efficient influence function for θ_a is

$$\begin{aligned} D_{P,a}(O_i) = {} & \frac{\mathbf{1}_a(A_i)}{\pi_a(X_i)} r_a(M_i, X_i, Z_i)\{Y_i - \mu_{Y|a,M,X,Z}(M_i, X_i, Z_i)\} \\ & + \frac{\mathbf{1}_a(A_i)}{\pi_a(X_i)}\{\eta_{\mu|a,Z,X}(X_i, Z_i) - \xi_{\eta|a,X}(X_i)\} \\ & + \frac{\mathbf{1}_{a,1}(A_i, \Delta_i)}{\bar{\pi}_0(X_i)}\left\{\eta_{\mu|a,M,X}(M_i, X_i) - \xi_{\eta|a,X}(X_i)\right\} + \xi_{\eta|a,X}(X_i) - \theta_a. \end{aligned}$$

$\sqrt{n}$-Convergence

Theorem (Asymptotic normality)

Under the following assumptions,

(i) *Positivity of estimates:* $\pi_{n,a} > \epsilon_1$ *for some* $\epsilon_1 > 0$, $\bar{\pi}_{n,0} > \epsilon_2$ *for some* $\epsilon_2 > 0$, *and* $\frac{p_{n,M|\Delta=1,0,X}}{p_{n,M|a,X,Z}} < \epsilon_3$ *for some* $\epsilon_3 < \infty$;

(ii) $n^{1/2}$*-convergence of second order terms...*

(iii) $L^2(P)$*-consistent influence function estimate:*

$$\int [\{D_{a,P_\ell}(o) - D_{a,P_n}(o)\}^2] dP(o) = o_P(1),$$

where P_ℓ *denotes the limit of* P_n *as* $n \to \infty$.

(iv) *Donsker influence function estimate:* D_{a,P_n} *falls in a* P*-Donsker class with probability tending to 1.*

Then,

$$\theta^+_{n,a} - \theta_a = \frac{1}{n} \sum_{i=1}^{n} D_{a,P}(O_i) + o_P(n^{-1/2})$$

and

$$n^{1/2}(\theta^+_{n,a} - \theta_a) \Rightarrow N(0, E[D_{P,a}(O)^2]) \,.$$

Multiple robustness

	$\mu_{n,Y\|A,M,X,Z}$	$\eta_{n,\mu\|A,M,X}$	$\xi_{n,a,\eta\|X}$	$\bar{\pi}_{n,0}$	$\pi_{n,a}$	$p_{n,M\|\Delta=1,A,X}$	$p_{n,M\|A,X,Z}$
(B2.1)					✓	✓	✓
(B2.2)			✓			✓	✓
(B2.3)	✓	✓		✓	✓		
(B2.4)	✓				✓	✓	
(B2.5)	✓		✓			✓	

Table: Theorem: multiple robustness. Each row indicates a setting for consistency, where check marks indicate the nuisance parameters which, when they converge to true functions, then $E[D_{P',a}(O)] = 0$, and $\theta^{+}_{n,a} \to \theta_a$.

Estimation, 1/3

1. *Estimate mean functional connectivity* $\mu_{Y|A,M,X,Z}$. Fit a super learner regression using Y as the outcome and including A, M, X, and Z as predictors. Evaluate the fitted value for $i = 1, \ldots, n$ and for $a = 0, 1$.

2. *Estimate motion distributions* $p_{M|A,X}$, $p_{M|\Delta=1,A,X}$, $p_{M|A,X,Z}$, and $p_{M|\Delta=1,A,X,Z}$. Estimate densities using the highly adaptive LASSO and evaluate for $a = 0, 1$ and $i = 1, \ldots, n$.

3. *Estimate motion-standardized functional connectivity* $\eta_{\mu|A,Z,X}$. Create the pseudo-outcome $\hat{Y}_{M,i} = \mu_{n,Y|A,M,X,Z}(A_i, M_i, X_i, Z_i) \times \frac{p_{n,M|\Delta=1,A,X}(M_i|0,X_i)}{p_{n,M|\Delta=1,A,X,Z}(M_i|A_i,X_i,Z_i)}$. Using only observations with $\Delta_i = 1$, fit a super learner regression using $\hat{Y}_M$ as the outcome and A, Z, and X as predictors. Set A to a evaluate for $i = 1, \ldots, n$.

Estimation, 2/3

4. *Estimate Z-standardized functional connectivity* $\eta_{\mu|A,M,X}$. Create the pseudo-outcome $\hat{Y}_{Z,i} = \mu_{n,Y|A,M,X,Z}(A_i, M_i, X_i, Z_i) \times \frac{p_{n,M|A,X}(M_i|A_i,X_i)}{p_{n,M|A,X,Z}(M_i|A_i,X_i,Z_i)}$. Fit a super learner regression using $\hat{Y}_Z$ as the outcome and including M, X, and A as predictors. Set A to a and evaluate for $i = 1, \ldots, n$.

5. *Estimate motion- and Z-standardized functional connectivity* $\xi_{a,\eta|X}$. Fit a super learner regression using $\eta_{n,\mu|A,Z,X}$ as the outcome and including A and X as predictors. For $a = 0, 1$, evaluate the fitted value for $i = 1, \ldots, n$.

6. *Calculate plug-in estimate.* Compute the plug-in estimate $\theta_{n,a} = n^{-1} \sum_{i=1}^{n} \xi_{n,a,\eta|X}(X_i)$.

Estimation, 3/3

7. Estimate propensities.

i. *Estimate diagnosis distribution* π_a Fit a super learner regression using A as the outcome and including X as predictors. Evaluate for $i = 1, \ldots, n$ and set $\pi_{n,0}(X_i) = 1 - \pi_{n,1}(X_i)$.

ii. *Estimate inclusion probability* $\pi_{\Delta=1|A,X}$. Fit a super learner using Δ as the outcome and including A and X as predictors. Set A to 0 to obtain $\pi_{n,\Delta=1|A,X}(0, X_i)$ for $i = 1, \ldots, n$. Compute $\bar{\pi}_{n,0}(X_i) = \pi_{n,0}(X_i)\pi_{n,\Delta=1|A,X}(0, X_i)$ for $i = 1, \ldots, n$.

8. *Evaluate estimated efficient influence function* $D_{n,a}(O_i)$. For $a = 0, 1$ and each $i = 1, \ldots, n$, evaluate $D_{n,a}(O_i)$ by substituting the fitted values based on the estimated nuisance parameters obtained in steps 1-7 into equation (2).

9. *Compute the one-step estimator.* For $a = 0, 1$, compute $\theta^+_{n,a} = \theta_{n,a} + n^{-1} \sum_{i=1}^{n} D_{n,a}(O_i)$.

Simulation: Confirming theoretical properties of estimators

- Simulation Setting
  $X \sim \text{Bin}(1, \frac{1}{2})$
  $A \sim \text{Bin}(1, \text{expit}(X - \frac{1}{4}))$
  $Z \sim \text{Bin}(1, \text{expit}(\frac{5}{4}A - \frac{1}{2}))$
  $M \sim \text{N}(1 + A + X/2 - Z/4, 1)$
  $Y \sim \text{N}(-1 + X/2 - Z/3 - A/4 + M/5, 1)$
  sample size $n \in \{200, 500, 1000, 2000, 4000\}$
- evaluate proposed estimators of θ_0 and θ_1

Simulation: Confirming theoretical properties of estimators

Case I: all nuisance parameters are consistently estimated at appropriate rates

n	$\theta^{\text{cf}}_{n,0}$				$\theta^{\text{cf}}_{n,1}$			
	$n^{1/2}$ bias	$n^{1/2}$ sd	sd ratio	cover	$n^{1/2}$ bias	$n^{1/2}$ sd	sd ratio	cover
200	-0.235	2.063	1.075	0.929	-0.246	2.358	1.430	0.851
500	-0.150	1.938	0.986	0.951	-0.310	2.369	1.205	0.900
1000	-0.141	2.003	1.028	0.940	-0.113	2.333	1.110	0.922
2000	-0.026	1.977	1.033	0.940	-0.077	2.328	1.056	0.931
4000	-0.006	1.913	1.014	0.950	0.076	2.074	0.914	0.979

Table: All nuisance parameters are consistently estimated at appropriate rates with the use of cross-fitting