

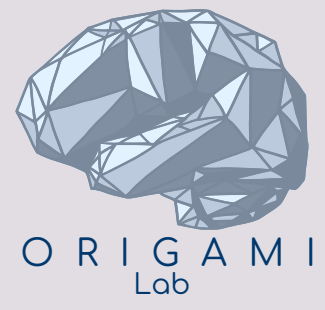
Benchmarking federated learning approaches against siloed and mega-analysis regimes

Michelle Wang¹, Nikhil Bhagwat¹, Gabriel D. Pinilla-Monsalve², Yuan Song², Oury Monchi², Jean-Baptiste Poline¹

¹ McGill University, Montreal, Canada

² Centre de recherche de l'Institut universitaire de gériatrie de Montréal, Montreal, Canada

Email: michelle.wang6@mail.mcgill.ca, jean-baptiste.poline@mcgill.ca



Introduction

Many clinical neuroscience datasets **cannot be openly shared** due to institutions and countries adopting strong data privacy frameworks (Marelli & Testa, 2018): these datasets remain in so-called "**data silos**".

Decentralized data processing tools and **federated** analysis methods could enable large-scale, multi-site studies that make use of these datasets while **respecting data sharing constraints**.

We compare a simple **federated** machine learning setup (i.e. sharing only fitted models) with two traditional experimental setups – **siloed** (no sharing) and **mega-analysis** (sharing data).

Hypothesis: model generalizability improves as we go from siloed to federated to mega-analysis setups.

Neuroinformatics tool for multicentric analyses

Nipoppy framework for the organization and **decentralized processing** of datasets.



Neurobagel ecosystem for distributed data **harmonization and search**.



Methods

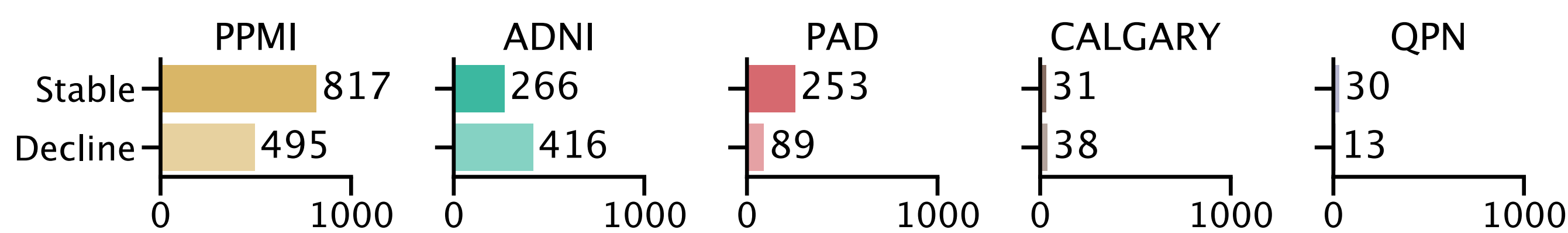
Datasets

We used **demographic** information, structural **neuroimaging** features (cortical thickness [CT] and subcortical volumes [SV]) and **cognitive** assessment scores (Montreal Cognitive Assessment [MoCA] or Mini-Mental State Examination [MMSE]) from five Parkinson's disease (PD) or Alzheimer's disease (AD) datasets:

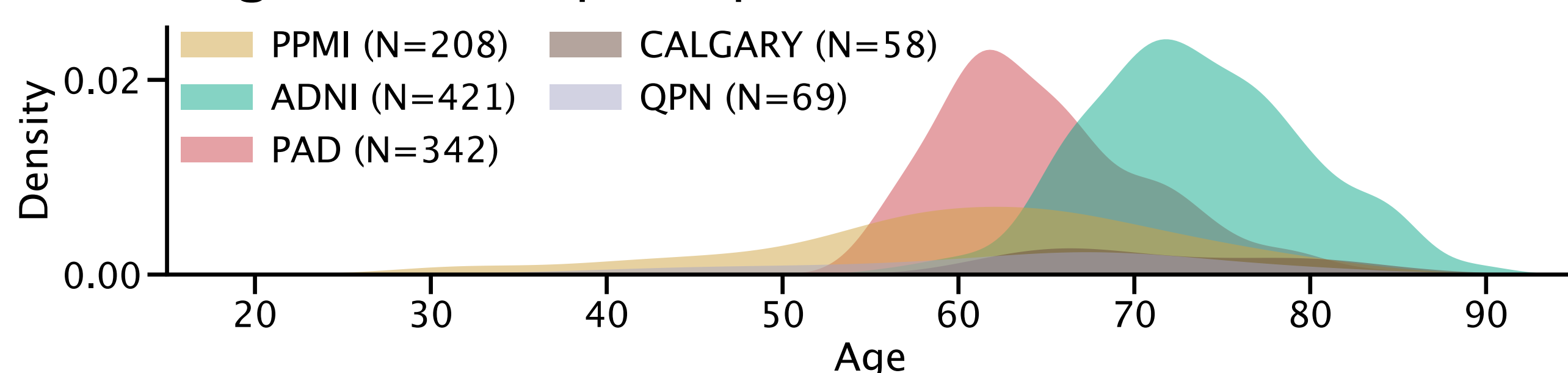
- Parkinson's Progression Markers Initiative (**PPMI**) (Marek et al., 2018)
- Alzheimer's Disease Neuroimaging Initiative (**ADNI**) (Jack et al., 2008)
- Pre-symptomatic Evaluation of Experimental or Novel Treatments for Alzheimer's Disease (**PAD**)
- University of Calgary PD and mild cognitive impairment (**Calgary**)
- Quebec Parkinson Network (**QPN**) (Gan-Or et al., 2020)

Prediction tasks

1) Predict **cognitive decline** (≥ 1 -point/year loss in MoCA or MMSE scores within 5 years from baseline) in patients from **age + sex + CT**



2) Predict **age** of control participants from **sex + SV**



Model training setups

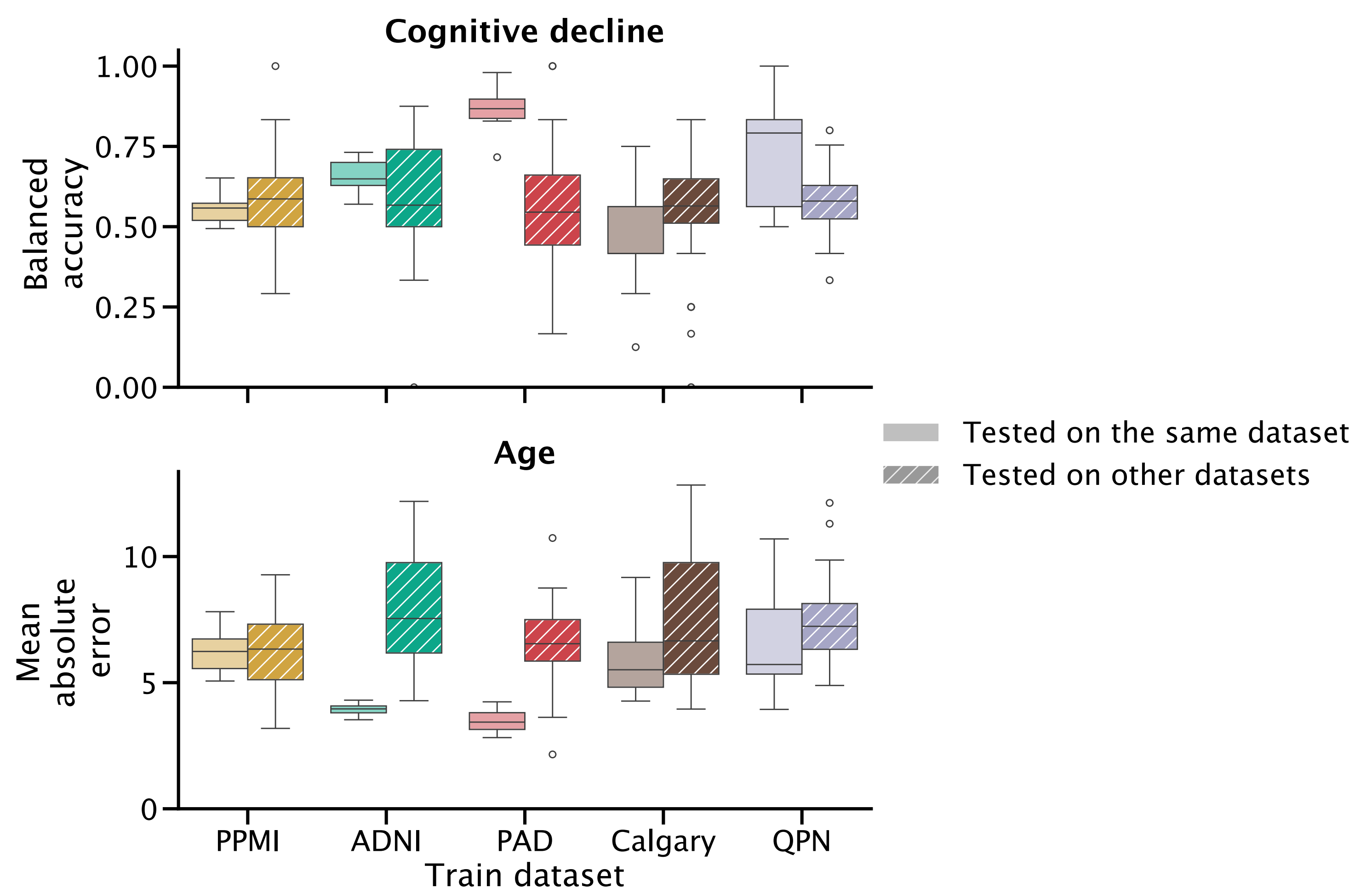
Model: Z-scoring + stochastic gradient descent classifier/regressor

Setup	Train data	Model
Siloed (PPMI)	PPMI	$M_{\text{siloed, ppmi}}$
Siloed (ADNI)	ADNI	$M_{\text{siloed, adni}}$
Siloed (PAD)	PAD	$M_{\text{siloed, pad}}$
Siloed (Calgary)	Calgary	$M_{\text{siloed, calgary}}$
Siloed (QPN)	QPN	$M_{\text{siloed, qpn}}$
Federated	PPMI	$M_{\text{federated}}$ (weighted avg. of params.)
	ADNI	
	PAD	
	Calgary	
	QPN	
Mega-analysis	PPMI, ADNI, PAD, Calgary, QPN	$M_{\text{mega-analysis}}$

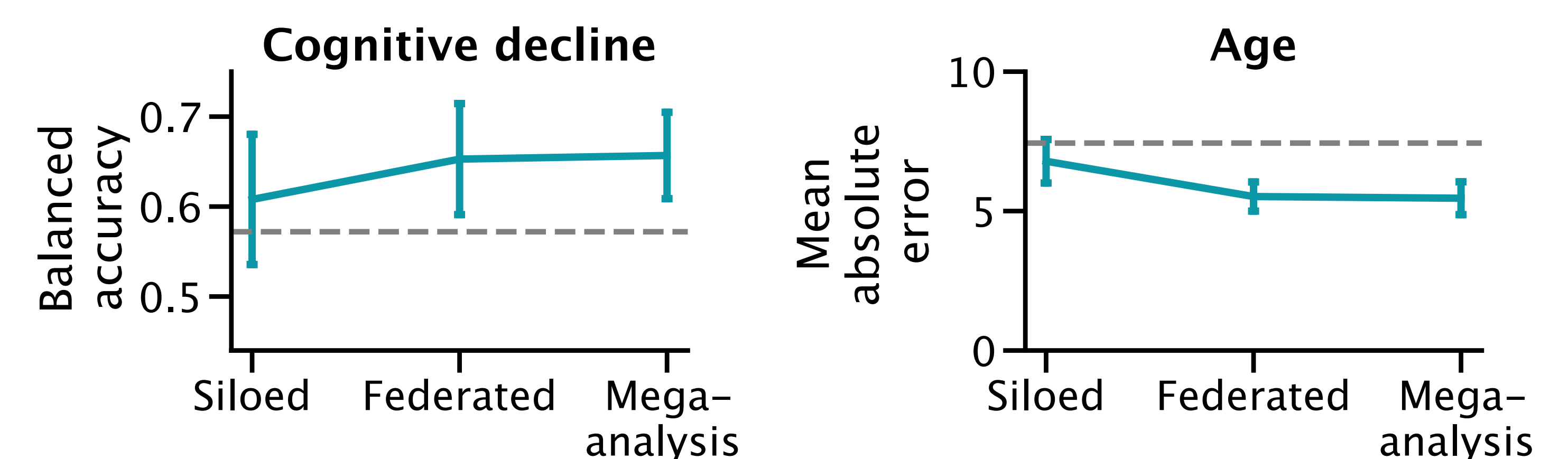
Model performance estimated using **10-fold cross-validation**

Results

Poor cross-dataset generalizability in Siloed setups (dataset shift)



Test performance scores on **PPMI+ADNI+PAD+Calgary+QPN** show **improvement from Siloed to Federated setup** in both tasks



Dashed lines: 95th percentile Mega-analysis null models

Conclusion

Our results suggest that, for the datasets and use-cases investigated, **federated setups show generalizable performances that are comparable to performances with mega-analysis setups**. Hence, federated learning is a promising direction for multisite studies with data-sharing and governance constraints.

References

- Gan-Or, Z. et al. (2020). The Quebec Parkinson Network: A Researcher-Patient Matching Platform and Multimodal Biorepository. *Journal of Parkinson's Disease*, 10(1), 301–313.
- Jack, C. R. et al. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: JMIR*, 27(4), 685–691.
- Marek, K. et al. (2018). The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, 5(12), 1460–1477.
- Marelli, L., & Testa, G. (2018). Scrutinizing the EU General Data Protection Regulation. *Science*, 360(6388), 496–498.

Acknowledgements

