Understanding (and mitigating) the impact of preprocessing pipelines on neuroimaging analyses (Can *you** reproduce neuroimaging analyses?)

Nikhil Bhagwat[1], Amadou Barry[2], Erin W. Dickie[3], Shawn T. Brown[1], Gabriel A. Devenyi[4,5], M. Mallar Chakravarty[4,5], Celia M. T. Greenwood[2,8], Bratislav Misic[1], David N. Kennedy[6], Jean-Baptiste Poline[1,6,8]

1. Montreal Neurological Institute & Hospital, McGill University, Montreal, QC, Canada
2. Lady Davis Institute for Medical Research, McGill University, Montreal, QC, Canada
3. Kimel Family Translational Imaging-Genetics Research Lab, CAMH, Toronto, ON, Canada
4. Computational Brain Anatomy Laboratory, Douglas Mental Health Institute, Verdun, QC, Canada
5. Department of Psychiatry, McGill University, Montreal, QC, Canada
6. Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada
7. Child and Adolescent Neurodevelopment Initiative, University of Massachusetts, Worcester, USA
8. Ludmer Centre for Neuroinformatics & Mental Health, McGill University, Montreal, QC, Canada

Contact: nikhil153@gmail.com, jean-baptiste.poline@mcgill.ca

Code: https://github.com/neurodatascience/compare-surf-tools

*** Take our short survey! Help us to help you!**
https://forms.gle/SQiZibd1rUdWqWYX8

INTRODUCTION

- Implementation choices in the computational workflow introduce variability in the neuroimaging analysis - which consequently impacts reproducibility of scientific findings.
- We highlight this issue with a sample sMRI dataset processed using several pipelines (see Fig. 1).
- We compare the processed features and the performance variation of statistical models for biological inference and individualize prediction.

METHODS

- Dataset: Autism Brain Imaging Data Exchange (ABIDE) Preprocessing project
- Phenotype: Cortical thickness
- Comparisons
  - Software: Freesurfer (6.0, 5.3, 5.1) , CIVET (2.1.0), ANTs (RRID:SCR_004757)
  - Parcellations: Desikan-Killiany Atlas (31x2), Destrieux (74x2), Glasser (180x2)
  - Quality control (QC): manual, automatic outlier detection
- Analyses
  - Neurobiological inference: diagnostic group differences
  - Individualized prediction: Dx, Age prediction

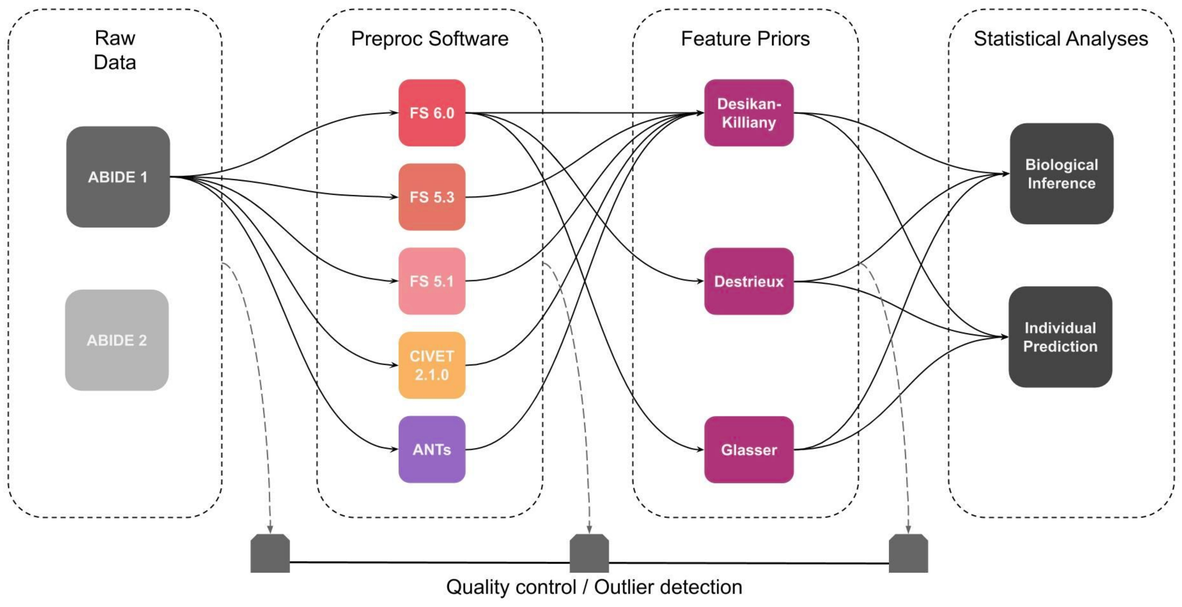
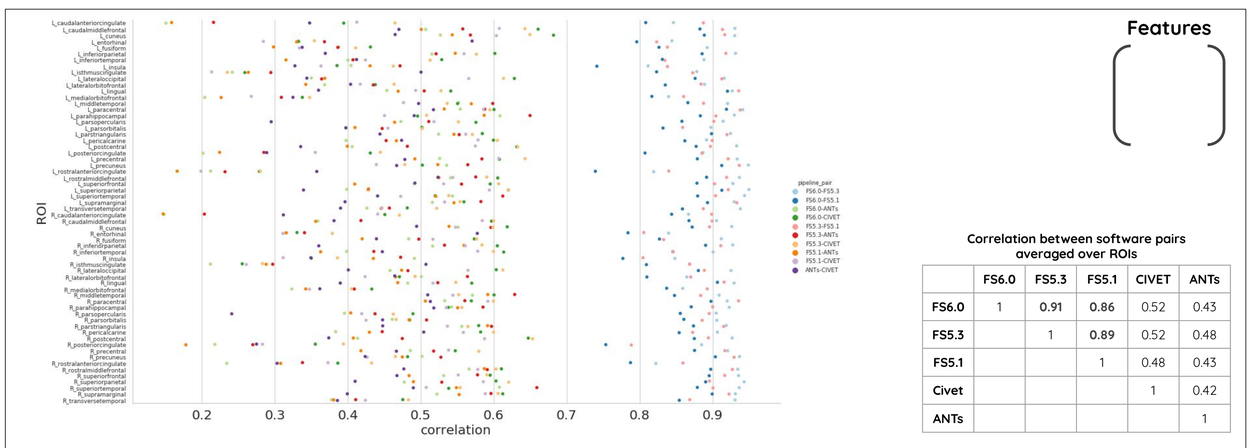


Figure 1:
Computational workflow building blocks and potential permutations for typical structural MR image analyses. Only a subset of possible pipelines analyzed in this work is shown with connected arrows.

RESULTS

Neurobiology ⟷ *Individuals*



Correlation between software pairs averaged over ROIs

	FS6.0	FS5.3	FS5.1	CIVET	ANTs
FS6.0	1	**0.91**	**0.86**	0.52	0.43
FS5.3		1	**0.89**	0.52	0.48
FS5.1			1	0.48	0.43
Civet				1	0.42
ANTs					1

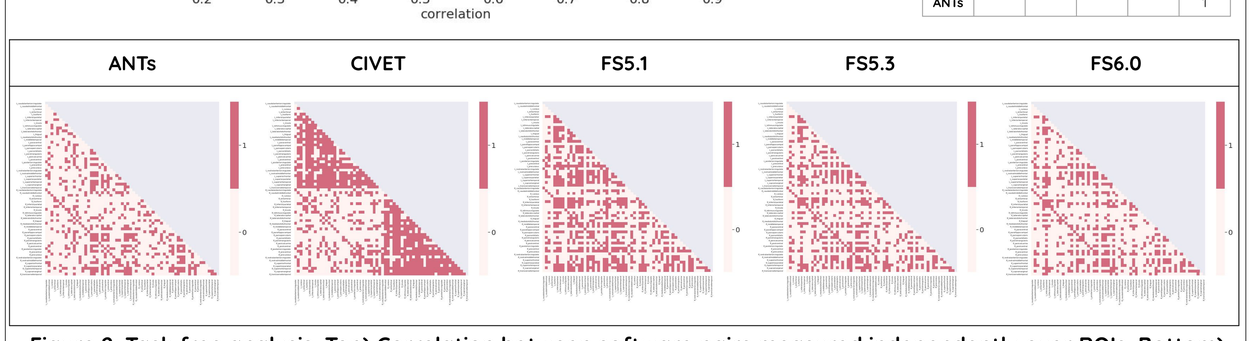
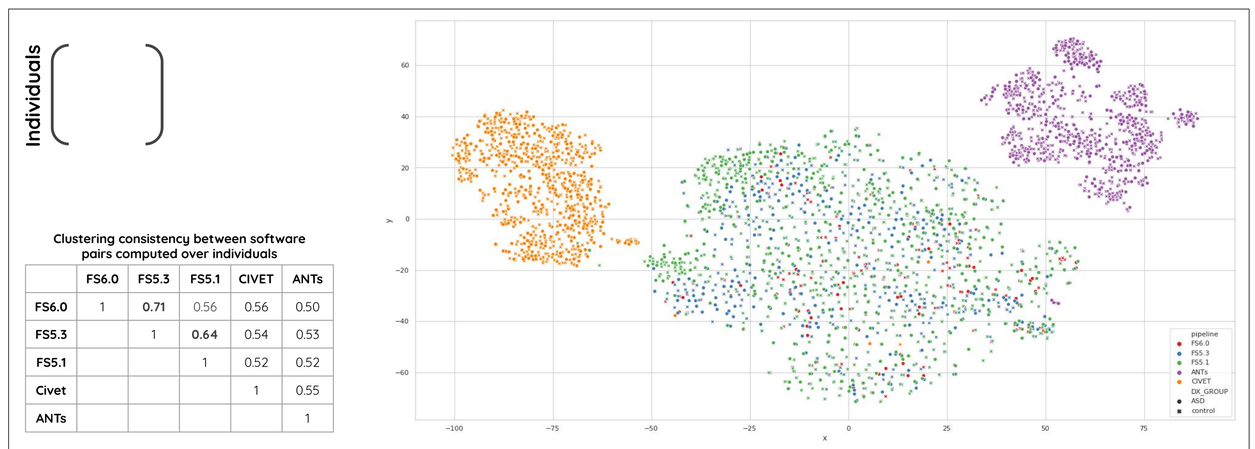


Figure 2: Task free analysis. Top) Correlation between software pairs measured independently over ROIs. Bottom) Structural covariance per software measured as inter-ROI correlation. The ROIs are defined by DK parcellation.



Clustering consistency between software pairs computed over individuals

	FS6.0	FS5.3	FS5.1	CIVET	ANTs
FS6.0	1	**0.71**	0.56	0.56	0.50
FS5.3		1	**0.64**	0.54	0.53
FS5.1			1	0.52	0.52
Civet				1	0.55
ANTs					1

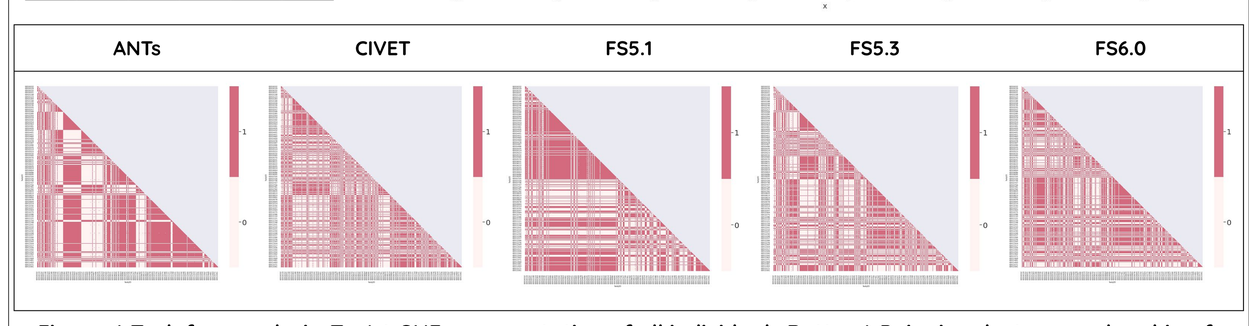
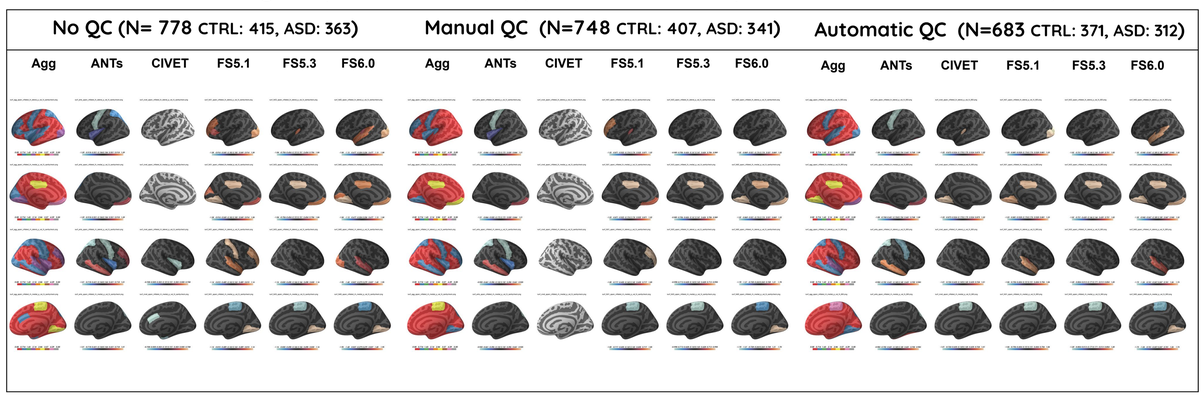


Figure 4: Task free analysis. Top) t-SNE representation of all individuals Bottom) Pairwise cluster-membership of all individuals. Binary membership is assigned based on two clusters derived from hierarchical clustering.

Dx group differences (Software Comparisons (with DK parcels))



Dx group differences (Parcellation Comparisons (with FS6.0))

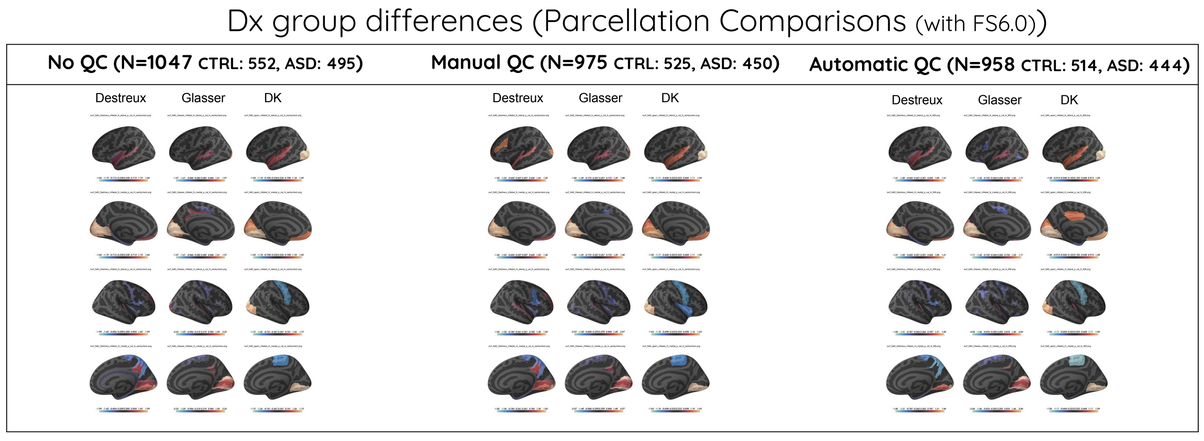
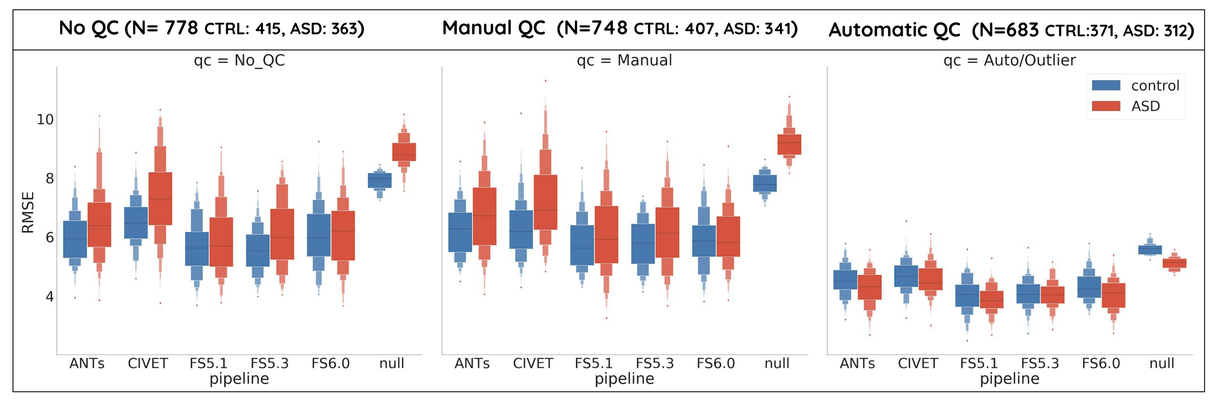


Figure 3: Task driven analysis. Top) Significantly ROI differences with various software and QC levels. The 'agg' column represents ranking based on overlap across pipelines. Bottom) Significantly ROI differences with various parcellations and QC levels. The ROI colors denote beta values.

Age prediction (Software Comparisons (with DK parcels))



Age prediction (Parcellation Comparisons (with FS6.0))

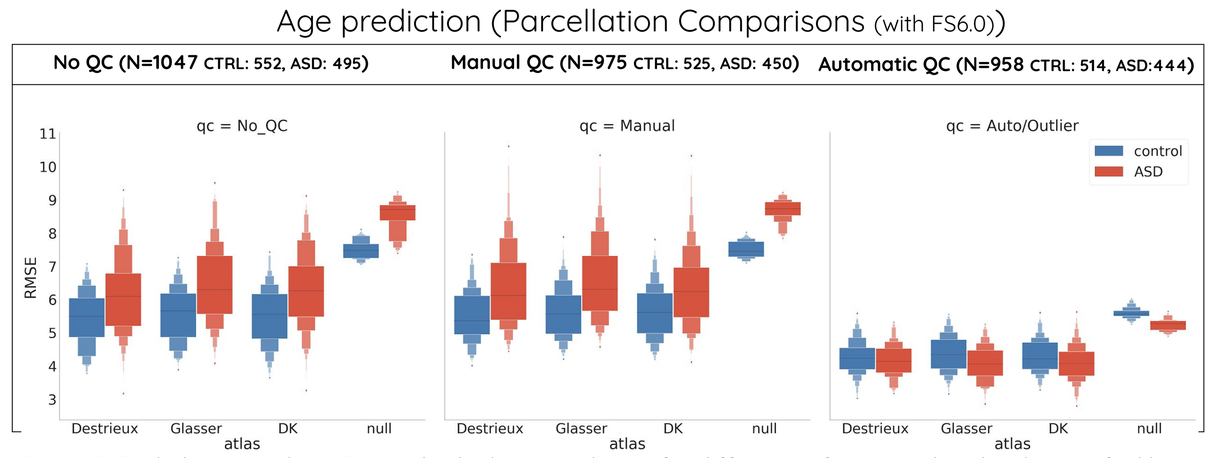


Figure 5: Task driven analysis. Top) Individual age prediction for different software and QC levels stratified by diagnosis. Bottom)) Individual age prediction for different parcellations and QC levels stratified by diagnosis. Performance is validated using a Random Forest model over 100 shuffle-split iterations stratified by Dx.

CONCLUSIONS

- Large differences across software produce weakly correlated features and offer poor clustering consistency across individuals.
- Consensus across parcellations can improve the confidence and specificity of biomarkers; however no significant impact is seen on individual predictions.
- QC procedure has strong impact on biological and individual findings, as well as the underlying null distributions.
- We encourage reporting of variation in scientific findings with respect to implementation choices of software, priors (e.g. parcellations), and QC filtering.

ACKNOWLEDGMENTS

This project was partially funded by NIH-NIBIB P41 EB019936 (ReproNim) NIH-NIMH R01 MH083320 (CANDIShare) and NIH 5U24 DA039832 (NIF), as well as a Discovery Grant from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative. We thank Gleb Bezgin, John Lewis, and David Kennedy's group for compiling manual QC lists used in this work. We thank Lindsey Lewis and Claude Lepage for advising us on surface analysis and comparability across Freesurfer and Civet. We also thank Satrajit Ghosh for helping us extend this work as a stand-alone module for vibration analysis in future neuroimaging workflows.

REFERENCES

1. Craddock, R. Cameron, and Pierre Bellec. 2015. "Preprocessed Connectomes Project: Abide." *Available Online at: Http://preprocessed-Connectomes-Project. Github. Io/abide.* http://preprocessed-connectomes-project.org/abide/.
2. Dickie, Erin, Steven M. Hodge, R. Cameron Craddock, Jean-Baptiste Poline, and David N. Kennedy. 2017. "Tools Matter: Comparison of Two Surface Analysis Tools Applied to the ABIDE Dataset." *Riogrande Odontologico* 3 (May): e13726.
3. Tustison, Nicholas J., Philip A. Cook, Arno Klein, Gang Song, Sandhitsu R. Das, Jeffrey T. Duda, Benjamin M. Kandel, et al. 2014. "Large-Scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements." *NeuroImage.*