

Counteracting Selection Bias and Improving Statistical Power in Functional Connectivity Studies of Autism

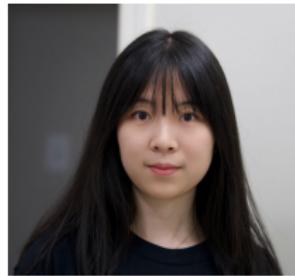
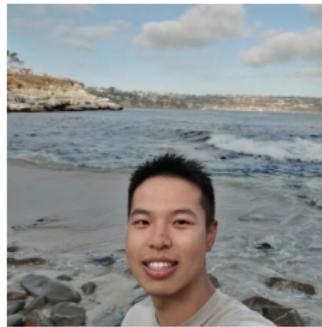
Benjamin Risk

benjamin.risk@emory.edu



Collaborators

The project team includes students Liangkang Wang, Jialu Ran, Zihang Wang, and Jinyu Wang, and co-investigators David Benkeser, Cheryl Klaiman, Razieh Nabi, Deqiang Qiu, and Sarah Shultz. This talk is based on based on a master's thesis by **Liangkang Wang**.

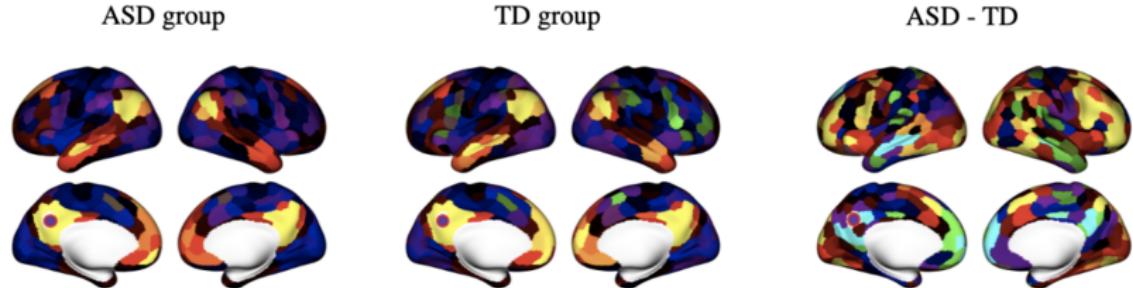


Goal of this talk

- Overview of our approach for accounting for selection bias and improving statistical power in fMRI studies.
- Simulations comparing AIPWE to DRTMLE in smaller sample sizes.
- Propose a computationally scalable permutation test.
- Results on school-age children in ABIDE.

Background on fMRI

- Resting-state fMRI: for about six minutes, lie awake and motionless in a scanner staring at a cross-hair
- Measure spontaneous brain activity across time.
- Undirected correlations between brain regions are called **functional connectivity**.
- seed region: correlation between one region and all other regions.



Problem part I: Motion has huge impacts in fMRI

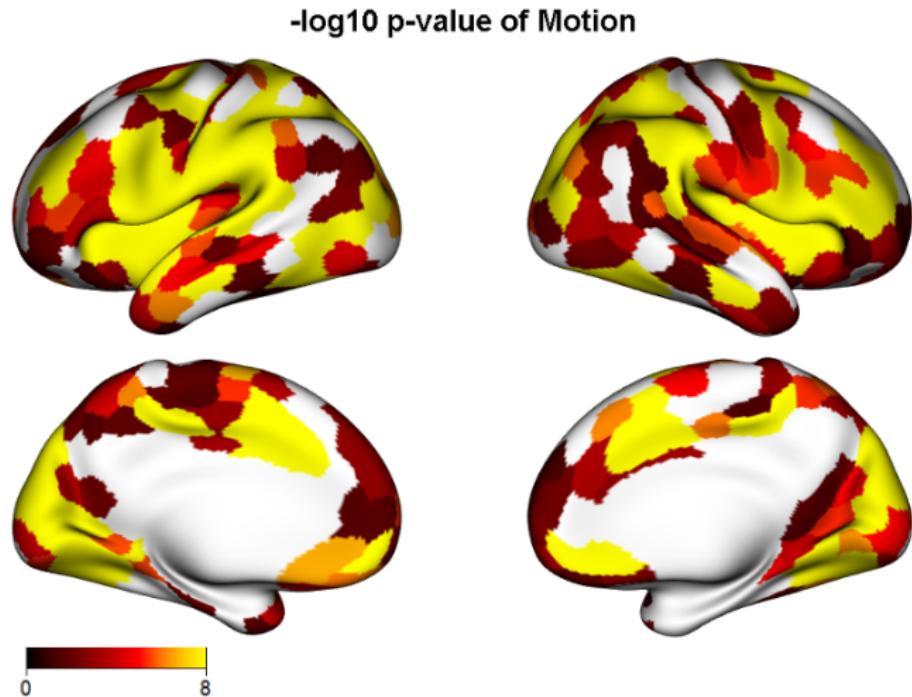


Figure: -log10 pvalues of effect of motion on functional connectivity from GAMs. The current view is that regression-based techniques are insufficient.

Motivation: Removal of high motion participants

- Motion in the scanner produces artifacts. Regression techniques for removal are insufficient ([Power et al., 2012](#)).
- Studies recommend excluding “high” motion participants.
- ABCD study with school-aged children removed 60 – 75% of children due to excessive motion ([Marek et al., 2022](#); [Nielsen et al., 2019](#)). Our previous study on ASD children removed 20 – 60% ([Nebel et al., 2022](#)).
- Our previous study found more extensive differences between ASD and TD when using DRTMLE ([Nebel et al., 2022](#)), relying upon asymptotic inference and using a lenient criteria.
- In the current study, we examine methods for smaller sample sizes with both strict and lenient criteria.

School-age children data set

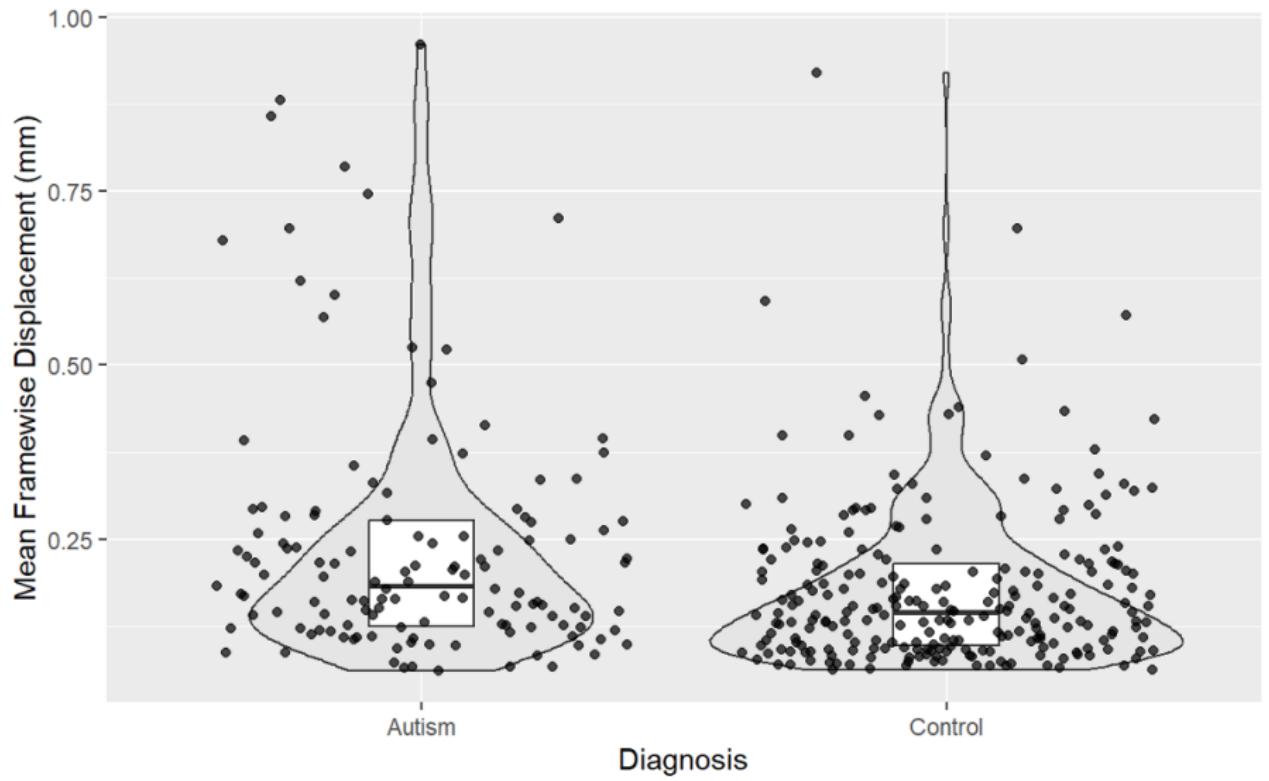
- Autism spectrum disorder is a neurodevelopmental disorder characterized by challenges in social interactions, communication, and repetitive behaviors.
- We are interested in school-aged children (8-13 years).
- We selected school-age children from the ABIDE I and II datasets, focusing on the two sites that had > 20 school-age children.

Demographic Table for the Current Study

	TD (N=252)	ASD (N=144)	Total (N=396)	p value
Age				0.680
Mean (SD)	10.400 (1.347)	10.338 (1.594)	10.378 (1.440)	
Range	8.010 - 13.720	8.014 - 13.950	8.010 - 13.950	
Gender				0.003
Male	174 (69.0%)	119 (82.6%)	293 (74.0%)	
Female	78 (31.0%)	25 (17.4%)	103 (26.0%)	
FIQ				< 0.001
Mean (SD)	114.627 (11.507)	103.625 (17.477)	110.626 (14.926)	
Range	80.000 - 144.000	63.000 - 148.000	63.000 - 148.000	
Handedness				0.289
Right	235 (93.3%)	130 (90.3%)	365 (92.2%)	
Left	17 (6.7%)	14 (9.7%)	31 (7.8%)	
ADOS				< 0.001
Mean (SD)	0.000 (0.000)	13.403 (5.245)	4.874 (7.186)	
Range	0.000 - 0.000	6.000 - 35.000	0.000 - 35.000	
Currently on Stimulants				< 0.001
No	252 (100.0%)	116 (80.6%)	368 (92.9%)	
Yes	0 (0.0%)	28 (19.4%)	28 (7.1%)	
Currently on NonStimulants				< 0.001
No	251 (99.6%)	118 (81.9%)	369 (93.2%)	
Yes	1 (0.4%)	26 (18.1%)	27 (6.8%)	
Site ID				< 0.001
ABIDEI-KKI	33 (13.1%)	22 (15.3%)	55 (13.9%)	
ABIDEI-NYU	44 (17.5%)	43 (29.9%)	87 (22.0%)	
ABIDEII-KKI_1	155 (61.5%)	56 (38.9%)	211 (53.3%)	
ABIDEII-NYU_1	20 (7.9%)	23 (16.0%)	43 (10.9%)	
ciric				< 0.001
Unusable	39 (15.5%)	45 (31.2%)	84 (21.2%)	
Usable	213 (84.5%)	99 (68.8%)	312 (78.8%)	
powerpt2				< 0.001
Unusable	126 (50.0%)	110 (76.4%)	236 (59.6%)	
Usable	126 (50.0%)	34 (23.6%)	160 (40.4%)	

Table: Socio-demographic characteristics

Motion distributions in ASD and TD children



Selection bias

- $Y(\Delta = 1)$, i.e., $Y(1)$, is the counterfactual that a participant's scan is usable. Define associational parameter (Nebel et al., 2022):

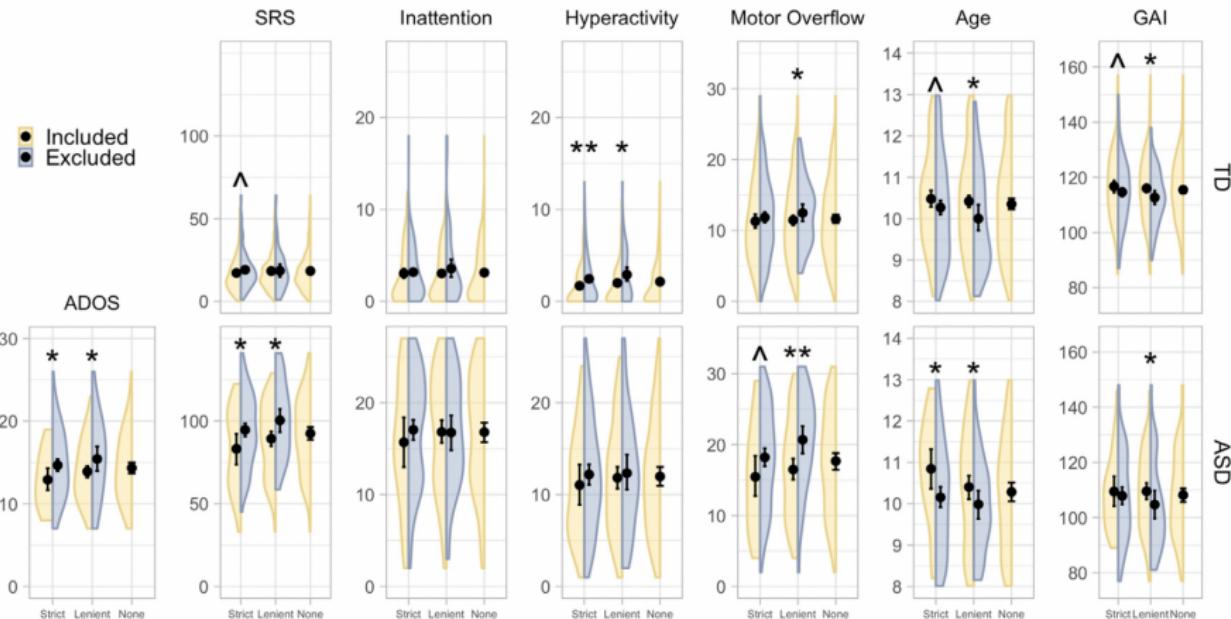
$$\begin{aligned}\psi^* &= E^*(Y(1)|A = 1) - E^*(Y(1)|A = 0) \\ &= E^*\{E^*(Y(1)|A = 1, W)|A = 1\} \\ &\quad - E^*\{E^*(Y(1)|A = 0, W)|A = 0\}.\end{aligned}$$

- Compare this to the naive mean difference:

$$\begin{aligned}\psi_{naive} &= E\{E(Y|\Delta = 1, A = 1, W)|\Delta = 1, A = 1\} \\ &\quad - E\{E(Y|\Delta = 1, A = 0, W)|\Delta = 1, A = 0\}.\end{aligned}$$

- Selection bias: $\psi_{naive} \neq \psi^*$.
- Key: lack of exchangeability between usable and unusable data.
- Bias can arise when $\Delta \leftrightarrow W$, $W \leftrightarrow Y$. Then
 $E^*[Y(1)|A = 1] \neq E[Y|\Delta = 1, A = 1]$

Selection bias in previous study



Target Parameter and Identifiability Assumptions

- Define our target parameter, the debiased group difference:

$$\begin{aligned}\psi = & E\{E(Y \mid \Delta = 1, A = 1, W) \mid A = 1\} \\ & - E\{E(Y \mid \Delta = 1, A = 0, W) \mid A = 0\}.\end{aligned}$$

- Identifiability assumptions: $\psi^* = \psi$ if

- (A1.1) *Mean exchangeability (no missing confounders)*:
for $a = 0, 1$, $E^*\{Y(1) \mid A = a, W\} = E^*\{Y(1) \mid \Delta = 1, A = a, W\}$.
- (A1.2) *Positivity*: for $a = 0, 1$ and all possible w ,
 $P(\Delta = 1 \mid A = a, W = w) > 0$.
- (A1.3) *Causal Consistency*: for all i such that $\Delta_i = 1$, $Y_i(1) = Y_i$.

Notation

Notation:

- Let n_1 be the number of children with ASD
- $\{i \in \mathcal{S}_1\}$ denote the set of indices for ASD children.
- Similarly, define n_0 and $\{i \in \mathcal{S}_0\}$ for the TD children.

IPWE and G-Comp

- Inverse probability weighted estimator:
 - Use **ensemble of machine learning methods** to fit propensity model: $\hat{p}(\Delta_i \mid A_i, W_i)$.
 - Inverse probability weighted estimator for correlation between DMN seed region and **regions $j=1, \dots, 400$** :

$$\hat{\psi}_{j,IPWE} = \frac{1}{n_1} \sum_{i \in \mathcal{S}_1} \left(\frac{\Delta_i}{\hat{p}(A_i, W_i)} Y_{ij} \right) - \frac{1}{n_0} \sum_{i \in \mathcal{S}_0} \left(\frac{\Delta_i}{\hat{p}(A_i, W_i)} Y_{ij} \right).$$

- G-Computation estimator:
 - Fit outcome model with **superlearner** ([Van Der Laan et al., 2007](#)):

$$\hat{Y}_{ij} = \bar{Q}_j(A_i, W_i).$$

- Predict values for all i (usable and unusable):

$$\hat{\psi}_{j,GComp} = \frac{1}{n_1} \sum_{i \in \mathcal{S}_1} \hat{Y}_{ij} - \frac{1}{n_0} \sum_{i \in \mathcal{S}_0} \hat{Y}_{ij}.$$

- The Augmented Inverse Probability Weighted Estimator combines IPWE and G-Computation in missing data ([Bang and Robins, 2005](#)):

$$\hat{\psi}_{j,AIPWE} = \frac{1}{n_1} \sum_{i \in \mathcal{S}_1} \left(\left[\frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] Y_{ij} + \left[1 - \frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] \bar{Q}_j(A_i, W_i) \right) \\ - \frac{1}{n_0} \sum_{i \in \mathcal{S}_0} \left(\left[\frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] Y_{ij} + \left[1 - \frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] \bar{Q}_j(A_i, W_i) \right).$$

- Doubly robust estimate of mean: if either propensity or outcome model is correct, consistent estimator.
- Standard errors are not consistent if model mis-specified (but we will use superlearner to flexibly model propensity and outcome models).

Doubly robust targeted minimum loss based estimation

- Benkeser et al. (2017) developed a doubly robust targeted minimum loss-based estimator: if *at least* one of the two regressions is consistently estimated, both $\hat{\psi}$ and its SE are consistently estimated.
- ➊ Fit propensity model.
 - ➋ Fit outcome model.
 - ➌ Apply DRTMLE to propensities and predicted outcomes. Involves a special iterative logistic regression.
 - ➍ Great theoretical properties – use when you have thousands of participants.

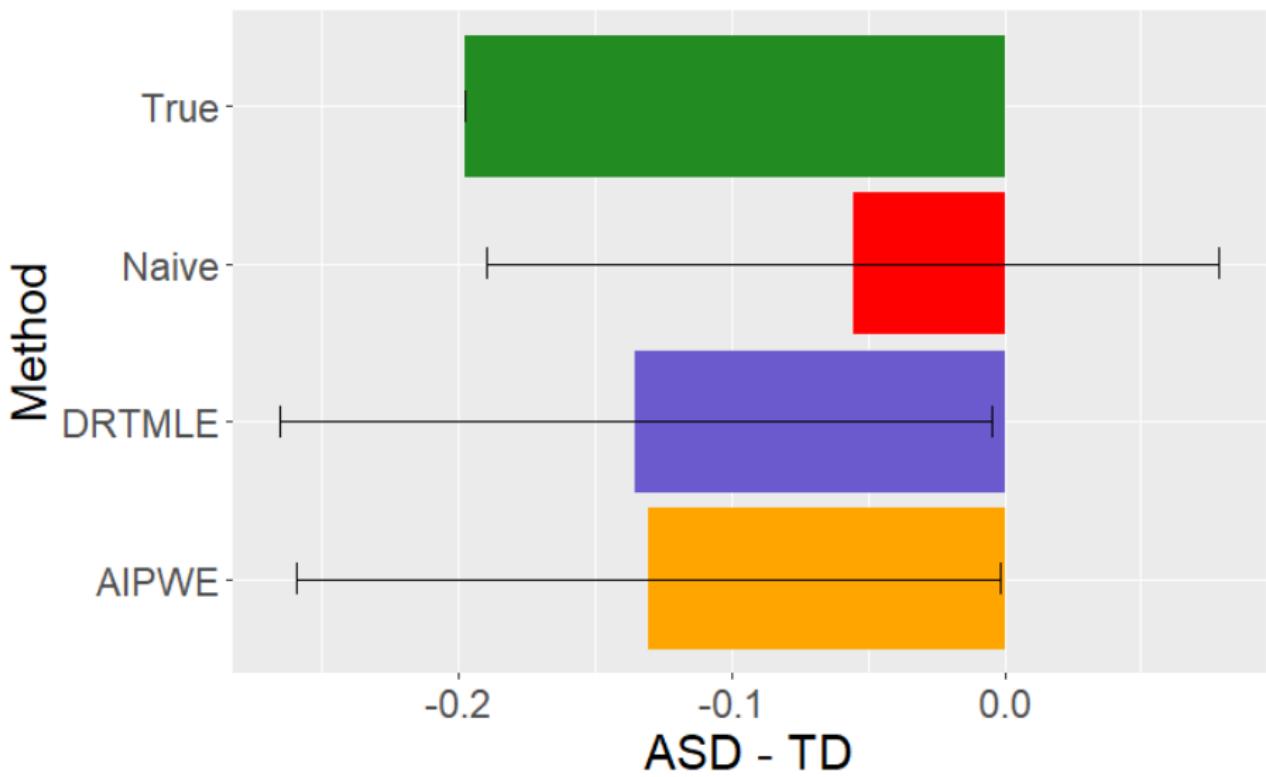


Figure: Estimate of functional connectivity from DRTMLE and AIPWE compared to naive data removal from a toy simulation.

Perm Tests and Computational considerations

- Permutation tests are popular for finite-sample inference in neuroimaging.
- The propensity and outcome models require fitting superlearner with 10-fold CV for propensity and outcome models: `SL.earth`, `SL.glmnet`, `SL.gam`, `SL.glm`, `SL.ranger`, `SL.ridge`, `SL.step`, `SL.step.interaction`, `SL.svm`, `SL.xgboost`.
- CV is sensitive to random seed, fit 20 times for propensity model, average $\hat{p}(A_i, W_i)$, fit 20 times for each of 400 locations, average $\bar{Q}_j(A_i, W_i)$.
- Single permutation: outcome model involves fitting regressions to approximately 400 locations with 20 random seeds and 10 learners and 10-fold CV $\approx 800,000$.

Novel Permutation test

- Computationally scalable: permute membership of \mathcal{S}_1 and \mathcal{S}_0 , call $\mathcal{S}_1^{(k)}$ and $\mathcal{S}_0^{(k)}$.

$$\hat{\psi}_{j,AIPWE}^{(k)} = \frac{1}{n_1} \sum_{i \in \mathcal{S}_1^{(k)}} \left(\left[\frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] Y_{ij} + \left[1 - \frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] \bar{Q}_j(A_i, W_i) \right)$$
$$- \frac{1}{n_0} \sum_{i \in \mathcal{S}_0^{(k)}} \left(\left[\frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] Y_{ij} + \left[1 - \frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] \bar{Q}_j(A_i, W_i) \right).$$

- Standardize by asymptotic standard error to generate z_j .
- **Family-wise error rate control:**

$$p_{j,fwer} = \frac{1}{K} \sum_{k=1}^K I \left(\left\{ \max_j |z_j^{(k)}| \right\} > |z_j| \right)$$

Novel Permutation test

- Under the null hypothesis,

$$\begin{aligned} & E\{E(Y_j \mid \Delta = 1, A = 1, W) \mid A = 1\} \\ & - E\{E(Y_j \mid \Delta = 1, A = 0, W) \mid A = 0\} = 0. \end{aligned}$$

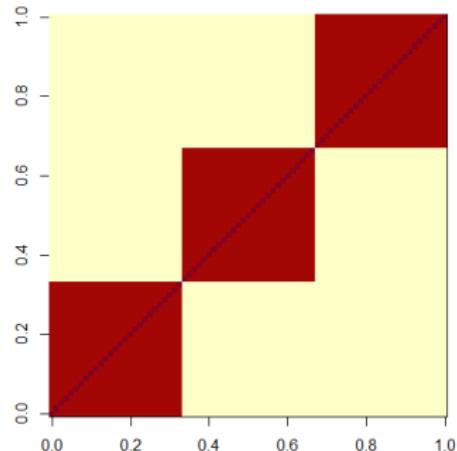
- In the permutation test, we preserve the inner conditional expectation, $E(Y_j \mid \Delta = 1, A, W)$ by using the estimates

$$\left[\frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] Y_{ij} + \left[1 - \frac{I(\Delta_i = 1)}{\hat{p}(A_i, W_i)} \right] \bar{Q}_j(A_i, W_i)$$

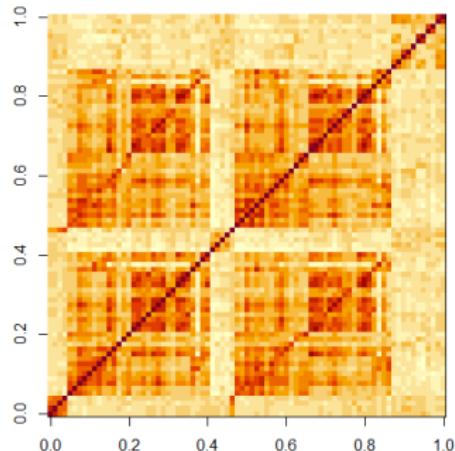
- We obtain a null distribution for our finite sample since

$$\begin{aligned} E\hat{\psi}_{j,AIPWE}^{(k)} &= E\{E(Y_j \mid \Delta = 1, A = 1, W)\} \\ &- E\{E(Y_j \mid \Delta = 1, A = 0, W)\}. \end{aligned}$$

Simulation overview



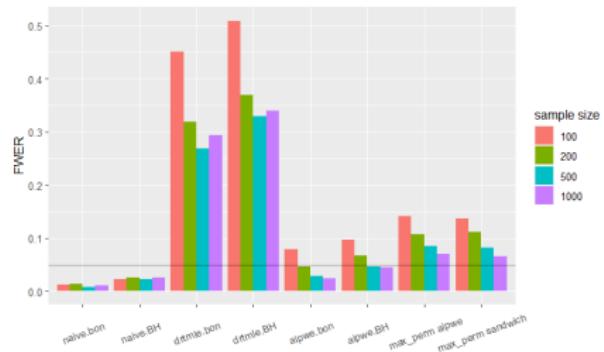
(a) Strong block-wise correlation



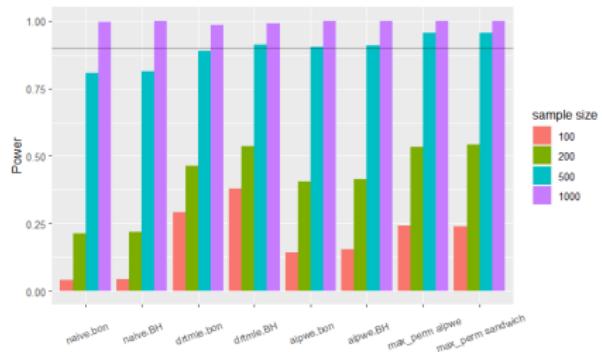
(b) Correlation from a seed-based analysis

Figure: Block structure used in simulation. a) Strong block-wise correlation between 81 regions. b) Correlation from data.

Simulations: strong correlation



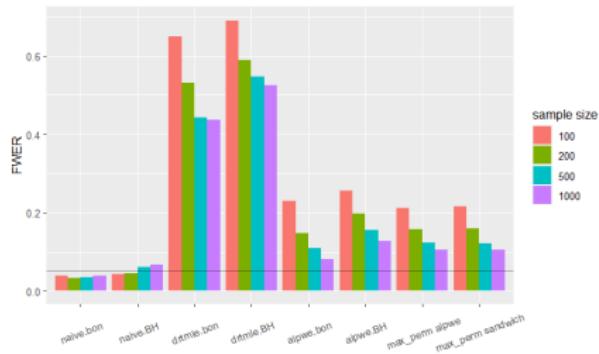
(a) FWER from 81 regions



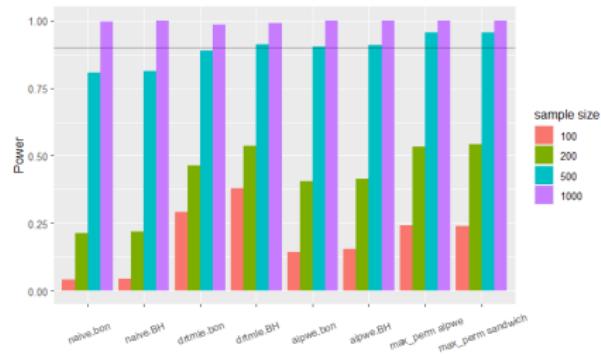
(b) Power

Figure: Simulation setting: within-block correlation = 0.9, three blocks, 81 regions. Inflated FWER in permutation test may be due to exchangeability violations.

Simulations: data-based correlation



(a) FWER from 81 regions



(b) Power

Figure: Simulation setting: data-based correlation between errors.

Simulations summary

- AIPWE has fewer false positives than DRTMLE – the improved (asymptotic) robustness in DRTMLE comes at a cost for $n < 1000$.
- AIPWE has improved power relative to naive removal of scans.
- Max perm didn't show many benefits over AIPWE. Some power gains but mixed results with type 1 errors.
- Adequate FDR control for AIPWE (not shown), still problems with FWER control in some settings.

Previous study: motion exclusion criteria in functional MRI can cause sampling bias

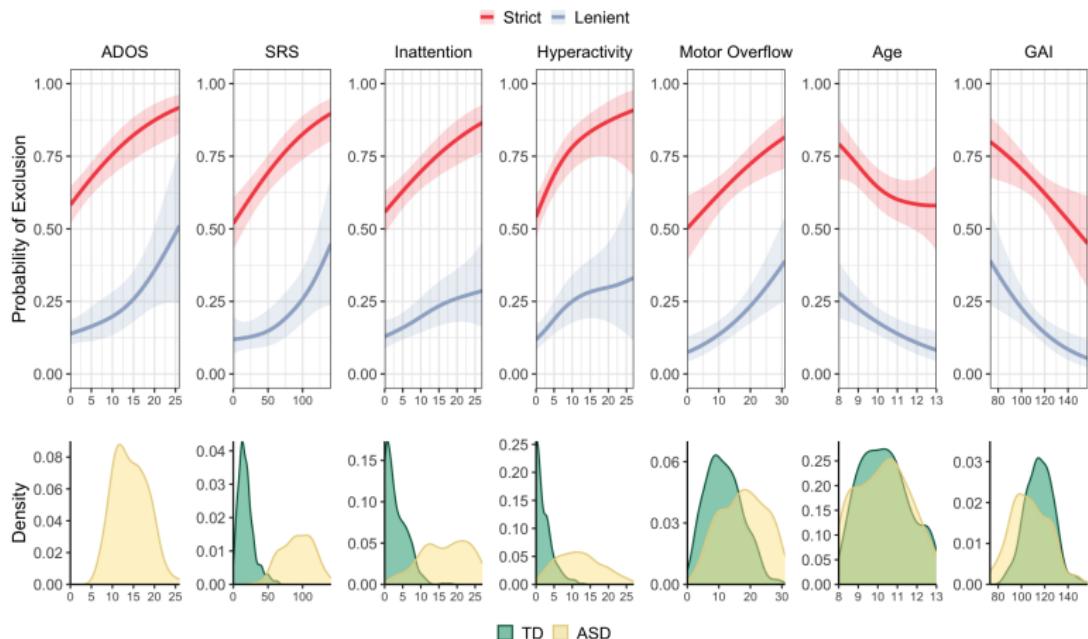
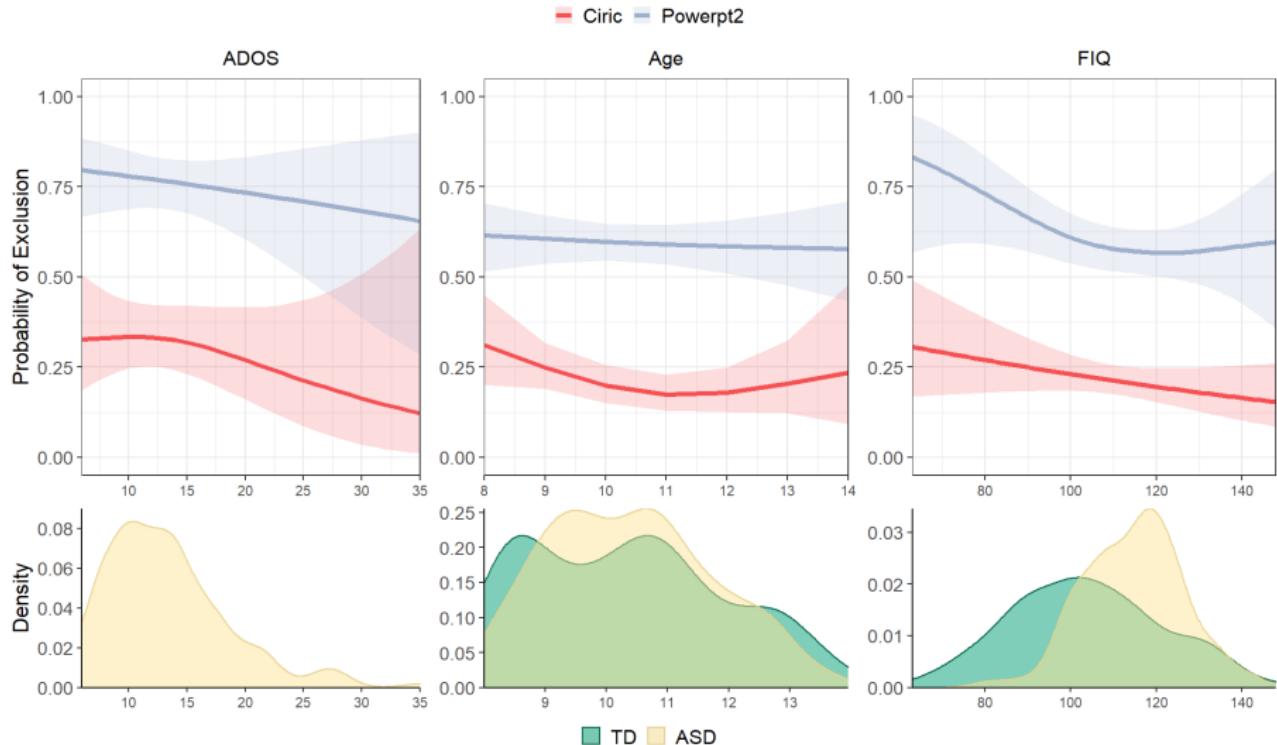


Figure: From Nebel et al. (2022).

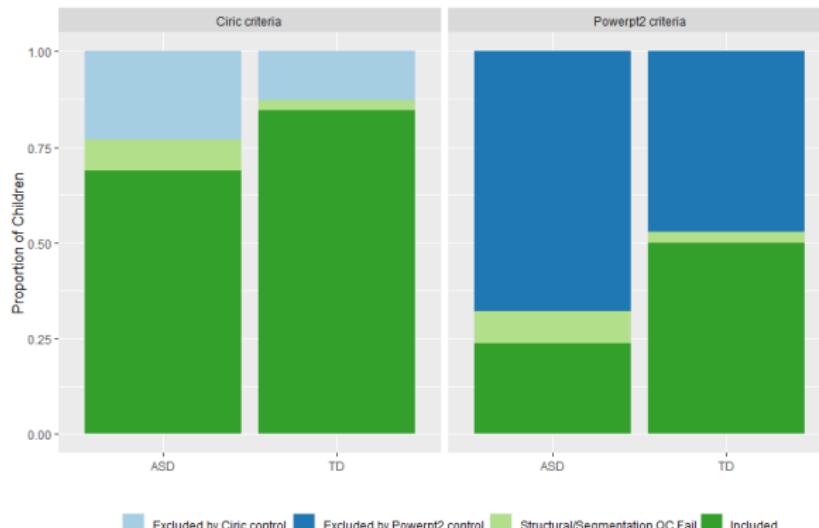
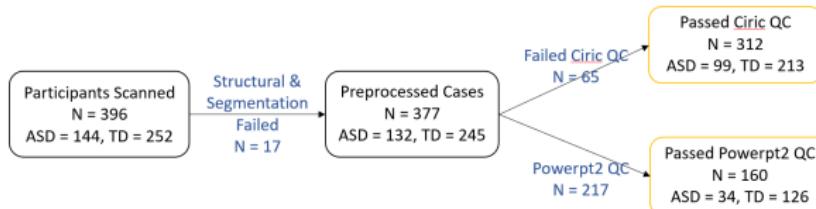
Current study: no relationship between motion exclusion criteria and ADOS, Age, or FIQ



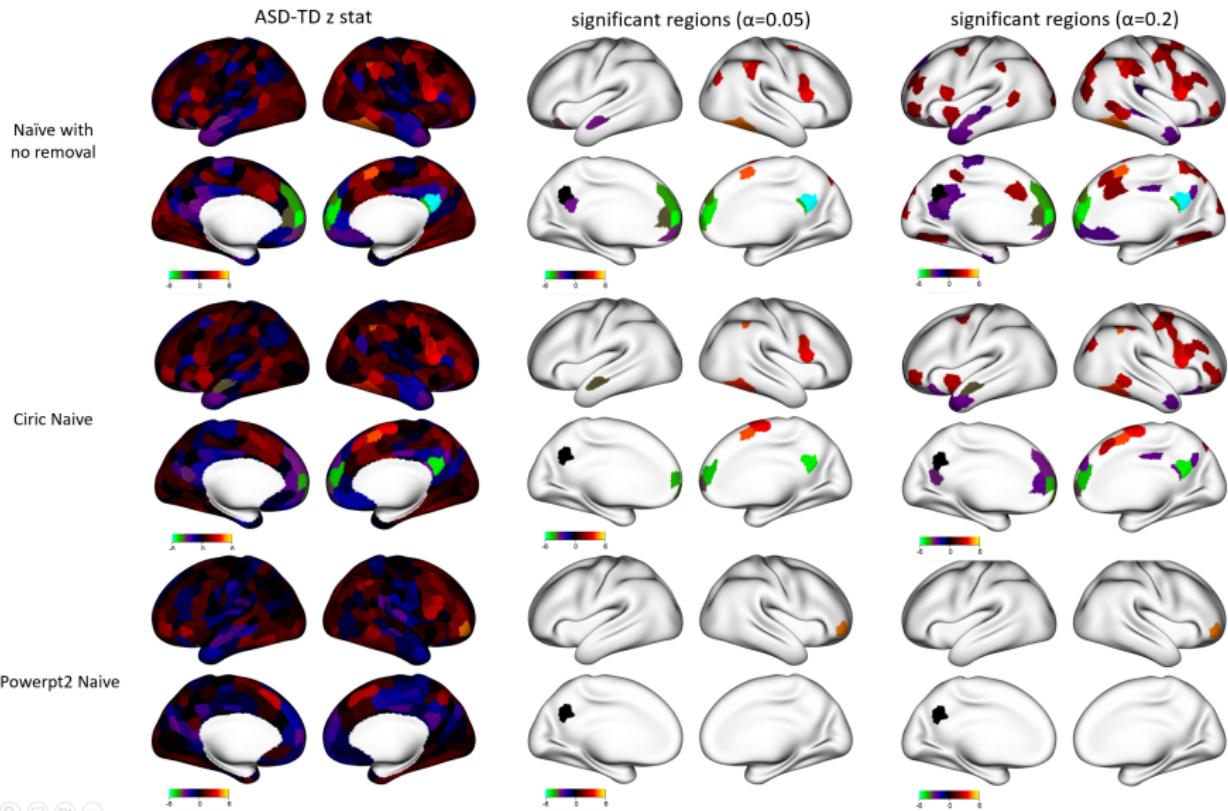
Resting-state fMRI analysis

- We performed preprocessing using fmriprep with `--cifti-output`.
- Visually inspected segmentation, 17 had issues.
- Used 400-node parcellation from ([Schaefer et al., 2018](#)) and `ciftiTools` following the tutorial in ([Pham et al., 2022](#)).
- Used a region in the default mode network as a seed region and focused on its correlation with 399 regions, since DMN is thought to be important in ASD ([Di Martino et al., 2014](#)).
- COMBAT for siteXacquisitionXheadcoil harmonization ([Fortin et al., 2017](#)).
- SuperLearner with 10-fold CV for propensity and outcome models:
`SL.earth`, `SL.glmnet`, `SL.gam`, `SL.glm`, `SL.ranger`,
`SL.ridge`, `SL.step`, `SL.step.interaction`, `SL.svm`,
`SL.xgboost`: diagnosis, ADOS, FIQ, stimulants, non-stimulants, age, sex, and handedness.

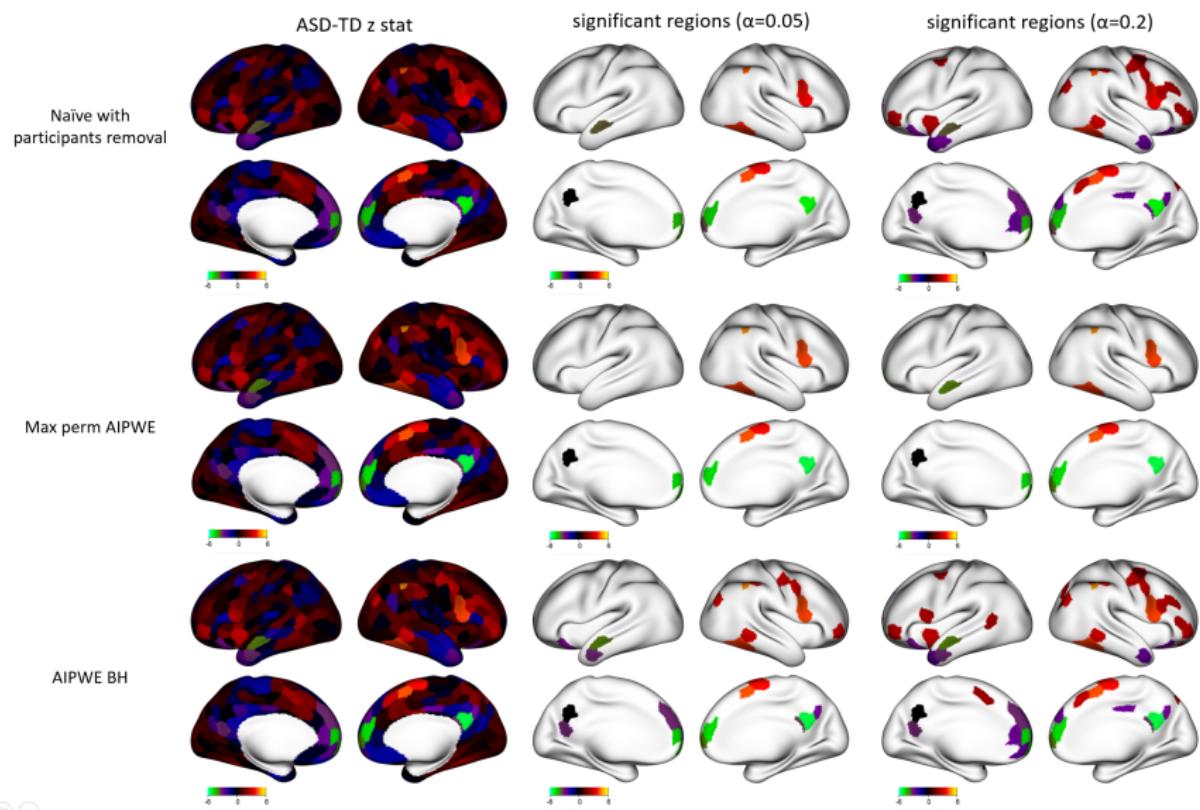
QC of school-age children in ABIDE I and II



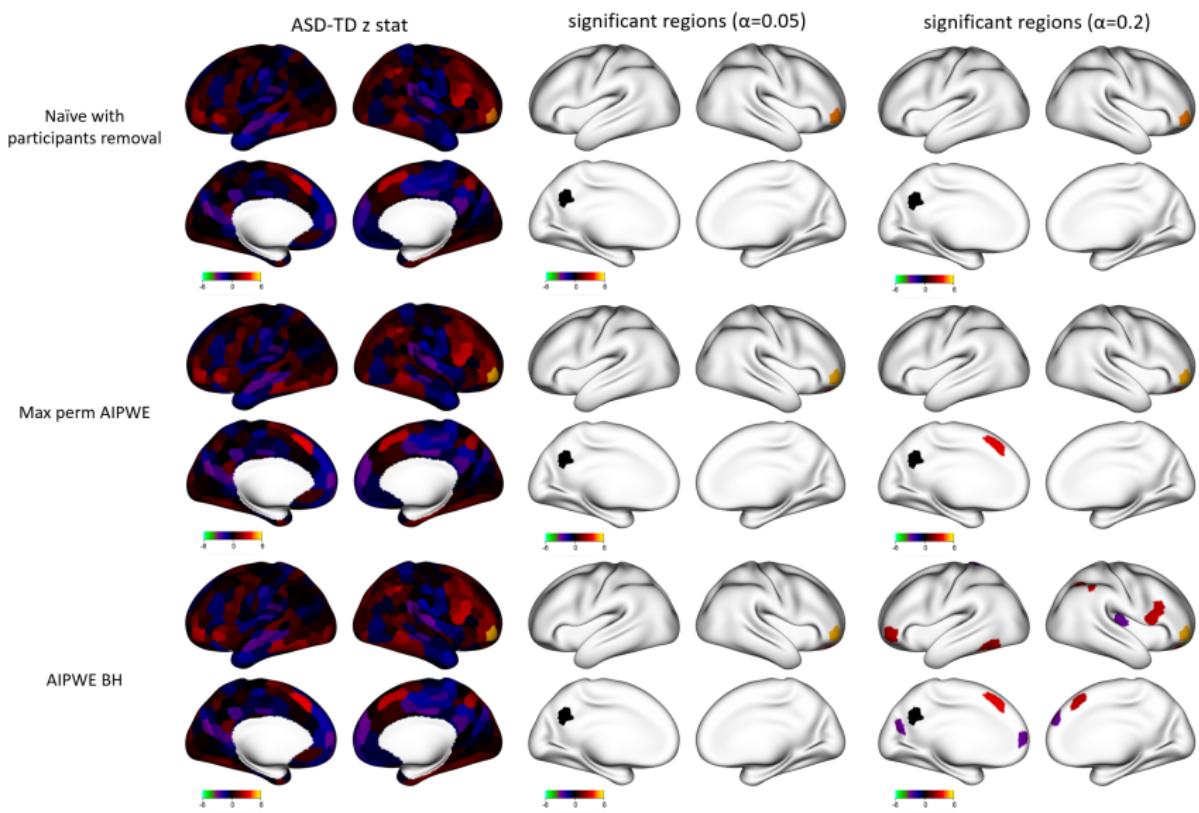
Comparison of naive motion removal strategies



AIPWE motion control: lenient (Ciric) criteria



AIPWE motion control: strict (powerpt2) criteria



Returning to our goals

- Overview of our approach for accounting for selection bias and improving statistical power in fMRI studies.
AIPWE can reduce sample bias and improve power.
- Compare AIPWE to DRTMLE in smaller sample sizes.
In simulations, AIPWE had better type-1 error control, but still inflated in some settings.
- Propose a computationally scalable permutation test.
Results are mixed.
- Results on school-age children from Autism Brain Imaging Data Exchange.
With 99 usable ASD scans, we found some evidence of differences between ASD and TD. With 34 usable scans ASD scans, we found limited evidence of differences between ASD and TD.

Limitations

- Additional research to disentangle false positives from true positives.
- Preliminary results using bootstraps are promising, but we are having issues with computational scalability.
- Our current model suffers from probable violations of the assumption of no missing confounders.
- Examine other applications where we expect selection bias – ABCD developmental differences between boys and girls, cortical thickness studies in Alzheimer's, ADHD, prospective Autism study in the brisklab with richer phenotyping.

Acknowledgments

This research was supported by the 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.

Thank you!

- Thank you!
- <https://github.com/thebrisklab>
- We are looking for a post doc or research scientist: email benjamin.risk@emory.edu.



References I

- Bang, H. and Robins, J. M. (2005). Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, 61(4):962–973.
- Benkeser, D., Carone, M., Laan, M. J. V. D., and Gilbert, P. B. (2017). Doubly robust nonparametric inference on the average treatment effect. *Biometrika*, 104(4):863–880.
- Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D. A., Gallagher, L., Kennedy, D. P., Keown, C. L., Keysers, C., Lainhart, J. E., Lord, C., Luna, B., Menon, V., Minshew, N. J., Monk, C. S., 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. *Mol. Psychiatry*, 19(6):659–667.

References II

- Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., and Shinozaki, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *Neuroimage*, 161:149–170.
- 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., Feczkó, 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., Barch, D. M., Luna, B., Fair, D. A., and Dosenbach, N. U. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature* 2022 603:7902, 603(7902):654–660.
- Nebel, M. B., Lidstone, D. E., Wang, L., Benkeser, D., Mostofsky, S. H., and Risk, B. B. (2022). Accounting for motion in resting-state fMRI: What part of the spectrum are we characterizing in autism spectrum disorder? *NeuroImage*, 257:119296.

References III

- Nielsen, A. N., Greene, D. J., Gratton, C., Dosenbach, N. U., Petersen, S. E., and Schlaggar, B. L. (2019). Evaluating the prediction of brain maturity from functional connectivity after motion artifact denoising. *Cerebral Cortex*, 29(6):2455–2469.
- Pham, D. D., Muschelli, J., and Mejia, A. F. (2022). ciftitools: A package for reading, writing, visualizing, and manipulating CIFTI files in R. *Neuroimage*, 250(118877):118877.
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3):2142–2154.
- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., and Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex*, 28(9):3095–3114.
- Van Der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1).