

Towards scanner-independent MRI markers: a harmonization service for the RIN Network

Damiano Archetti^a, Silvia De Francesco^a, Anna Nigri^b, Maria Grazia Bruzzone^b, Fabrizio Tagliavini^b, Raffaele Lodi^c, Alberto Redolfi^a, RIN – Neuroimaging Network

a – Neuroinformatics Lab, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli; b - Neuroradiology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Via Celoria 11, 20133-Milan, Italy ; c - IRCCS Istituto delle Scienze Neurologiche di Bologna, Via Altura 3, 40139-Bologna, Italy .

- MRI harmonization models can effectively mitigate scanner effects also when shared acquisition protocols are in effect
- Generalizability of harmonization model requires additional training on diverse data

P4-0524

CONTACT

darchetti@fatebenefratelli.eu
sdefrancesco@fatebenefratelli.eu
aredolfi@fatebenefratelli.eu
<https://www.reteneuroscienze.it/en>



INTRODUCTION	METHODS	RESULTS	CONCLUSIONS																																																																														
MRI acquisition parameters influence image quality and markers derived from scans, and even if this can be mitigated with shared acquisition protocols, some residual scanner bias remains detectable. Post-hoc harmonization algorithms can help reduce inter-scanner variability. Here, we present two harmonization models based on data from the Italian Neuroscience Network (RIN)¹, a consortium of Italian Research hospitals that supports and promotes Neuroscience research, with the aim of removing residual variability and testing the ability to harmonize MRI-derived volumetric and thickness markers from independent datasets to RIN data.	Training data consisted of T1w-3D MRI scans from 297 RIN participants. Data for independent validation were acquired from various datasets, totaling 748 T1w-3D MRI scans. Eligible individuals had a diagnosis of Alzheimer’s Dementia, Frontotemporal Dementia, Dementia with Lewy bodies or Healthy controls, Tab.1 shows participant details. Volumes and thicknesses of relevant cortical and subcortical regions were computed with Freesurfer V7.3.2 and used to train harmonization models based on Combat¹ and Neuroharmony², imposing a condition to conserve sex-, age- and disease-related variability. Harmonization performance was assessed via a measure of statistical concordance, cross-set similarity was assessed via a distance metric (i.e.: difference of the respective Euclidean distances).	<table><tr><th></th><th>AGE (yrs)</th><th>SEX (F/M)</th><th colspan="4">Diagnostic classes</th><th colspan="3">Scanners</th></tr><tr><th></th><th></th><th></th><th>HC</th><th>AD</th><th>DLB</th><th>FTD</th><th>GE</th><th>Philips</th><th>Siemens</th></tr><tr><td>Train Set</td><td>68.6±8.5</td><td>152/145</td><td>118</td><td>106</td><td>27</td><td>46</td><td>64</td><td>118</td><td>115</td></tr><tr><td>Test Set</td><td>71.2±8.6</td><td>350/398</td><td>172</td><td>299</td><td>108</td><td>169</td><td>143</td><td>363</td><td>242</td></tr></table> Table 1: train and test sets characteristics. The results show that harmonization models remove scanner effects from the train set but generally fail to harmonize external data (Tab.2). Combat was able to align independent data to the RIN reference only when few acquisition parameters differed from the RIN protocol, while Neuroharmony also worked in very different settings (Fig.1). <table><tr><th></th><th></th><th>HC</th><th>AD</th><th>DLB</th><th>FTD</th></tr><tr><td rowspan="3">Train Set</td><td>Unharmonized</td><td>95%</td><td>99%</td><td>100%</td><td>99%</td></tr><tr><td>Combat</td><td>100%</td><td>100%</td><td>100%</td><td>100%</td></tr><tr><td>Neuroharmony</td><td>100%</td><td>100%</td><td>100%</td><td>100%</td></tr><tr><td rowspan="3">Test Set</td><td>Unharmonized</td><td>93%</td><td>94%</td><td>44%</td><td>71%</td></tr><tr><td>Combat</td><td>87%</td><td>93%</td><td>52%</td><td>64%</td></tr><tr><td>Neuroharmony</td><td>93%</td><td>92%</td><td>49%</td><td>69%</td></tr></table> Table 2: Train harmonization model performance in terms of marker concordance. Acronyms: F = female; M = male; HC = healthy control; AD = Alzheimer’s disease; DLB = Dementia with Lewy bodies; FTD = frontotemporal dementia Figure 1: difference in distance between unharmonized data and RIN reference and distance between harmonized data and RIN reference for Combat (top) and Neuroharmony (bottom) models as a function of the number of acquisition parameters of test set scans in disagreement with the RIN acquisition protocol.		AGE (yrs)	SEX (F/M)	Diagnostic classes				Scanners						HC	AD	DLB	FTD	GE	Philips	Siemens	Train Set	68.6±8.5	152/145	118	106	27	46	64	118	115	Test Set	71.2±8.6	350/398	172	299	108	169	143	363	242			HC	AD	DLB	FTD	Train Set	Unharmonized	95%	99%	100%	99%	Combat	100%	100%	100%	100%	Neuroharmony	100%	100%	100%	100%	Test Set	Unharmonized	93%	94%	44%	71%	Combat	87%	93%	52%	64%	Neuroharmony	93%	92%	49%	69%	Post-hoc harmonization can further reduce residual scanner-related variability also where consistent acquisition protocols are applied. However, generalizability remains limited. Our models could not consistently harmonize data from other cohorts, likely due to large differences in acquisition protocols, scanner characteristics, and data distributions. This highlights the need for caution when applying pre-trained harmonization models across datasets and for the development of even more fine grained, cohort-specific models.
	AGE (yrs)	SEX (F/M)	Diagnostic classes				Scanners																																																																										
			HC	AD	DLB	FTD	GE	Philips	Siemens																																																																								
Train Set	68.6±8.5	152/145	118	106	27	46	64	118	115																																																																								
Test Set	71.2±8.6	350/398	172	299	108	169	143	363	242																																																																								
		HC	AD	DLB	FTD																																																																												
Train Set	Unharmonized	95%	99%	100%	99%																																																																												
	Combat	100%	100%	100%	100%																																																																												
	Neuroharmony	100%	100%	100%	100%																																																																												
Test Set	Unharmonized	93%	94%	44%	71%																																																																												
	Combat	87%	93%	52%	64%																																																																												
	Neuroharmony	93%	92%	49%	69%																																																																												
REFERENCES																																																																																	
[1] Nigri A, et al. (2022) Quantitative MRI Harmonization to Maximize Clinical Impact: The RIN-Neuroimaging Network. <i>Front Neurol</i> 13:855125. doi: 10.3389/fneur.2022.855125 [2] Fortin JP, et al (2018). Harmonization of cortical thickness measurements across scanners and sites. <i>Neuroimage</i> 167:104-120. doi: 10.1016/j.neuroimage.2017.11.024. [3] Garcia-Dias R, et al (2020). Neuroharmony: A new tool for harmonizing volumetric MRI data from unseen scanners. <i>Neuroimage</i> 220:117127. doi: 10.1016/j.neuroimage.2020.117127.																																																																																	