Artificial
Intelligence in
Medicine

Harmonization strategies for MRI features: summary of PI, BA and BO analyses and future goals

AIM1.T1

A. Retico (PI), BA, BO

AIM General Meeting - Oct 15-16, 2020

Outline

Contributions of the Pisa, Bari and Bologna groups to AIM1.T1 - Multi-site data harmonization in MRI.

- Pisa: highlighted the confounding effect of the "data acquisition site" in machine learning classification of MRI features and developed a Confounding Index (CI) to quantify the impact of confounding factors in machine learning binary classification problems
- Bologna: implemented a statistical harmonization technique based on ComBat and Surrogate Variable Analysis, and is now facing with an age prediction problem
- Bari: implemented the ComBat harmonization and extensively validated it in an age prediction problem with three ML models

- Additional materials available (previous meetings):
  - https://agenda.infn.it/event/21746/ (Collaboration meeting February 3rd, 2020)
  - https://agenda.infn.it/event/17879/ (Kickoff meeting January 30th, 2019)

AIM1.T1: multi-site data harmonization

data gathered by different sites and/or acquisition systems carries local "fingerprint", often to the detriment of the much more subtle information of interest.

this problem is akin to the management of **systematic errors**

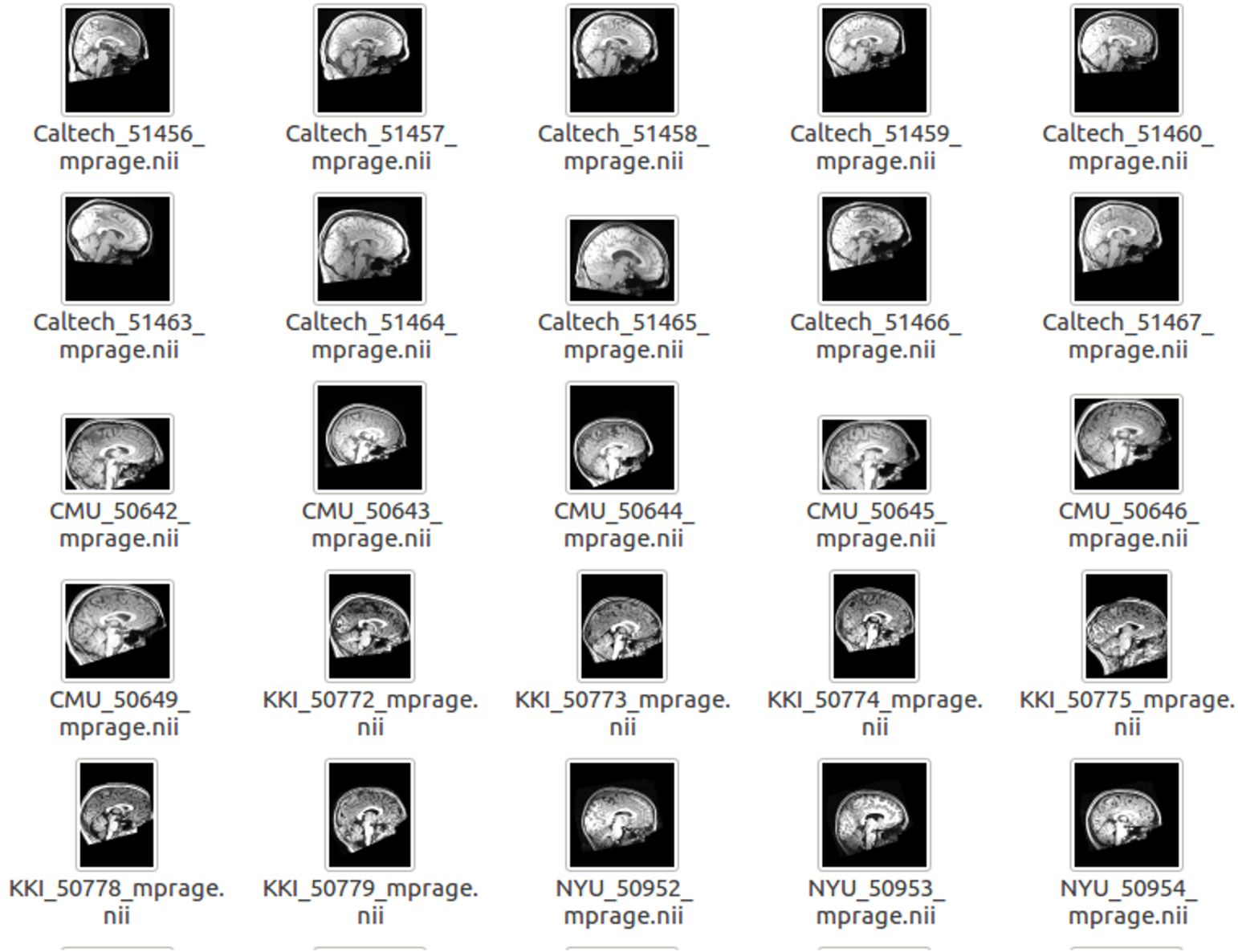
typical application cases: MRI, RX, PET, NPSY tests

2226 subjects			
1060 ASDs		1166 TDCs	
907 M	153 F	879 M	287 F
Age at Scan 5 – 64 years			
40 different acquisition sites			

ABIDE

Autism Brain Imaging Data Exchange

FreeSurfer



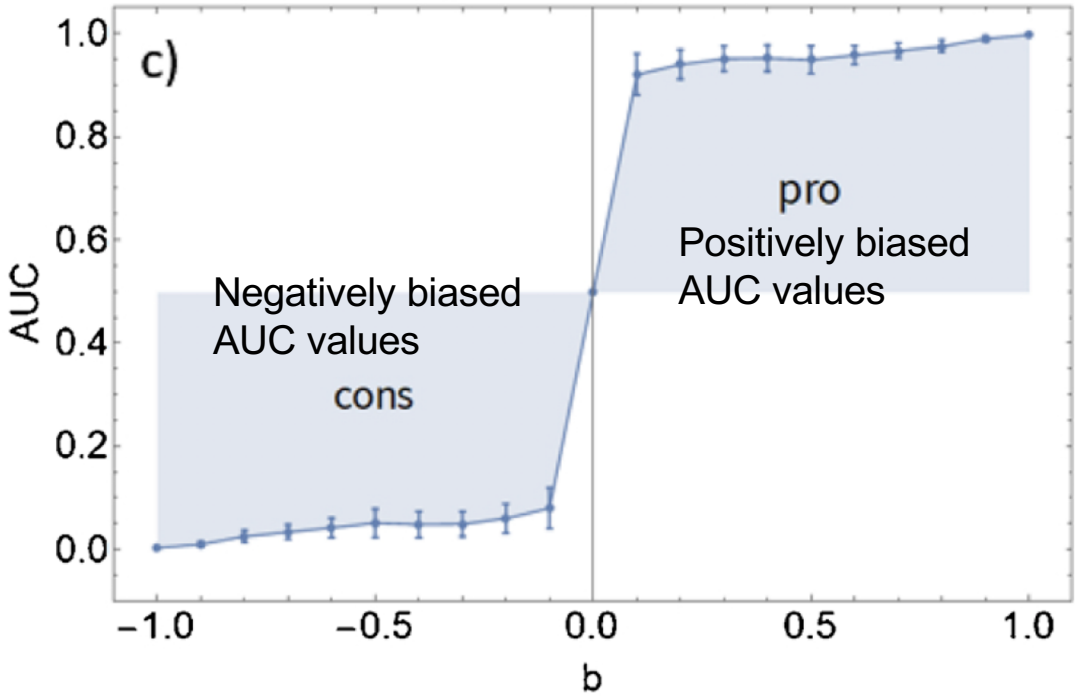
Confounding Index (CI)

- Machine learning (ML) models trained on brain MRI features (e.g. features extracted with FreeSurfer) are strongly affected by the confounding effect introduced by the acquisition site.
- In addition, other confounding variable have to be accounted for in a typical case-control classification (e.g. age, gender)

- A Confounding Index (CI) has been developed to quantify the impact of confounding factors with respect to a two-class ML classification task
  - Ferrari, E., Retico, A., & Bacciu, D. (2020). Measuring the effects of confounders in medical supervised classification problems: the Confounding Index (CI). Artificial Intelligence in Medicine, 103(Ci), 101804. https://doi.org/10.1016/j.artmed.2020.101804
  - Ferrari, E., Bosco, P., Calderoni, S., Oliva, P., Palumbo, L., Spera, G., Fantacci, M. E., & Retico, A. (2020). Dealing with confounders and outliers in classification medical studies: The Autism Spectrum Disorders case study. Artificial Intelligence in Medicine, 108(July), 101926. https://doi.org/10.1016/j.artmed.2020.101926

Confounding Index (CI)

- The proposed CI founds on measuring the variation of the AUC obtained using different, engineered biases during training, and thus depends on how the confounder and the class labels affect the input features.
- The CI ranges from 0 to 1 and allows:
  - to test the effect of a confounding variable on a specific binary classifier;
  - to rank variables with respect to their confounding effect;
  - to evaluate the effectiveness of a normalization procedure and assess the robustness of a training algorithm against confounding effects.



CI = Area_pro – Area_cons

Data normalization strategy implemented

- To reduce inter-individual variation due to head size and inter-image differences caused by voxel-scaling variations, the data of each subject have been normalized to global quantities of the same subject:
  - volumetric features are divided by the Estimated Total Intracranial Volume (eTIV), i.e. the Freesufer measure of the intracranial volume;
  - cortical surfaces are divided by the area of the total white matter
  - cortical thicknesses are divided by the mean cortical thickness across the entire brain.
- After "self-normalization", z-score is applied to bring all features in the same range.

Variable			Non-normalized Data		Normalized Data	
			CI	Error	*CI*	Error
Handedness			0.01	0.06	0.02	0.06
Sex			0.43	0.03	0.23	0.04
AM Acquisition Modality	$\neq$ scanners and protocols	NYU-I vs KKI_{8ch}-II	0.54	0.02	0.62	0.03
		NYU-I vs UM_1-I	0.63	0.02	0.70	0.02
	= scanners and protocols	NYU-I vs NYU_1-II	0.32	0.04	0.39	0.03
		UM_1-I vs UM_2-I	0.29	0.02	0.35	0.02

The normalization procedure does not mitigate the site effect

BO contribution to MRI data Harmonization

Claudia Sala e Daniel Remondini

Data: 961 features from NMR images (extracted with FreeSurfer) of 2364 subjects, measured in **16 different sites**.

Three main steps:

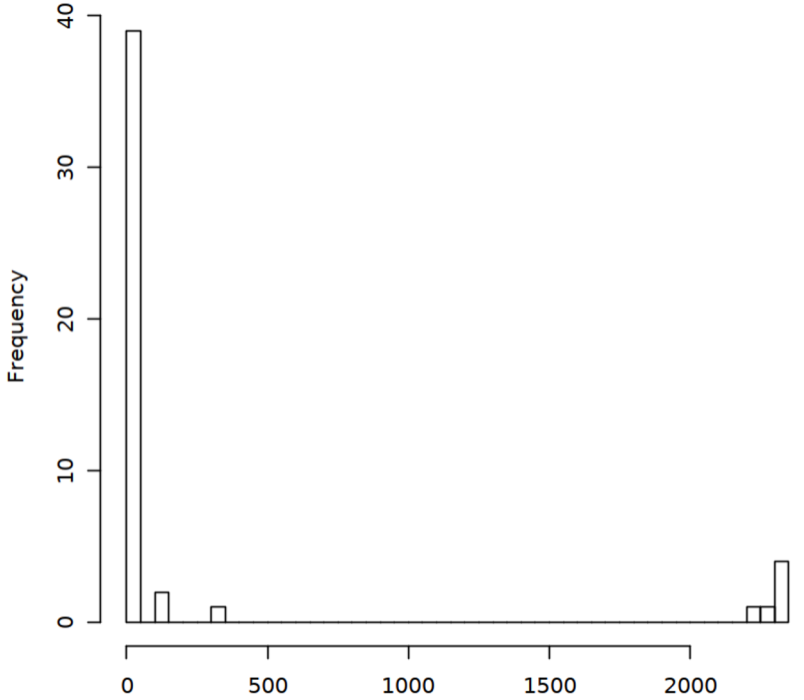
1) Data cleaning

2) Outlier removal

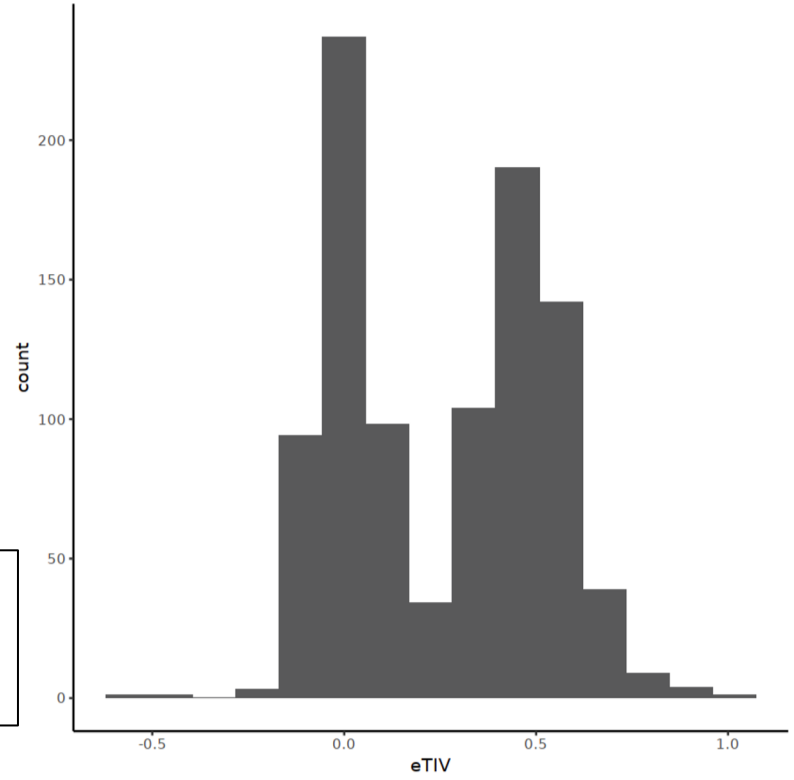
3) Batch identification and adjustment

Data cleaning

- Features with zeros in most samples were removed.
- Features were divided by eTIV (estimated Total Intracranial Volume) only when such transformation improved their correlation with age.



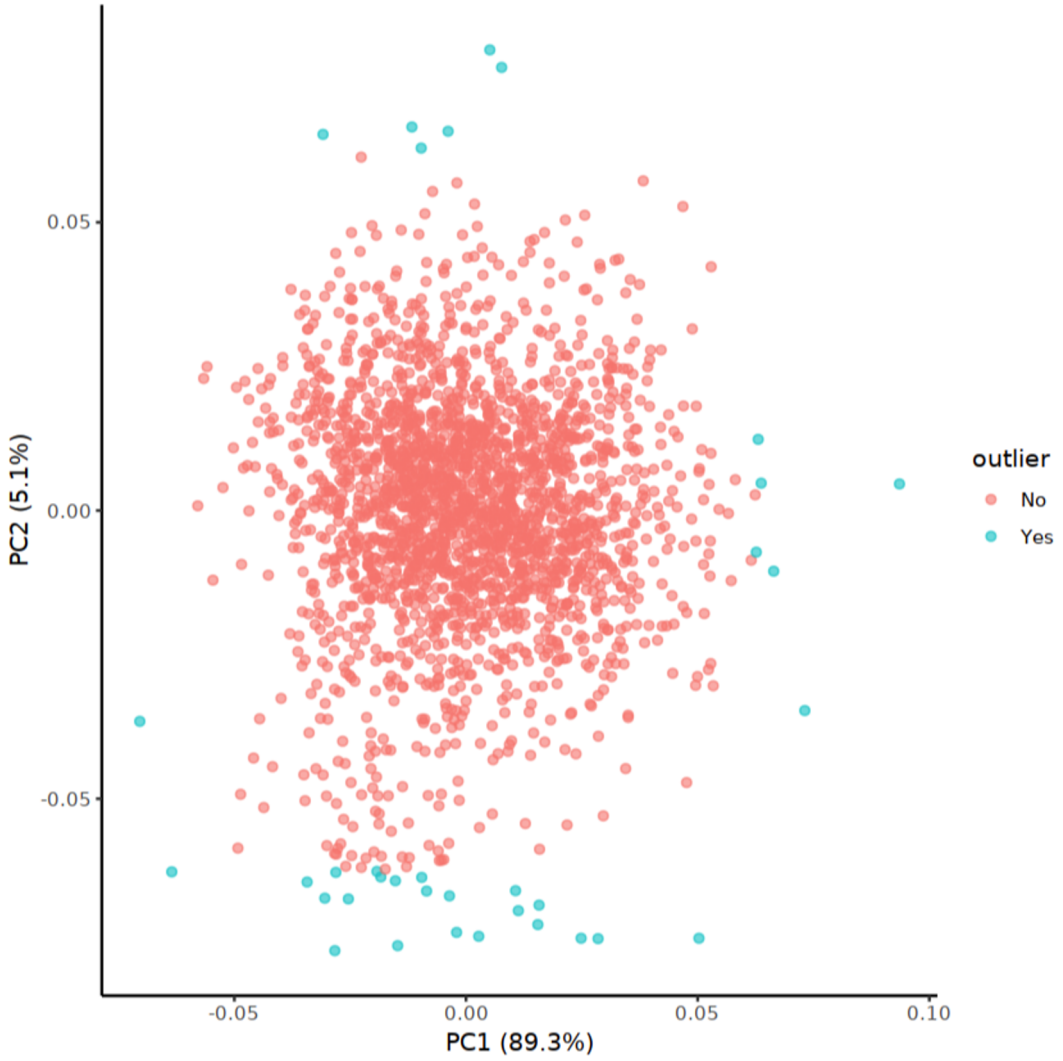
Histogram of the values of features with at least one zero. Some features are zero in most samples.



Histogram of correlations between each features and eTIV. Not all features are correlated with eTIV.

Removing outliers

- Principal Component Analysis (PCA) was computed taking all features into account.
- Considering the first two Principal Components (PC1, PC2), 35 samples that were more than 3σ from the mean of PC1 or PC2 were removed.
- This step improved the performance of the following batch effects identification and adjustment.



Batch identification and adjustment

1) adjust for know batches using COMBAT [1];
2) then identify and adjust unknown batch effects with SVA [2].

Alternative methods for the identification of unknown confounding factors or batches are RUV [3] and CATE [4].

[1] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007 Jan 1;8(1):118-27.

[2] Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007 Sep 28;3(9):e161.

[3] Gagnon-Bartsch JA, Jacob L, Speed TP. Removing unwanted variation from high dimensional data with negative controls. Berkeley: Tech Reports from Dep Stat Univ California. 2013 Dec 2:1-12.

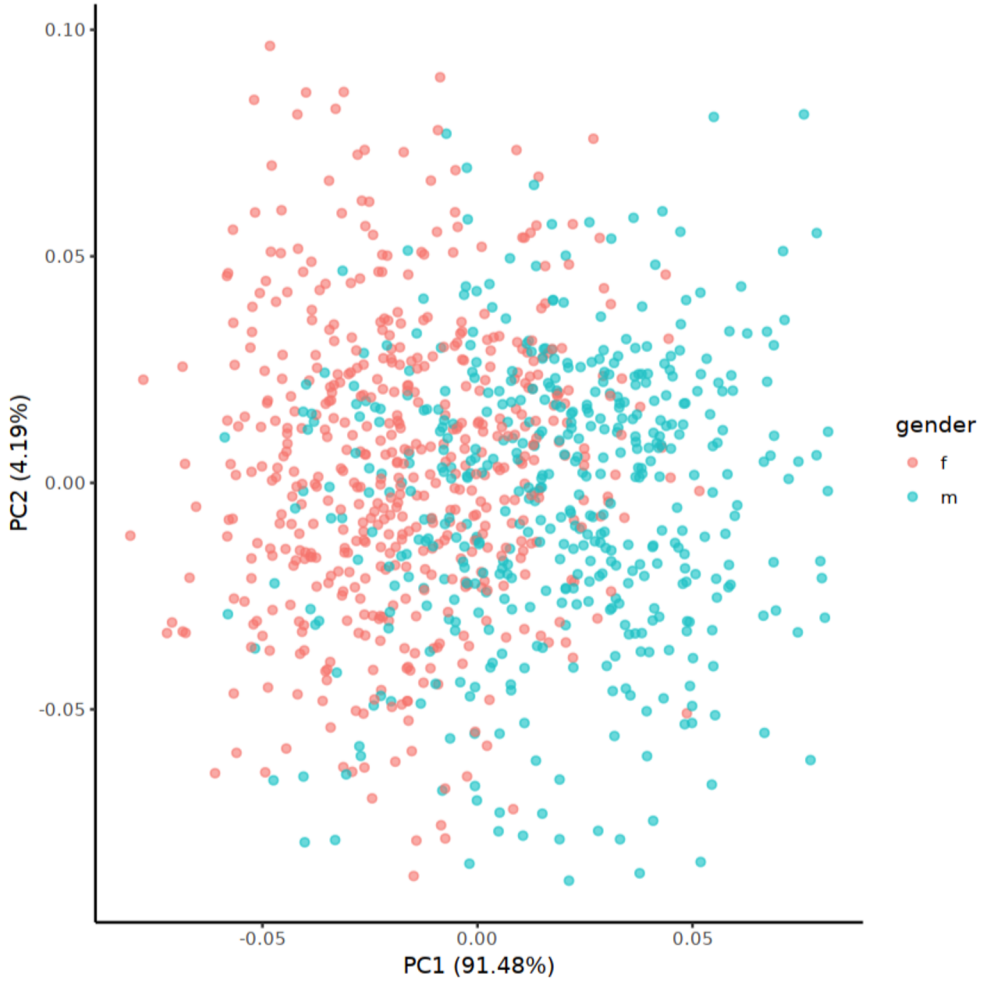
[4] Wang J, Zhao Q, Hastie T, Owen AB. Confounder adjustment in multiple hypothesis testing. Annals of statistics. 2017 Oct;45(5):1863.

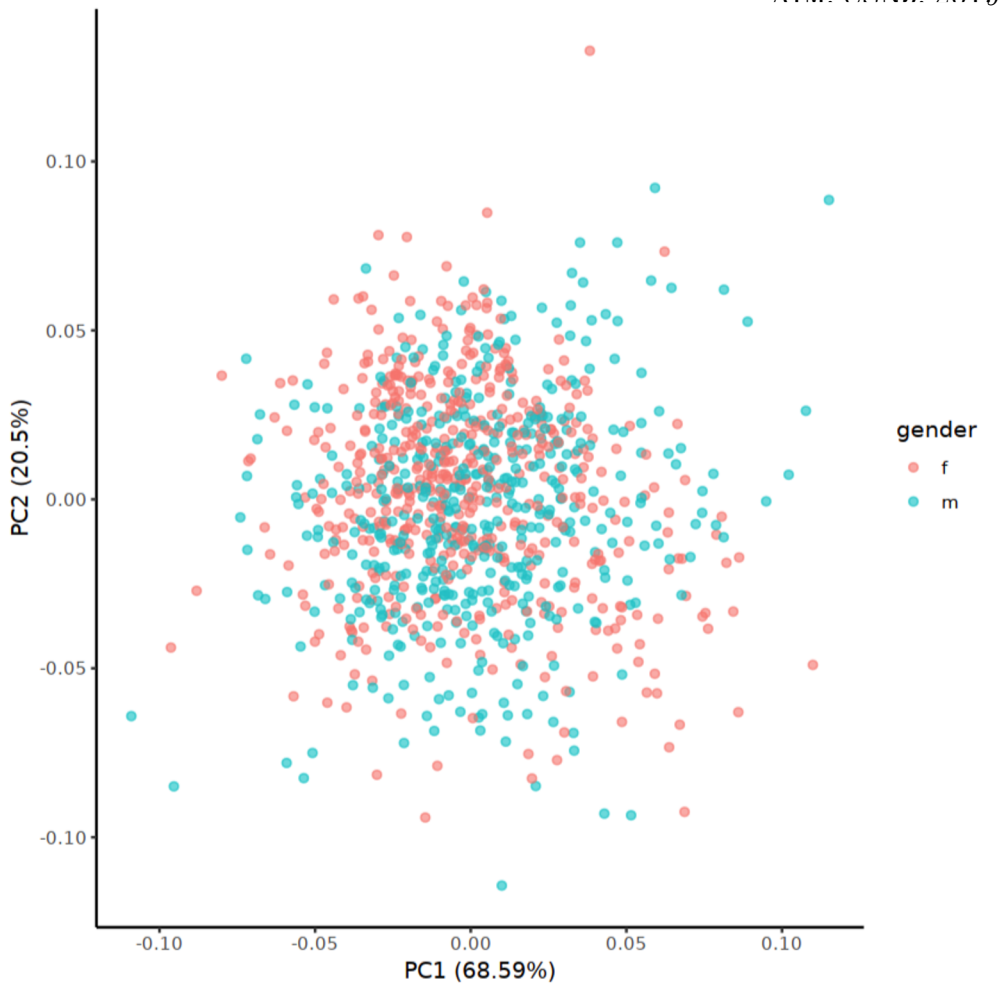
Batch identification and adjustment

To test the ability of SVA to identify and adjust unknown batches, we ignored the sex and site information.

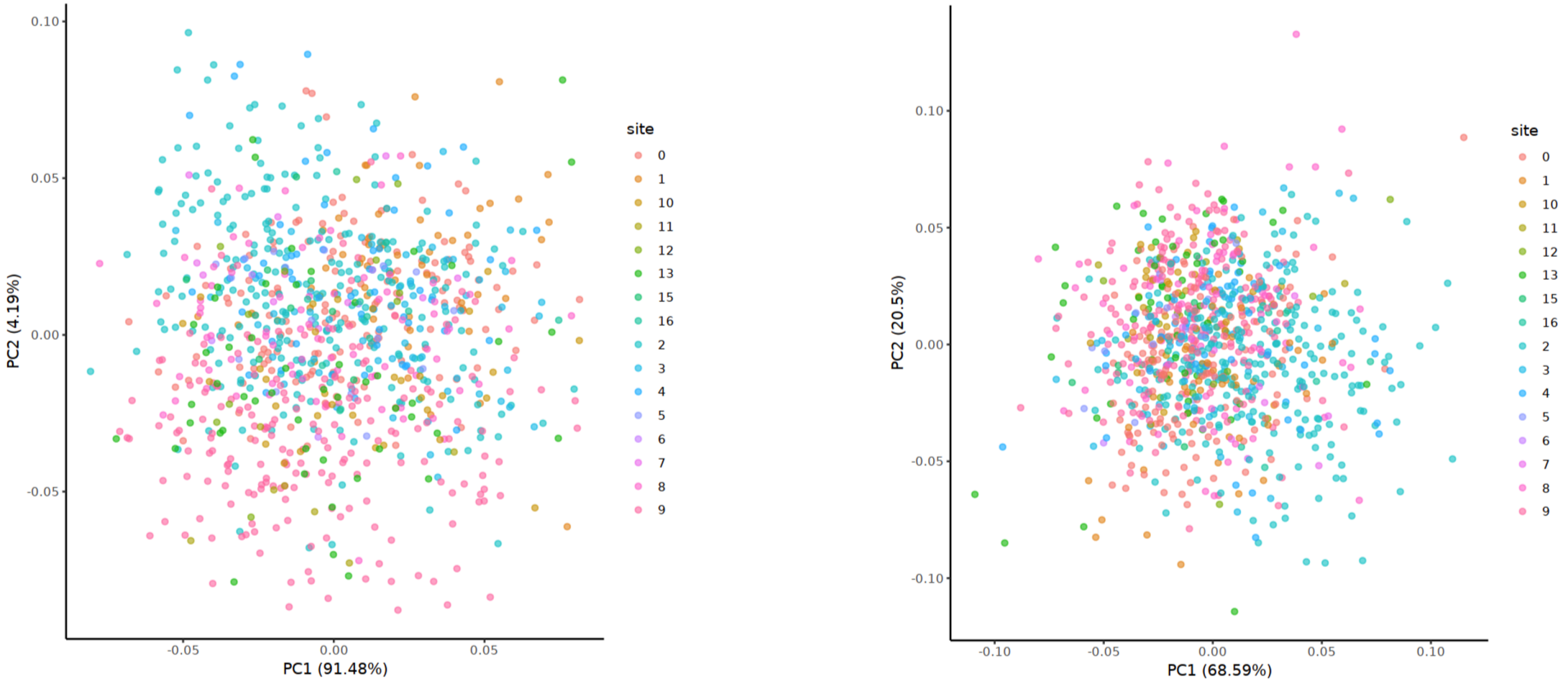
The method was trained on one part of the data (training set) and tested on the rest (test set).

In the following figures, we plot the first 2 PCs of the PCA computed on the test set before and after applying SVA.





First 2 Principal Components of the test set before (left) and after (right) applying SVA. Subjects (points) were colored according to **Sex**: increased harmonization is obtained



First 2 Principal Components of the test set before (left) and after (right) applying SVA. Subjects colored according to Site: here some batch effects seem to persist.

BA contribution to MRI data harmonization

Lombardi A, Amoroso N., Diacono D., Monaco A., Tangaro S., Bellotti R. Extensive Evaluation of Morphological Statistical Harmonization for Brain Age Prediction, *Brain Sciences*, **2020**, *10*(6), 364

Harmonization for age models

- **ComBat** is a batch-effect correction tool used in genomics, that has been also adapted for harmonizing cortical thickness measurements and multi-site DTI studies.
- **ComBat** was found to be an effective harmonization technique that both removes unwanted variation associated with site and preserves biological associations in the data.

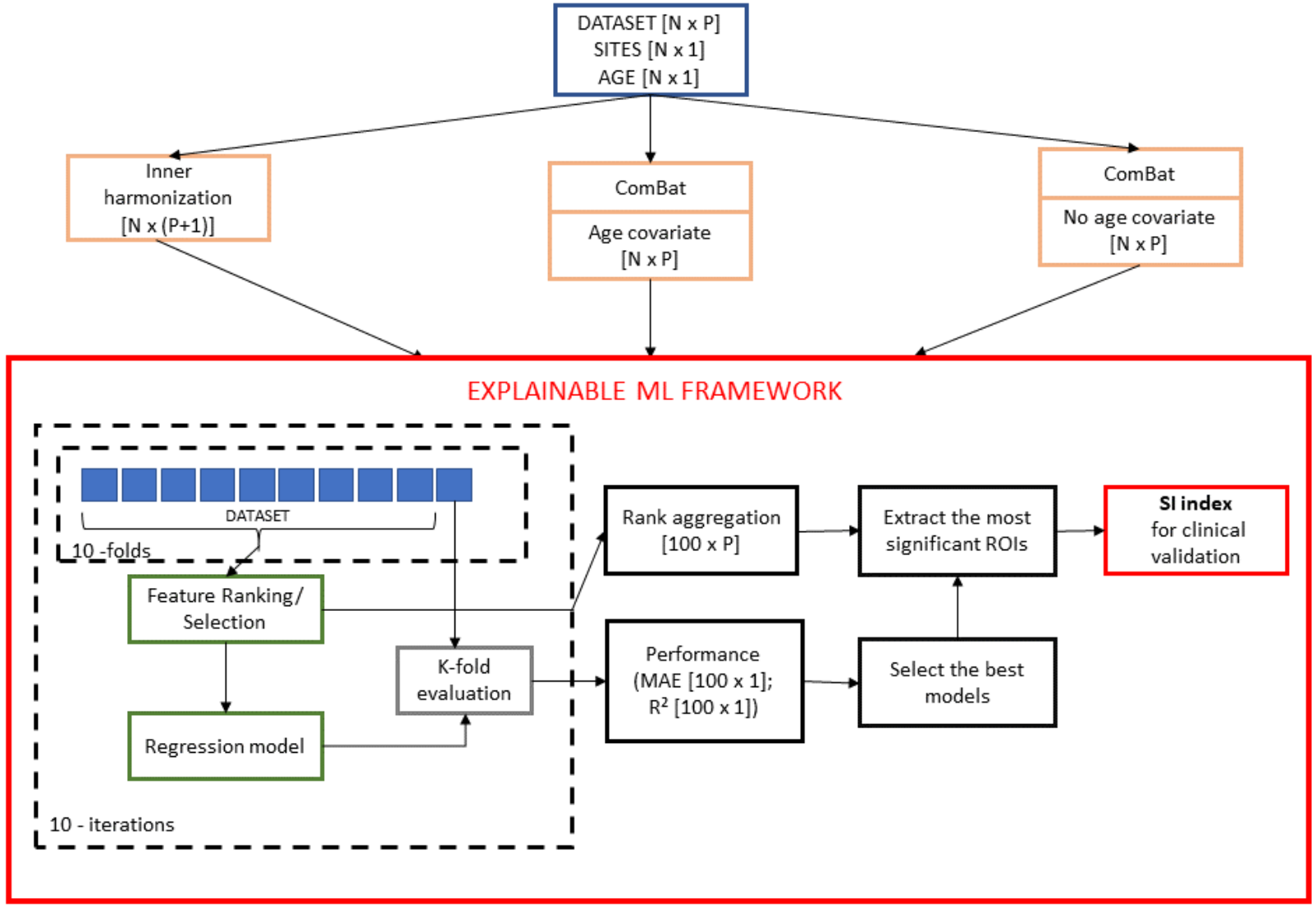
1) What is the best strategy?
2) Is the accuracy of age models related to a specific harmonization technique?
3) Is there a consensus among different harmonization strategies?

Presented by Angela Lombardi in February's AIM meeting

Johnson,W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007, 8, 118–127.

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.; et al. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 2017, 161, 149–170.

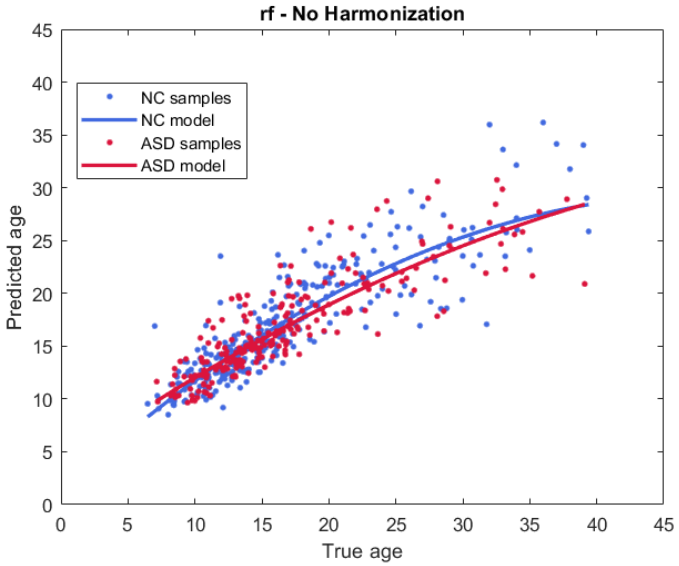
Explainable framework

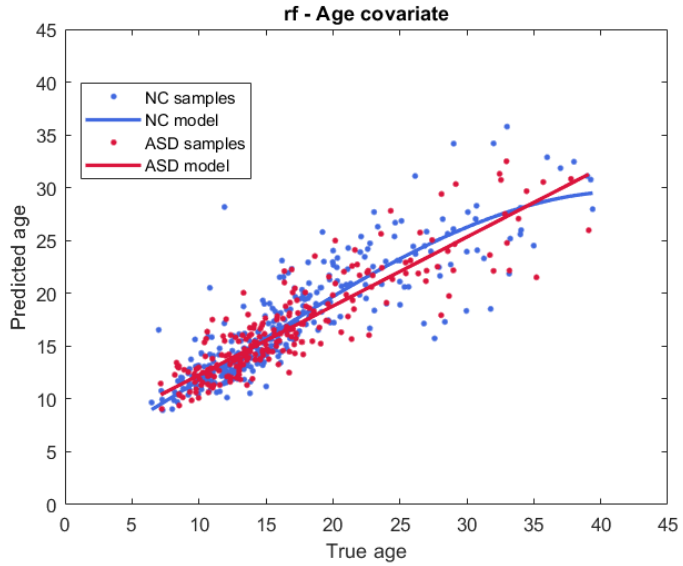


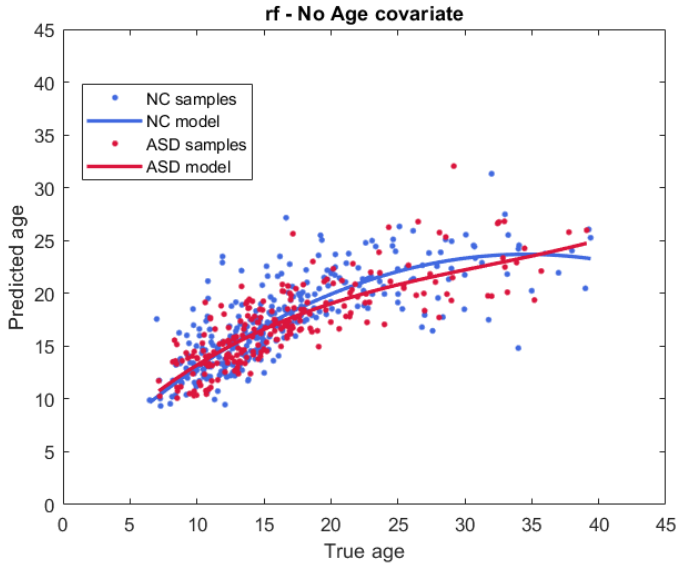
Results in age prediction

MAE

Harmonization technique	NC			ASD		
	SVR	RF	Lasso	SVR	RF	Lasso
No harmonization	2.66 ± 0.38	2.59 ± 0.37	2.79 ± 0.37	2.56 ± 0.49	2.54 ± 0.45	2.80 ± 0.51
Age covariate	2.70 ± 0.36	2.62 ± 0.32	2.68 ± 0.37	2.46 ± 0.40	2.46 ± 0.39	2.68 ± 0.36
No age covariate	4.40 ± 0.55	3.67 ± 0.54	4.82 ± 0.53	4.28 ± 0.58	3.29 ± 0.55	4.56 ± 0.59







Clinical interpretability

In order to verify if the harmonization strategies affected the most significant age-related regions of interest, we evaluated the overlap between the two sets of selected ROIs resulting from two different harmonization strategies:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Index of stability of the most significant anatomical regions for the age prediction with respect to the adopted harmonization strategy:

NC

$$J(A,B) = 0.45,\ p = 0.01$$

ASD

$$J(A,B) = 0.29,\ p = 0.005$$

The results show that the age covariate harmonization and no-harmonization techniques yield comparable results in terms of performance for both groups of subjects, while the statistical harmonization seems to affect the most age-related predictive features.
The proposed framework provides a robust set of relevant features by means of an objective comparison of the outcomes resulting from different harmonization strategies: it could strengthen the relevance of clinical considerations.

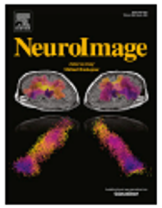
Recently...

NeuroImage 208 (2020) 116450

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan

Raymond Pomponio [a,*], Guray Erus [a], Mohamad Habes [a,b], Jimit Doshi [a], Dhivya Srinivasan [a], Elizabeth Mamourian [a], Vishnu Bashyam [a], Ilya M. Nasrallah [a,g], Theodore D. Satterthwaite [l], Yong Fan [a], Lenore J. Launer [c], Colin L. Masters [d], Paul Maruff [d], Chuanjun Zhuo [e,f], Henry Völzke [h], Sterling C. Johnson [i], Jurgen Fripp [j], Nikolaos Koutsouleris [k], Daniel H. Wolf [l], Raquel Gur [g,l], Ruben Gur [g,l], John Morris [m], Marilyn S. Albert [n], Hans J. Grabe [o], Susan M. Resnick [p], R. Nick Bryan [q], David A. Wolk [b], Russell T. Shinohara [a,r,s], Haochang Shou [a,r,2], Christos Davatzikos [a,**,1,2]

[a] Center for Biomedical Image Computing and Analytics, Department of Radiology, University of Pennsylvania, USA
[b] Department of Neurology, University of Pennsylvania, USA
[c] Laboratory of Epidemiology and Population Sciences, National Institute on Aging, USA
[d] Florey Institute of Neuroscience and Mental Health, University of Melbourne, Australia
[e] Tianjin Mental Health Center, Nankai University Affiliated Tianjin Anding Hospital, Tianjin, China
[f] Department of Psychiatry, Tianjin Medical University, Tianjin, China
[g] Department of Radiology, University of Pennsylvania, USA
[h] Institute for Community Medicine, University of Greifswald, Germany
[i] Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, USA

Dataset of 10477 typical subjects [3-96 years]

ComBat-GAM

Not limited to linear model for age trends, but introduces Generalized Additive Model (GAM)

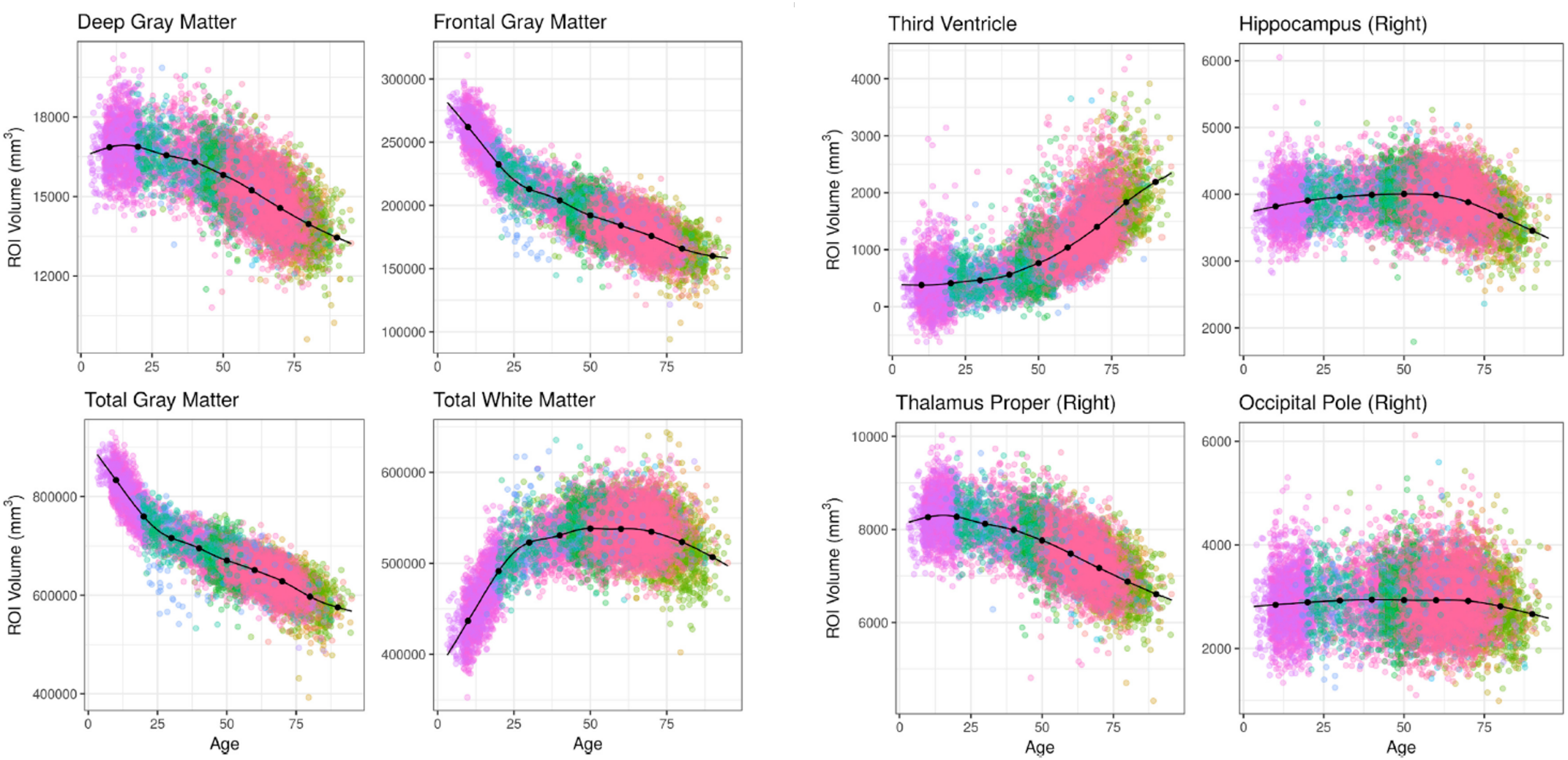
$$Y^*_{ijk} = (Y_{ijk} - f_k(x_{ij}, z_{ij}, w_{ij}) - g^*_{ik}) / d^*_{ik} + f_k(x_{ij}, z_{ij}, w_{ij})$$

location

scale

non-linear function of age, sex, TIV

Age trends for selected ROI (Pomponio et al. 2020)



Final considerations from Pomponio et al. 2020

- The authors made publicly-available:
  - visualization tool we provide as a product of the LIFESPAN dataset (https://rpomponio.shinyapps.io/neuro_lifespan/)
  - a package that enables users to apply ComBat-GAM on their own datasets (https://github.com/rpomponio/neuroHarmonize)

Our analyses have focused primarily on typically-developing and typically-aging participants, establishing age trends of brain regions for healthy controls. We included participants without neurological or psychiatric disorders; however, to harmonize studies which have a specific neurological or psychiatric disease as a focus, data from an appropriate control population is required. Patient data should then follow the same harmonization transformations, but patients should not be used in the calculation of the harmonization model. This is because the underlying assumption behind our approach is that each cohort's measurements were drawn from the same distribution of values, albeit differing by age, sex, and intra-cranial volume (ICV). Patients with structural brain alterations could violate this assumption and, further, including them in the harmonization would attenuate disease-related effects. Hence, the age trend that we provided through the web-interface can serve as a reference based on large control population over a wide age range, and assuming a sufficient control sample is available, could assist with the harmonization task of relatively small pathologic studies, which is otherwise unfeasible.

Conclusions

- Milestone 2021
  - **AIM.1: Valutazione dell'impatto delle diverse strategie implementate per l'armonizzazione dei dati e identificazione delle strategie ottimali rispettivamente per studi MRI/Mammografici/PET multicentrici**

- Outline di un possibile lavoro da sviluppare in sinergia:
  - Dataset ABIDE + ADNI (cases and controls matched for gender and age)
  - Modello ComBat-GAM sui controlli (Location and Scale factors used to harmonize cases across sites)
  - Evaluation of the site CI after site harmonization
  - Evaluation of the possible improvement in predicting subjects' age (MAE), and study of age trajectories in patients' cohorts.
  - Evaluation of the possible improvement in case-control separation (AUC) in "separable" (AD vs. controls) and "barely separable" (ASD vs. controls) problems.

thank you