Physica Medica 110 (2023) 102577

Contents lists available at ScienceDirect

Physica Medica

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

Quality assessment, variability and reproducibility of anatomical measurements derived from T1-weighted brain imaging: The RIN–Neuroimaging Network case study

Paolo Bosco [a], Marta Lancione [a], Alessandra Retico [b], Anna Nigri [c], Domenico Aquino [c], Francesca Baglio [d], Irene Carne [e], Stefania Ferraro [c,f], Giovanni Giulietti [g,h], Antonio Napolitano [i], Fulvia Palesi [j,k], Luigi Pavone [l], Giovanni Savini [m], Fabrizio Tagliavini [n], Maria Grazia Bruzzone [c], Claudia A.M. Gandini Wheeler-Kingshott [j,k,o], Michela Tosetti [a,*], Laura Biagi [a], the RIN – Neuroimaging Network

[a] *Laboratory of Medical Physics and Magnetic Resonance, IRCCS Stella Maris Foundation, Pisa, Italy*
[b] *Pisa Division, INFN – National Institute for Nuclear Physics, Pisa, Italy*
[c] *Neuroradiology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy*
[d] *IRCCS Fondazione Don Carlo Gnocchi Onlus, Milan, Italy*
[e] *Neuroradiology Unit, IRCCS Istituti Clinici Scientifici Maugeri, Pavia, Italy*
[f] *MOE Key Laboratory for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China*
[g] *Neuroimaging Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy*
[h] *SAIMLAL Department, Sapienza University of Rome, Rome, Italy*
[i] *Medical Physics, IRCCS Istituto Ospedale Pediatrico Bambino Gesù, Rome, Italy*
[j] *Neuroradiology Unit, IRCCS Mondino Foundation, Pavia, Italy*
[k] *Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy*
[l] *IRCCS Neuromed, Pozzilli, Italy*
[m] *Neuroradiology Unit, IRCCS Humanitas Research Hospital, Milan, Italy*
[n] *Scientific Direction, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy*
[o] *NMR Research Unit, Department of Neuroinflammation, Queen Square MS Centre, UCL Queen Square, Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom*

ARTICLE INFO

Keywords:
Magnetic Resonance Imaging
T1-weighted MRI
Multicentric study
Reproducibility
Standard Operating Procedures
Neurodevelopment
Neurodegeneration
Brain

ABSTRACT

Initiatives for the collection of harmonized MRI datasets are growing continuously, opening questions on the reliability of results obtained in multi-site contexts.

Here we present the assessment of the brain anatomical variability of MRI-derived measurements obtained from T1-weighted images, acquired according to the Standard Operating Procedures, promoted by the *RIN-Neuroimaging Network*. A multicentric dataset composed of 77 brain T1w acquisitions of young healthy volunteers (mean age = 29.7 ± 5.0 years), collected in 15 sites with MRI scanners of three different vendors, was considered. Parallelly, a dataset of 7 "traveling" subjects, each undergoing three acquisitions with scanners from different vendors, was also used. Intra-site, intra-vendor, and inter-site variabilities were evaluated in terms of the percentage standard deviation of volumetric and cortical thickness measures. Image quality metrics such as contrast-to-noise and signal-to-noise ratio in gray and white matter were also assessed for all sites and vendors.

The results showed a measured global variability that ranges from 11% to 19% for subcortical volumes and from 3% to 10% for cortical thicknesses. Univariate distributions of the normalized volumes of subcortical regions, as well as the distributions of the thickness of cortical parcels appeared to be significantly different among sites in 8 subcortical (out of 17) and 21 cortical (out of 68) regions of i nterest in the multicentric study.

The Bland-Altman analysis on "traveling" brain measurements did not detect systematic scanner biases even though a multivariate classification approach was able to classify the scanner vendor from brain measures with an accuracy of 0.60 ± 0.14 (chance level 0.33).

* Corresponding author at: Viale del Tirreno, 331, 56128 Pisa, Italy.
E-mail address: michela.tosetti@fsm.unipi.it (M. Tosetti).

https://doi.org/10.1016/j.ejmp.2023.102577
Received 11 July 2022; Received in revised form 1 March 2023; Accepted 5 April 2023
Available online 29 April 2023

1. Introduction

In the last decades, non-invasive anatomical measurements derived from Magnetic Resonance Imaging (MRI) of the brain played a pivotal role in the assessment of many diseases such as neurodevelopmental, neurodegenerative, psychiatric and rare conditions. Many of these measurements already demonstrated to be well-suited neuroimaging anatomical biomarkers for the early diagnosis and assessment of Alzheimer's Disease [1–3], frontotemporal dementia [4,5], Parkinson's Disease [6,7] and for the differential diagnosis of other forms of dementia such as Lewy Body Dementia [8,9]. In psychiatric and neurodevelopmental disorders, brain anatomical measurements have been shown either relevant or at least promising in the study of many diseases such as schizophrenia [10,11], major depressive disorder [12,13], autism spectrum disorders [14–16], childhood apraxia of speech [17–19].

However, there are still many challenges to be tackled for the advances in the detection of structural brain biomarkers. There is a strong need for large sample sizes to provide sufficient statistical power for the investigation of groups and subgroups and to deal with relatively small pathology effect size, hence multi centric studies are more and more necessary for the development of both pharmacological and non-pharmacological interventions.

In this context, in most cases there are unclear recommendations for MRI image acquisition and analysis details for multivendor protocols using the standard equipment available in hospitals. Moreover, there are no clear quality control guidelines and reference values of different markers and brain measurements extracted from T1-weighted imaging together with unclear recommendations for retrospective harmonization of already existing data acquired with different protocols [20]. These factors hinder the advances in the field since they represent sources of variability which, in addition to the heterogeneity of the population under exam, often hamper the detection of subtle pathological changes, even in the context of the recent development of advanced and powerful artificial intelligence techniques [21].

For these reasons many initiatives were promoted for the harmonization of MRI acquisitions protocols and data analyses such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/) [22] and Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) [23].

With the same aim, the *RIN - Neuroimaging Network*, an Italian national consortium dedicated to share large-scale multimodal quantitative MRI datasets, promoted the development of guidelines for the data acquisition and processing [24,25]. In this study, we present the results obtained on brain structural MRI measures. In particular, we aimed to measure the anatomical variability of different brain structures, taking into account the influence on these measures of scanner vendor along with different hardware solutions used for data acquisition. In addition, we explored the variability of some image quality metrics which may have indirect effects on the image-derived anatomical measurements.

2. Material and methods

2.1. Description of the datasets

Data were acquired in fifteen sites of the RIN – Neuroimaging Network, equipped with 3 T MRI scanners from three different vendors (Philips Healthcare, GE Healthcare, and Siemens Healthineers). Two distinct studies were conducted. The first considered data acquired in 14 sites (*multicentric study);* the second considered data acquired on a small number of subjects that repeated the acquisition on three sites selected on the basis of scanner vendor and geographical area (*traveling brain study)*.

2.2. Multicentric study

In order to assess the image quality and the anatomical variability in a multicentric framework of cerebral measurements, derived from T1-weighted brain MRI, a dataset composed of 77 brain acquisition obtained in as many young healthy volunteers (45F/32 M, mean age = 29.7 ± 5.0 years, range [21–45] years) was considered. In particular, we collected 29 datasets from vendor 1 (mean age = 30.6 ± 5.4 years, 18F/11 M), 18 from vendor 2 (mean age = 30.4 ± 4.6 years, 10F/8M), and 30 from vendor 3 (mean age = 28.4 ± 4.8 years, 17F/13 M). Details on the subjects recruited in each participant center are reported in Table 1, along with the hardware information of the MR scanner (vendor code and number of channels of the receiving coils).

2.3. Traveling brain study

The inter scanner variability was performed with an additional dataset composed of 7 healthy traveling subjects who underwent three brain MRI acquisitions at three sites equipped with scanners from different vendors.

Two geographic areas (North Area, A_N, and South Area, A_S) were defined in Italy. Among the 7 traveling subjects, 4 subjects (mean age = 31.5 ± 2.2 years, 2F/2M) performed the T1-weighted MRI acquisitions in area A_N, at sites 1, 8, and 10; the remaining 3 subjects (mean age = 28.4 ± 11.0 years, 2F/1M) were acquired in area A_S, at sites 5, 9, and 15 (Table 1).

2.4. MRI brain imaging protocol

One of the main objectives of the *RIN - Neuroimaging Network* was the development of Standard Operating Procedures (SOPs) for the acquisition of a comprehensive MRI protocol for the brain. The complete set of scanning parameters for the acquisition of T1-weighted MRI imaging is reported in Table 2. The datasets, both for the *multicentric study* and the *traveling brain study*, were acquired according to the agreed SOPs.

3. Data analysis

3.1. Image segmentation

The FreeSurfer (FS, v.6.0) analysis pipeline [26,27] was used to carry out the segmentation of the brain in subcortical and cortical substructures (according to Desikan-Killiany parcellation [28]). Firstly, we converted 3D T1-weighted MR brain images from DICOM format to NIfTI format. Secondly, we used the FS pre-processing workflow, known as recon-all analysis pipeline, which processes the input structural MRI scan across several FS functions performing all cortical reconstruction through 31 processing steps. In order to carry out gray matter tissue segmentation, FS takes advantage of a lot of information such as image intensities, global position within the brain and relative position to neighboring brain regions. Based on this information, it uses a probabilistic atlas in which coordinates have anatomical meaning and a Markov Random Field (MRF) model is used to find local spatial relationships between labeled structures. FreeSurfer implements a model based on a mixture of a small number of Gaussians for each structure for each point in the space and a maximum posterior estimate of the model parameters to assign one of the Region of Interest (ROI) labels to each voxel. From the FS segmentations results, we extracted the volumes (mm^3) of the subcortical gray matter structures and the thicknesses (mm) of the cortical regions. Along with the brain structure, FS was used to measure the total intra-cranial volume, which is a well-established measurement for volume normalization across subjects [29]. In Fig. 1A an example of brain structural T1-weighted images is shown together with the overlay of FreeSurfer segmentation results (in false color) of subcortical and cortical gray matter structures (Fig. 1B).

3.2. Quality control

A pipeline for image quality control on T1-weighted dataset was implemented, by using the MRIQC protocol [30]. The following measures were extracted from each dataset for the evaluation of the main quality indicators:

- Contrast-to-Noise Ratio [31] (CNR): the CNR evaluates how separated the distributions of signal intensity of adjacent tissues are. CNR indicates specifically the contrast between GM and WM are. Higher values indicate a better gray matter structure definition with respect to the surrounding areas. Additionally, the contrast-to-noise ratio was evaluated between GM and CSF (CNR_{GMCSF}) in order to investigate the impact of this different contrast on the segmentation of GM structures surrounded by CSF.

- Signal-to-Noise Ratio (SNR): the SNR evaluates how much the signal intensity in a specific region is significant with respect to the noise fluctuations. It is calculated as the ratio between the mean intensity of the considered tissue and its standard deviation in the same region.

- Entropy Focus Criterion [32] (EFC): the EFC uses the Shannon entropy of voxel intensities as an indication of ghosting and blurring. Lower values indicate less artifacts and better image quality.

- Coefficient of Joint Variation (CJV): the CJV of gray (GM) and white matter (WM) was proposed for the evaluation of intensity non-uniformity. Higher values indicate worse image quality due to the presence of heavy head motion and large intensity non-uniformity artifacts.

3.3. Variability assessment

The variability of anatomical measurements of cortical and subcortical regions was assessed through the standard deviation of the measures on the whole multicentric data set, in terms of percentage with respect to the corresponding mean value. Moreover, the minimum and the maximum of the standard deviation were calculated for intra-site, inter-site and intra-vendor scenarios.

The distributions of quality control measurements were also calculated separately for different vendors, for different scanner models and number of elements of the receiving RF coils.

3.4. Statistical analyses

The comparisons of the mean values of the extracted measurements among the participant centers were performed with an ANOVA test and the statistical significance threshold of p-value = 0.01 was set (both uncorrected and with False Discovery Rate correction).

Table 2
Parameters of acquisition for T1-weighted MRI, differentiated for each scanner vendor, as reported in the SOPs developed by RIN – Neuroimaging Network.

Vendor	PHILIPS	GE	SIEMENS
Sequence type	3D FFE	3D FSPGR BRAVO	MP-RAGE
Slice orientation	sagittal	sagittal	sagittal
FOV [mm]	240 × 240	256 × 256	256 × 256
Resolution [mm^3]	1 × 1 × 1	1 × 1 × 1	1 × 1 × 1
Matrix (Base Resolution)	240 x 240	256 x 256	256 x 256
Slice thickness	1	1	1
Slice gap (mm)	–	–	–
Number of slices	175 – 180	175 – 180	175 – 180
Phase Encoding direction	AP	AP	AP
Slice order	Interleaved	Interleaved	Interleaved
NSA/Averages/NEX	1	1	1
TR [ms]	2300	not modifiable	2300
TE [ms]	2.96	3.2	2.96
TI [ms]	900	900	900
Flip angle	9°	9°	9°
Fat Suppression	No	No	No
k-space coverage (Halfscan/ Partial Fourier)	No	No	No
Acceleration factor	SENSE ≤ 2.3	ARC = 2	GRAPPA = 2
Filter	CLEAR on	PURE on	Prescan Normalize on
Bandwidth (Hz/pixel)	191	122	240
Duration	≈ 5 min 30 sec	≈ 5 min 30 sec	≈ 5 min 30 sec

FFE = Fast Field Echo; FSPGR = Fast SPoiled GRadient echo; BRAVO = BRAin VOlume imaging; MPRAGE = Magnetization Prepared Rapid Gradient Echo.

For the *traveling brain study*, Bland-Altman plots were considered to evaluate the agreement between the extracted measures with a different approach and to assess the variability at both subject and traveling brain cohort level. In order to assess potential biases limited to specific regions, paired t-tests were performed for each anatomical measurement for every couple of vendors under analysis.

The segmentation of the images, the quality control and the statistical analyses were performed at a single site, under the same operating system in order to avoid additional sources of variability [33].

3.5. Machine learning experiments on the multicentric data

In order to assess the residual dependency of the brain anatomical measurements on the acquisition characteristics after the application of the SOPs, a simple Support Vector Machine (SVM) classifier [34] was trained on the anatomical measures extracted by the segmentation algorithm to recognize the Vendor (Vendor 1, Vendor 2, Vendor 3) that

Table 1
Details on the technical characteristics of scanners (vendor, number of channels of the receiving head coils) of each site. Both site and scanner vendor were anonymized using a numerical code. For the multicentric study, the number of the subjects recruited at each site and their demographical data are reported. For the traveling brain study, columns indicate the geographical area of each site (North Area, A_N, and South Area, A_S), as well as the corresponding number of acquired subjects.

				Multicentric study			Traveling brain study	
Site	Vendor	Model	Rx Coil [ch]	#subjects	Age	Sex	Area	#subjects
1	1	a	32	5	31.8 ± 1.8	3F/2M	A_N	4
2	1	a	32	6	29.7 ± 4.3	5F/1M		
3	1	b	32	6	34.0 ± 7.1	3F/3M		
4	1	c	32	3	25.0 ± 2.0	1F/2M		
5	1	a	32	5	31.6 ± 6.3	4F/1M	A_S	3
6	1	b	32	4	28.5 ± 5.6	2F/2M		
7	2	d	32	7	29.3 ± 2.7	4F/3M		
8	2	d	16	6	29.3 ± 4.7	4F/2M	A_N	4
9	2	e	8	5	33.4 ± 6.0	2F/3M	A_S	3
10	3	f	64	6	26.3 ± 5.9	4F/2M	A_N	4
11	3	g	8	5	28.0 ± 2.3	1F/4M		
12	3	h	64	7	25.1 ± 3.3	6F/1M		
13	3	h	32	5	32.2 ± 3.0	3F/2M		
14	3	i	12	7	31.1 ± 4.7	3F/4M		
15	3	h	32				A_S	3

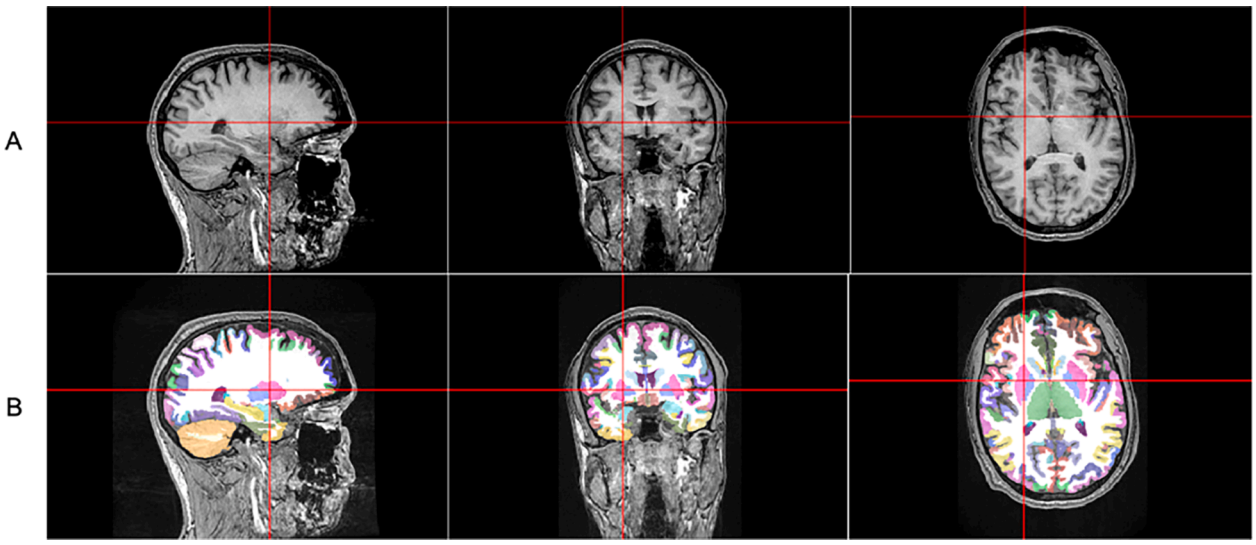

Fig. 1. A. Sagittal, coronal and axial view of a raw T1-weighted of a 3-D image of a representative subject of the dataset. B. overlay of FreeSurfer segmentation results (in false color) of subcortical and cortical gray matter structures.

manufactured the scanner.

The same classification problem was tackled by feeding a linear SVM classifier with the quality metrics (CJV, CNR, SNR_WM, SNR_GM, EFC) extracted through the MRIQC method.

In both the scenarios, the training and testing procedure was performed through a cross validation approach (8-fold cross validation). The classification performance was assessed by measuring the mean accuracy and the standard deviation of the accuracy on the 8 validation folds.

4. Results

Fig. 2 reports the images obtained with all vendors, from two subjects of the travelling brain study, one acquired in the North area (AN, left panel) and one in the South area (AS, right panel).

4.1. Multicentric variability

Fig. 3 shows two examples of distributions of volume and thickness measurements obtained across sites and vendors. Left and right hippocampal values (normalized to the total intracranial volume) are reported together with the thicknesses of the left and right precuneus cortices. The chosen structures are particularly suited for the study of neurodegeneration and aging, since they are strongly involved in cognition and memory.

For a more exhaustive description of the results, the values of volume

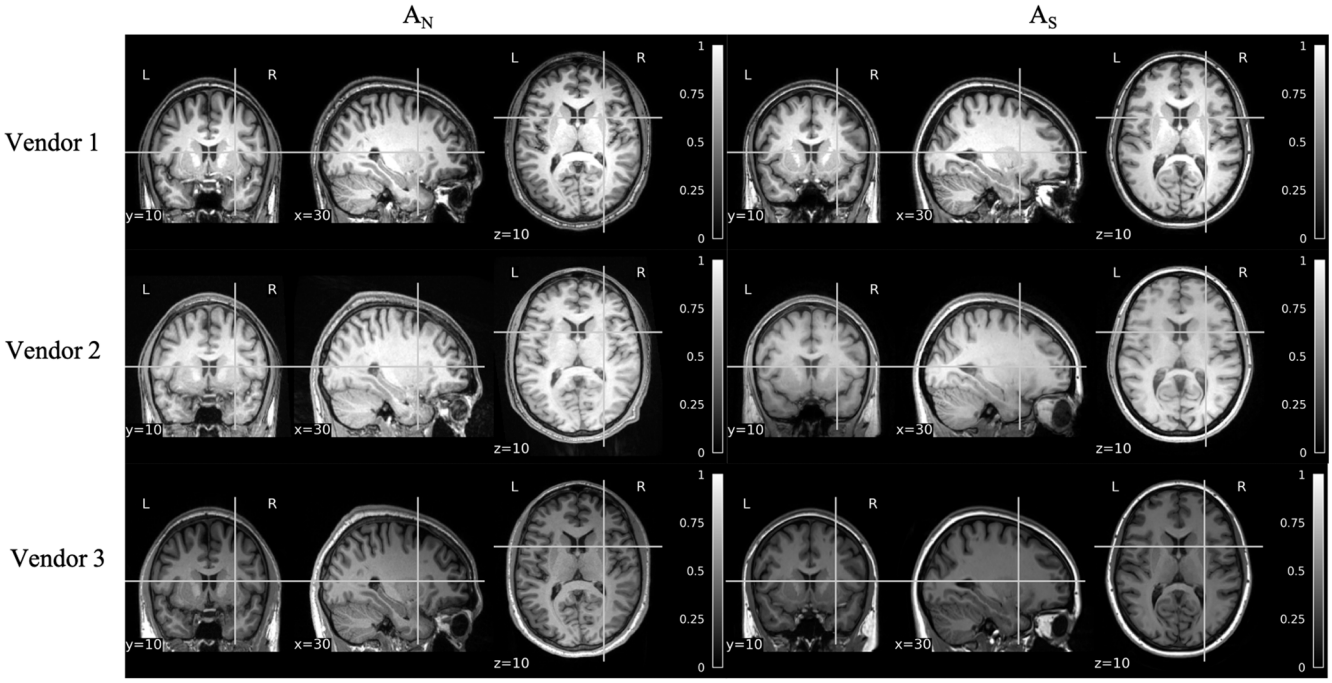

Fig. 2. T1w images acquired with the three vendors on the same two subjects: one subject in the North area, A_N (left column) and one subject in the South area (A_S, right column). The 3D images were registered to the MNI-152 (Montreal Neurological Institute) template and intensity rescaled between 0 and 1 by using as a min reference the 1st intensity percentile and as a max reference the 99th intensity percentile.

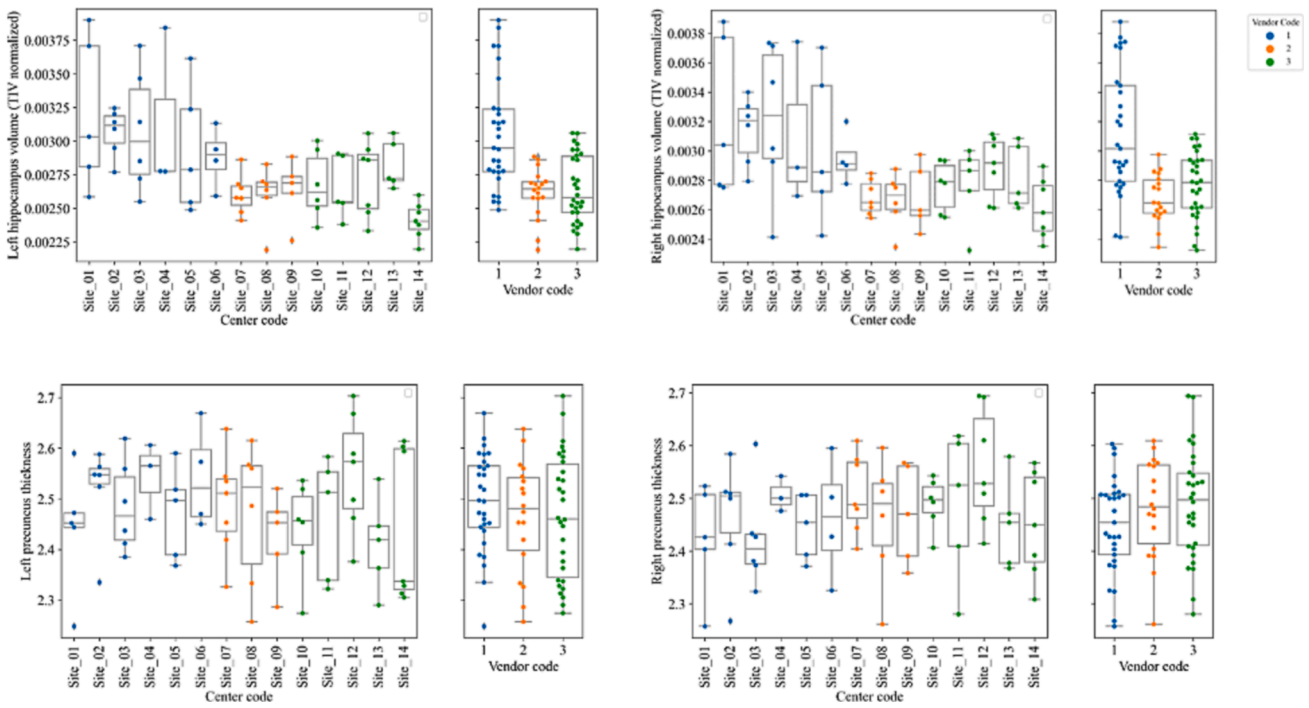

Fig. 3. Examples of box plots of anatomical measurement distributions across the different sites (left panels, site numerical codes on the x-axis) and the different vendors (right panels, vendor numerical codes on the x-axis). In the first row, the volumes of the left and right hippocampus (normalized to the total intracranial volume (TIV)) are reported. In the second row, the thicknesses of the left and right precuneus cortices are shown. The bottom and top edges of each box indicate the 25th and 75th percentile of the measure distribution respectively, and the central line indicates the median. Color code: blue for Vendor 1 (V1), orange for V2, and green for V3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

variability of all the segmented subcortical structures on the multicentric dataset are reported in Table 3. It reports, for each structure, the minimum, the maximum, and the mean of intra-site percentage variability, to assess the range of volume variation within a single site scenario. The variations were evaluated on volume values normalized to the total intra-cranial volume (TIV). Analogously, the intra-vendor and inter-site variabilities are reported. In addition, the global inter-site mean volume values and statistical significance of the ANOVA test on the compatibility of inter-site sampling are shown. The intra-site minimum variation ranges from 1.91% to 10.31%. The maximum ranges from 15.79% to 27.93% (global mean 11.36%). The intra-vendor variability calculated on the three separate datasets ranges from 5.44% (V2) to 17.70% (V3). Considering the average across all the subcortical areas, the mean variabilities among sites of the same vendor are 14.69% for V1, 8.72% for V2, 9.74% for V3; the latter values are comparable to the inter-site variability calculated on the entire dataset which ranges from 11.4% to 19.13%, with an average of 13.84%.

In all the considered cases the highest variability is found for the values of the nuclei accumbens which are small and difficult structures to be segmented by an automated tool.

Table 3

Anatomical variabilities of the measurements of the volume of subcortical structures. Intra-site analysis: minimum, maximum and mean standard deviations calculated on the 14 sites. Intra-vendor analysis: mean standard deviations calculated on datasets from sites of the same vendor (V1, V2, V3 as in Table 1). Inter-site analysis: global standard deviation and mean of volume measurements, and statistical significance of the ANOVA test on the mean compatibility across sites (*pvalue < 0.01). The variations were evaluated on volumes normalized to the Total Intracranial volume (TIV).

ROI		Intra-site variability (%SD)			Intra-vendor variability (%SD)			Inter-site			
		min	max	mean	V1	V2	V3	variability (%SD)	mean volume (mm^3)	pvalue (ANOVA)	pvalue FDR corr
Thalamus	L	2,7	21,5	10,6	15,3	7,4	6,4	11,9	7663	0,311	0,311
	R	1,9	22,6	9,4	14,4	5,4	6,3	11,4	7300	0,081	0,091
Caudate	L	6,1	16,9	10,5	13,3	11,0	8,2	12,2	3592	0,012	0,012
	R	5,9	15,8	9,8	12,9	9,1	8,3	12,8	3663	< 0,001*	< 0,001*
Putamen	L	5,0	26,7	11,6	16,1	8,7	9,4	15,1	5039	0,002*	0,006*
	R	3,9	26,5	10,7	16,4	6,2	8,1	13,9	5091	0,010	0,018
Pallidum	L	5,1	25,5	11,8	15,8	8,5	9,7	12,9	1979	0,294	0,311
	R	4,7	23,6	11,3	15,6	7,2	9,1	13,0	1938	0,069	0,084
Hippocampus	L	5,5	19,7	10,4	13,3	7,1	9,5	13,1	4173	0,001*	0,003*
	R	4,4	17,9	10,0	13,4	6,1	8,0	12,3	4309	0,005*	0,012
Amygdala	L	2,6	22,8	11,0	14,1	6,9	11,7	15,5	1659	< 0.001*	< 0.001*
	R	4,6	25,3	10,5	15,2	7,7	8,8	14,2	1770	0,002*	0,006*
Accumbens area	L	5,3	27,9	14,5	15,3	17,6	15,5	19,1	550	< 0,001*	< 0,001*
	R	10,3	25,9	16,9	16,0	14,1	17,7	18,6	586	0,022	0,031
VentralDC	L	4,2	20,6	11,2	14,2	8,2	10,0	13,2	4186	0,016	0,024
	R	6,1	21,3	11,2	13,9	8,9	9,7	13,2	4134	0,006*	0,012
Brain Stem		6,7	17,3	11,6	14,4	8,1	9,3	12,9	21,692	0,065	0,084

Similarly, Table 4 reports the measured cortical thickness variations across brain cortex parcels. The intra-site minimum variation ranges from 0.3% to 5.0%. The maximum ranges from 5.0% to 15.5% (global mean 5.1%). The intra-vendor variability of brain cortex thickness ranges from 3.1% (V1) to 12.2% (V2). Considering the average across all the cortical areas, the mean variabilities among sites of the same vendor are 5.3% for V1, 5.5% for V2, 5.4% for V3; the latter values are comparable to the inter-site variability of brain cortex thickness calculated on the entire dataset ranges from 3.3% to 10.4%, with an average of 5.7%.

The distributions of the normalized volumes of subcortical regions, as well as the distributions of the thickness of cortical parcels appeared to be significantly different among sites in 8 subcortical (out of 17) and 21 cortical (out of 68) ROIs.

4.2. Quality control measurements

In Fig. 4 the distributions across sites of each quality control metric are reported. For all of them, the intra-site distributions of the values appeared to be peculiar for each considered site, with a variability that is very different with respect to the global one.

In order to disentangle the contributions to the quality metrics variability, the contrast-to-noise ratio and the signal to noise ratio in brain gray matter were aggregated not only by scanner vendor, but also by scanner model and number of channels of the head coils (Fig. 5). The distributions were strongly dependent on the vendor, while neither the specific scanner model nor the number of elements of the head coils seemed to have a significant impact on the considered metrics.

4.3. Machine learning experiments

The linear SVM classifier, trained on the anatomical measurements extracted by the segmentation algorithm on the multicentric study, was able to classify the scanner vendor with an average accuracy of 0.60 $\pm$ 0.14. This value should be compared with the chance level, which is equal to 0.33 for a three-class classification. A similar SVM classifier, trained on the quality control metric extracted through the MRIQC protocol was able to classify the scanner vendor with an average accuracy of 0.87 $\pm$ 0.13 (chance level 0.33).

4.4. Traveling brain variability

Figs. 6 and 7 report the Bland-Altman plots for the assessment of the reproducibility across different vendors of subcortical volume and cortical thickness measurements, respectively. For each Figure, the first row reports the comparisons of the traveling data collected in A_N, while the second row shows the results for the traveling data collected A_S. In none of the cases the mean value of the difference significantly differs from 0 on the basis of a 1-sample *t*-test, indicating that there are no systematic biases.

The mean value of percentage of variations in measuring the volumes of deep structures varies from 0.2% to 1.3%, while the standard deviation ranges in both data sets from about 5% to 8%. The highest values of variation, associated with a lower reproducibility level, are related to nuclei accumbens, which are small and challenging structures to segment, as already stated (purple dots in Fig. 5).

Analogously, for the measurement of cortical thicknesses, the mean value of the percentage of variation goes from 0.2% to 3.1%, with the standard deviation ranging from about 4% to 7%. In these structures, the highest variations in the A_N traveling data set seem to be related to cortical regions with high thickness values such as some temporal regions (temporal pole, the transverse temporal cortex), the insula, and the entorhinal cortex. However, the same trend does not show in the A_S traveling subject data set, where the values outside the 95% distribution boundaries are more spread across the entire range of thickness values.

In the Supplementary material the uncorrected and FDR corrected p values of paired T-Test for the repeated subcortical volume and cortical thickness measurements with different vendors are reported for A_N and A_S subjects.

For the subcortical regions, no significant differences between each couple of vendors were obtained in A_N, while only the measurement of the volume of thalamus resulted statistically significant different between V1 and V2 of A_S.

For the cortical regions, significant differences were found between V1 and V2 of A_N (inferior temporal gyrus, lateral orbito-frontal gyrus, post-central gyrus, superior parietal gyrus) and of As (inferior temporal gyrus). Analogously, V1 and V3 show significant differences both in A_N (inferior temporal gyrus, post-central gyrus) and in A_S (inferior temporal gyrus). No statistically significant differences were obtained between V2 and V3 both in A_N and A_S.

5. Discussion

The work of the RIN – Neuroimaging Network started to address some of the urgent challenges for a full exploitation of MRI biomarkers for the diagnosis and prognosis in neurology field [20]. In particular, the project developed Standard Operating Procedures for the acquisition of T1-weighted MRI of the brain, adapted to multivendor scenarios and suitable for the equipment available in hospitals. On this basis, in this study a set of reference values for different anatomical cerebral structures was extracted from a population of young healthy subjects and the residual variability after the harmonization of the acquisition protocol was assessed in cortical and subcortical regions, by segmenting the images with one of the most widespread techniques (the FreeSurfer utility). We observed a residual intra-site minimum variation that ranges from about 2% to 10% and a maximum intra-site variation that ranges from 16% to 28%. The inter-site variability calculated on the entire multicentric dataset ranges from about 11% to 19% whilst the inter-vendor variability calculated on the entire dataset ranges from 5% to 18%. As expected, the volume variability changes considerably, depending on the intrinsic characteristics of the segmented subcortical structures, with some distributions across sites that are statistically different, even after total intracranial volume normalization, in specific structures such as hippocampus, amygdala, globus pallidus and nucleus accumbens. The same applies to cortical thickness measurements for which the percentage variation is lower since the intra-site minimum variation ranges from 0.3% to 5% and the maximum ranges from about 5% to 16%. The inter-site variability of brain cortex thickness calculated on the entire dataset ranges from 3% to 10% similarly to the inter-vendor variability which ranges from 3% to 12%. The thickness distributions of cortical parcels are statistically different across sites in about 30% of regions, equally distributed among hemispheres (11 ROIs in left and 10 ROIs in right hemisphere).

Multicentric studies, targeted to specific brain region alterations in terms of volume or thickness, usually plan the multicentric acquisition settings to minimize the variability due to acquisition parameters. We observed in this study that, even after an MRI definition of Standard Operating Procedures which minimizes the variability in the acquisition parameters, a complete image harmonization is not achieved. A residual not negligible variability is present due to the test–retest variability combined with the variations in T1-weighted images induced by the input parameters specific to each vendor. Thus, the expected pathological effect (e.g. the amount of cortical thinning in a specific region of interest or the volume enlargement in a deep structure due to pathophysiological mechanisms) in such studies must be compared to this residual variability in order to estimate appropriate sample sizes (both at global as well as at intra-site level).

Quality control measures analyses, indeed, confirmed that the T1-weighted MRI images of the brain are still strongly dependent on the vendor in terms of contrast to noise and signal to noise in different brain tissues even after the definition of Standard Operating Procedures for brain MRI acquisition, in part also observable in Fig. 2. On the other

Table 4
Anatomical variabilities of the measurements of the thicknesses of cortical structures: Intra-site analysis: minimum, maximum, and mean standard deviations calculated on the 14 sites. Intra-vendor analysis: mean standard deviations calculated on datasets from sites of the same vendor (V1, V2, V3 as in Table 1). Inter-site analysis: global standard deviation and mean of thickness measurements and statistical significance of the ANOVA test on the mean compatibility across sites (*pvalue < 0.01).

Cortical ROI		Intra-site variability (%SD)			Intra-vendor variability (%SD)			Inter-site			
		min	max	mean	V1	V2	V3	variability (%SD)	mean thickness (mm)	pvalue (ANOVA)	pvalue FDR corr
banks superior temporal	L	3,6	10,9	5,5	6,2	5,6	5,5	5,8	2,55	0,044	0,116
	R	2,5	7,9	4,6	4,6	5,7	5,0	5,1	2,65	0,108	0,179
caudal anterior cingulate	L	2,2	11,1	7,0	8,7	6,5	7,4	7,7	2,67	0,049	0,116
	R	3,3	11,4	6,5	5,9	6,3	7,8	6,8	2,52	0,374	0,439
caudal middle frontal	L	2,1	5,5	4,1	3,9	4,3	5,0	4,4	2,57	0,178	0,257
	R	0,3	6,4	3,8	3,1	4,7	4,1	4,0	2,54	0,253	0,324
cuneus	L	3,0	11,7	6,0	4,9	8,5	6,3	6,8	1,91	0,266	0,330
	R	1,5	12,0	5,8	4,7	9,3	6,3	7,0	1,95	0,113	0,182
entorhinal	L	3,9	12,9	7,3	8,7	5,9	8,0	8,2	3,37	0,069	0,150
	R	5,0	14,9	9,5	10,0	6,6	11,4	10,4	3,47	0,058	0,130
fusiform	L	1,6	5,0	3,0	3,6	3,3	2,9	3,3	2,79	0,046	0,116
	R	2,2	5,2	3,8	4,6	3,4	3,5	4,4	2,81	0,003*	0,011
inferiorparietal	L	1,6	6,3	3,7	3,8	3,5	5,3	4,3	2,50	0,004*	0,013
	R	2,4	5,2	3,8	3,6	4,6	3,9	4,1	2,52	0,050	0,116
inferiortemporal	L	1,7	7,9	4,2	4,6	4,8	4,3	5,4	2,77	<0,001*	0,001*
	R	2,2	8,0	4,3	4,6	5,4	3,3	5,3	2,79	<0,001*	0,001*
isthmuscingulate	L	2,2	8,9	6,1	6,6	5,4	7,3	6,6	2,45	0,420	0,468
	R	3,5	8,2	5,7	6,3	6,0	6,0	6,1	2,47	0,077	0,158
lateraloccipital	L	2,1	7,4	4,1	4,4	4,4	5,1	5,1	2,21	<0,001*	0,002*
	R	2,5	6,9	4,5	4,5	5,6	5,5	5,6	2,28	<0,001*	0,002*
lateralorbitofrontal	L	2,2	6,8	4,0	4,0	4,5	4,3	5,4	2,68	<0,001*	<0,001*
	R	2,0	7,7	4,5	4,5	4,1	5,4	5,4	2,63	0,001*	0,005*
lingual	L	1,8	8,4	4,9	4,2	6,4	5,1	5,5	2,11	0,155	0,235
	R	2,2	9,3	4,7	3,5	7,8	4,4	5,3	2,12	0,340	0,406
medialorbitofrontal	L	0,7	7,4	5,1	4,9	5,0	5,0	6,6	2,48	<0,001*	<0,001*
	R	1,9	7,7	5,0	4,6	5,2	6,5	6,3	2,51	0,001*	0,005*
middletemporal	L	1,2	5,9	4,2	4,8	4,3	4,6	5,1	2,89	0,001*	0,005*
	R	2,4	6,1	3,5	4,1	3,3	3,7	3,8	2,90	0,184	0,260
parahippocampal	L	4,7	11,2	8,0	8,5	8,6	7,7	8,5	2,89	0,091	0,167
	R	2,9	9,3	6,0	5,8	7,1	6,2	6,6	2,88	0,071	0,151
paracentral	L	2,1	7,0	4,7	4,1	5,9	5,6	5,2	2,47	0,156	0,236
	R	2,6	10,1	4,8	3,9	4,9	6,7	5,5	2,49	0,033	0,097
parsopercularis	L	1,6	7,4	4,7	5,1	5,0	4,8	4,9	2,62	0,381	0,439
	R	2,6	9,2	4,7	5,5	3,8	4,0	4,6	2,61	0,864	0,864
parsorbitalis	L	3,0	9,1	5,0	4,8	5,5	6,3	5,7	2,72	0,023	0,070
	R	3,3	9,0	5,8	5,6	5,8	6,6	6,3	2,72	0,097	0,174
parstriangularis	L	2,0	6,5	3,9	5,2	4,3	4,8	4,8	2,50	0,003*	0,011
	R	2,6	7,1	5,1	5,7	5,0	4,9	5,3	2,47	0,249	0,324
pericalcarine	L	3,8	9,5	6,6	6,4	8,4	6,7	8,1	1,68	0,001*	0,005*
	R	4,0	15,5	8,2	7,3	12,2	7,4	8,9	1,68	0,214	0,297
postcentral	L	1,3	7,1	4,3	4,6	5,0	5,1	5,5	2,12	<0,001*	0,002*
	R	3,0	9,7	5,2	4,4	5,9	6,3	5,9	2,09	0,243	0,324
posteriorcingulate	L	1,8	11,1	5,9	7,1	4,1	6,7	6,3	2,52	0,517	0,558
	R	2,6	6,0	3,8	5,1	2,9	4,0	4,5	2,50	0,001*	0,005*
precentral	L	1,1	5,9	3,4	3,3	3,7	4,5	4,1	2,63	0,040	0,109
	R	2,8	7,6	4,0	3,7	3,6	5,3	4,3	2,58	0,090	0,169
precuneus	L	3,0	6,2	4,3	3,8	4,5	5,0	4,4	2,48	0,576	0,602
	R	1,3	5,7	3,8	3,7	3,8	4,1	3,9	2,47	0,622	0,640
rostralanteriorcingulate	L	2,1	8,6	6,1	7,0	5,7	5,7	6,3	2,93	0,418	0,468
	R	4,0	12,3	6,7	8,3	4,8	6,2	7,4	2,94	0,088	0,167
rostralmiddlefrontal	L	1,5	5,2	3,2	4,2	3,9	3,6	4,5	2,39	<0,001*	<0,001*
	R	2,8	5,8	4,1	4,6	3,7	4,7	5,1	2,35	<0,001*	0,002*
superiorfrontal	L	1,8	8,1	4,4	4,8	4,3	4,8	4,7	2,74	0,100	0,174
	R	2,0	5,8	3,8	3,8	4,4	4,6	5,3	2,71	<0,001*	<0,001*
superiorparietal	L	2,8	6,0	4,3	4,0	4,1	5,1	4,6	2,25	0,080	0,160
	R	0,9	5,7	3,8	4,1	4,5	4,5	4,7	2,23	0,004*	0,015
superiortemporal	L	1,7	5,5	3,5	3,9	3,3	4,0	3,9	2,86	0,005*	0,016
	R	2,0	8,7	4,4	5,2	4,2	3,4	4,3	2,89	0,535	0,569
supramarginal	L	1,4	6,0	3,6	3,5	2,8	5,2	4,0	2,59	0,106	0,180
	R	0,8	5,4	3,4	4,1	3,8	3,8	4,2	2,58	0,001*	0,003*
frontalpole	L	2,0	12,1	7,6	7,5	8,1	7,2	7,9	2,72	0,454	0,498
	R	3,9	10,5	7,9	8,0	7,8	8,8	8,2	2,69	0,262	0,330
temporalpole	L	2,0	13,8	6,8	7,5	9,0	6,5	7,5	3,69	0,840	0,852
	R	0,6	11,6	6,2	8,1	6,3	6,7	7,1	3,76	0,139	0,219
transversetemporal	L	2,7	13,3	7,5	7,4	10,8	7,0	8,2	2,53	0,240	0,324
	R	4,1	14,0	7,1	7,2	9,9	5,6	7,4	2,57	0,296	0,360
insula	L	1,1	6,1	4,3	6,2	4,4	3,4	4,8	3,05	0,039	0,109
	R	2,8	7,8	4,5	5,5	4,7	3,9	4,6	3,07	0,168	0,249

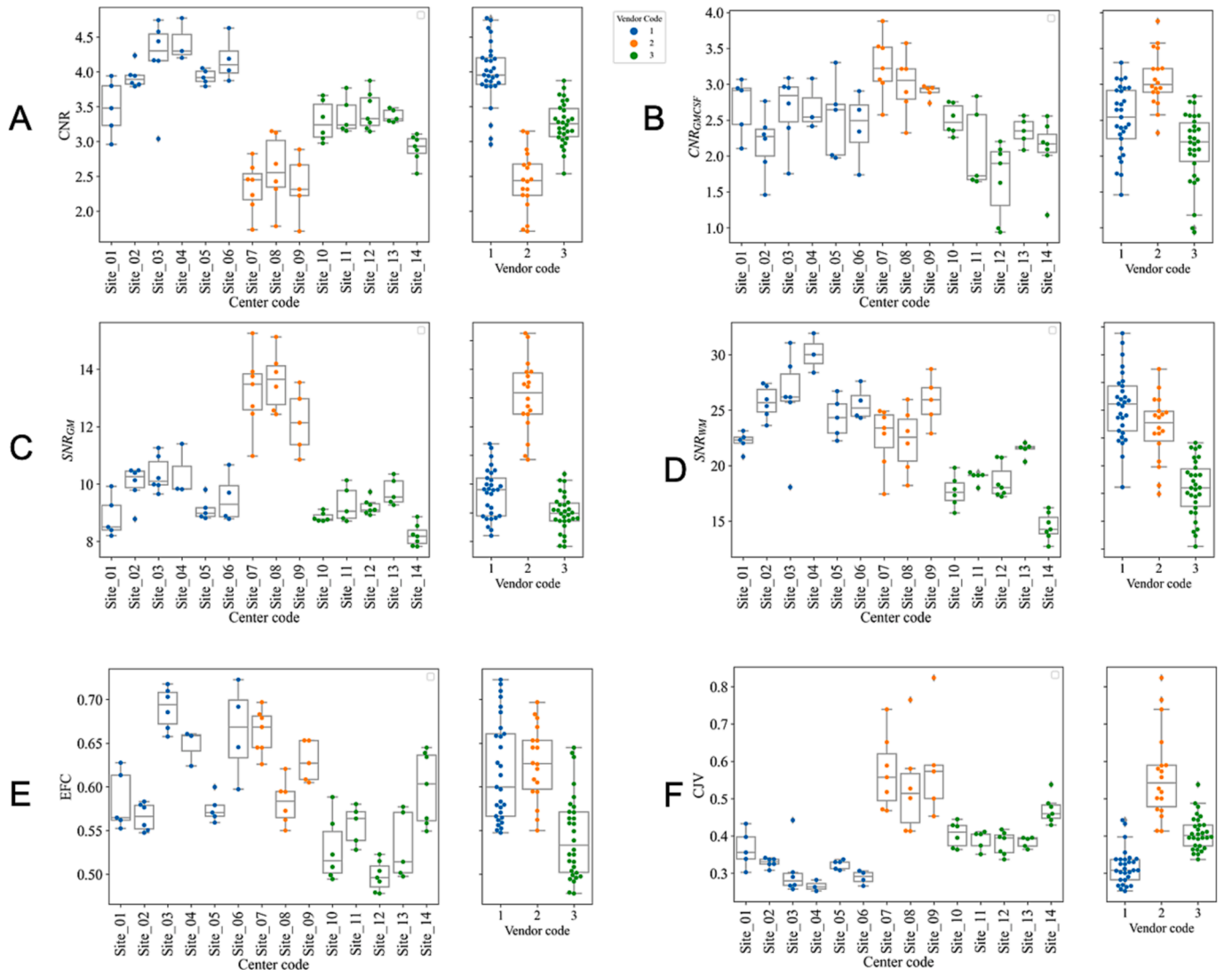

Fig. 4. Box plots of quality control metrics distributions across sites *(left panels, site codes on the x-axis) and vendors (right panels, vendor codes on the x-axis)*: (A) the contrast-to-noise-ratio (CNR), (B) the contrast-to-noise-ratio between GM and CSF (CNR_{GMCSF}) (C) the signal-to-noise ratio for gray (SNR_{GM}) and (D) white matter (SNR_{WM}), (E) the entropy focus criterion (EFC) and (F) the coefficient of joint variation (CJV), are reported. *Color code as in* Fig. 2 *(blue for Vendor 1 (V1), orange for V2, and green for V3).* (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hand, the same analyses ruled out the possibility that systematic signal alterations with a significant impact on the brain structures measurements were due to the number of channels of the head coils or to a specific scanner model of the same vendor (Fig. 5).

However, it is important to point out that differences in quality control metrics distributions could be generated also by the not perfect harmonization of T1-weighted sequences across vendors. An ideal match of different sequences with the same weighting, but from different vendors, would have required to change several variables, often not accessible to the radiographer. On the contrary, the SOPs were developed to help the operator to set the protocol on a commercial scanner, equipped with common sequences, changing simple parameters.

Even if beneficial, the definition of SOPs does not guarantee the similarity in quantitative volumetric measures. The CNR between gray and white matter seems to be the main driving feature for the automated gray matter structures segmentation. Indeed, even though the intensity range is visually well matched for vendor 1 and 2 (Fig. 2), the CNR is different (as shown in Figs. 4 and 5) and some discrepancies appear in intra-subject measurements (Figs. 6 and 7). Conversely, when the difference in CNR is smaller (vendor 1 and 3) despite a remarkable visual difference (Fig. 2), there is a better similarity in gray matter measures on the same subjects.

The residual impact of the scanner vendor on the brain measurements was detectable with a very simple machine learning experiment on vendor prediction which obtained accuracy values not compatible with the chance level. This is in line with previous studies [21] which demonstrated the impact of a not well-designed training set in causing sample and site dependent classifiers, originally thought for the detection of novel anatomical biomarkers of pathologies, which can show significantly positive performances due to underlying and not controlled capability in site classification. For these reasons, particular care should be taken in designing machine learning experiments on multivariate T1-weighted MRI derived measurements and in deep learning approaches which are even more sensitive to subtle intensity variations due to scanner properties even for images acquired with Standard Operating Procedures and well controlled protocols.

Traveling subjects' analyses showed a good agreement in both subcortical and cortical measurements obtained on the same subjects with scanners from different vendors. The residual variability in measuring the volumes of deep structures, calculated as the standard deviation of percentage variation in Bland-Altman plots, ranges on both

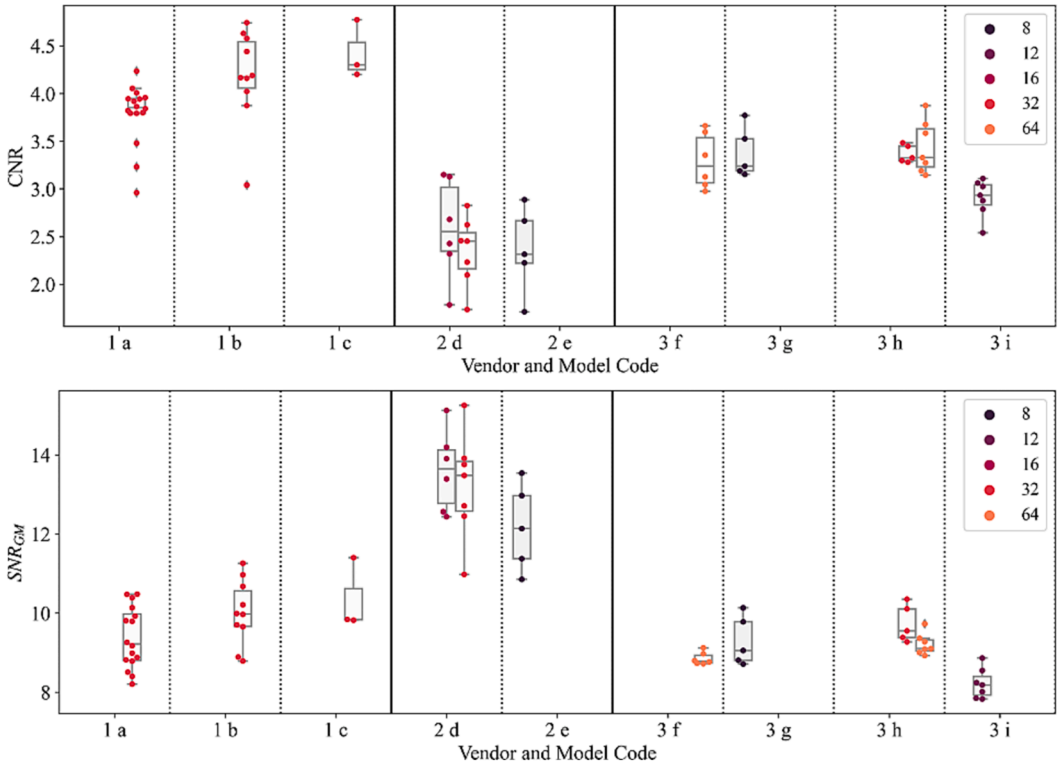

Fig. 5. Distributions of Contrast-to-noise ratio (CNR) and signal-to-noise ratio in gray matter (SNR_{GM}) distributions on data aggregated by vendor and scanner model: 3 models for V1, 2 models for V2 and 4 models for V3 The boxplot colors correspond to the number of channels of the head coils.

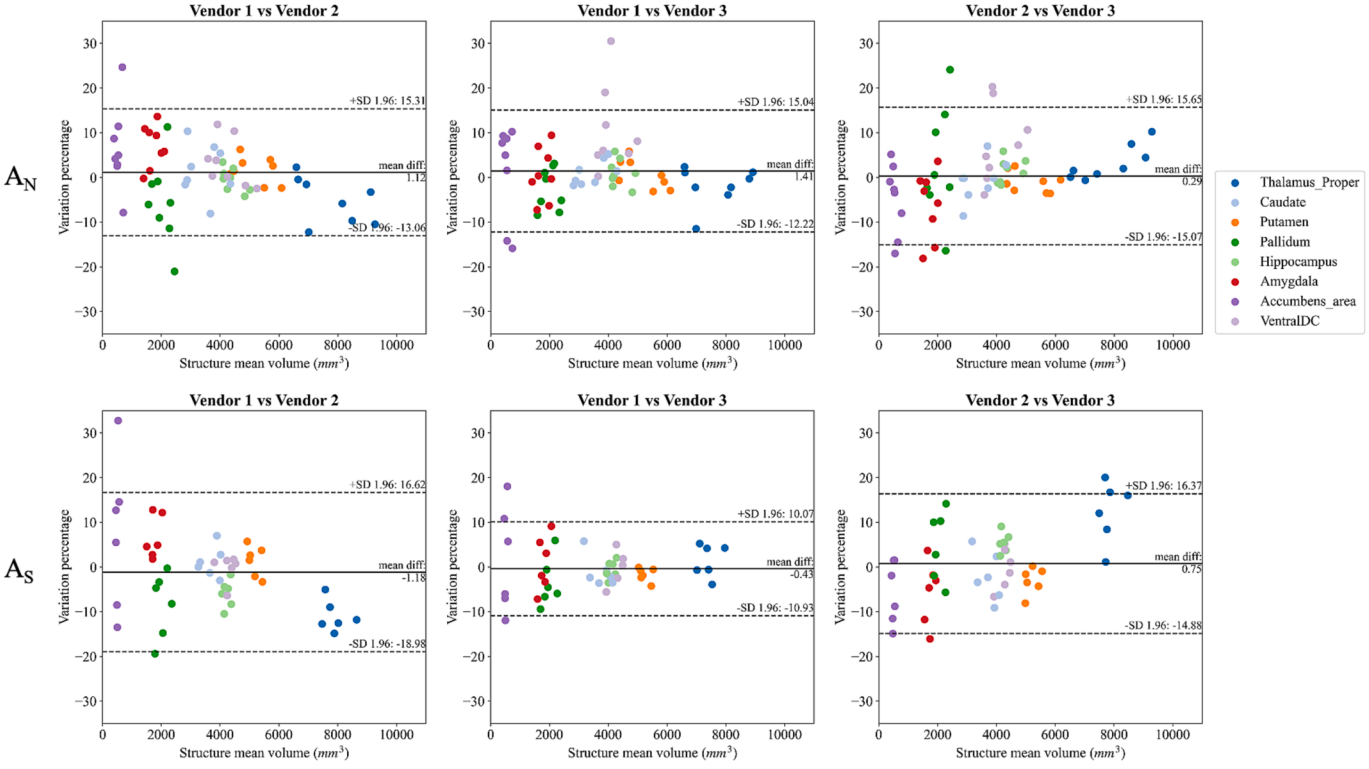

Fig. 6. Bland-Altman plots for the reproducibility assessment of the measurements of subcortical volumes across different vendors (V1, V2, V3), considering the traveling brain data collected in A_N (first row, 4 subjects) and A_S (second row, 3 subjects). To simplify reading, the homologous structures in left and right hemispheres were represented with the same color.

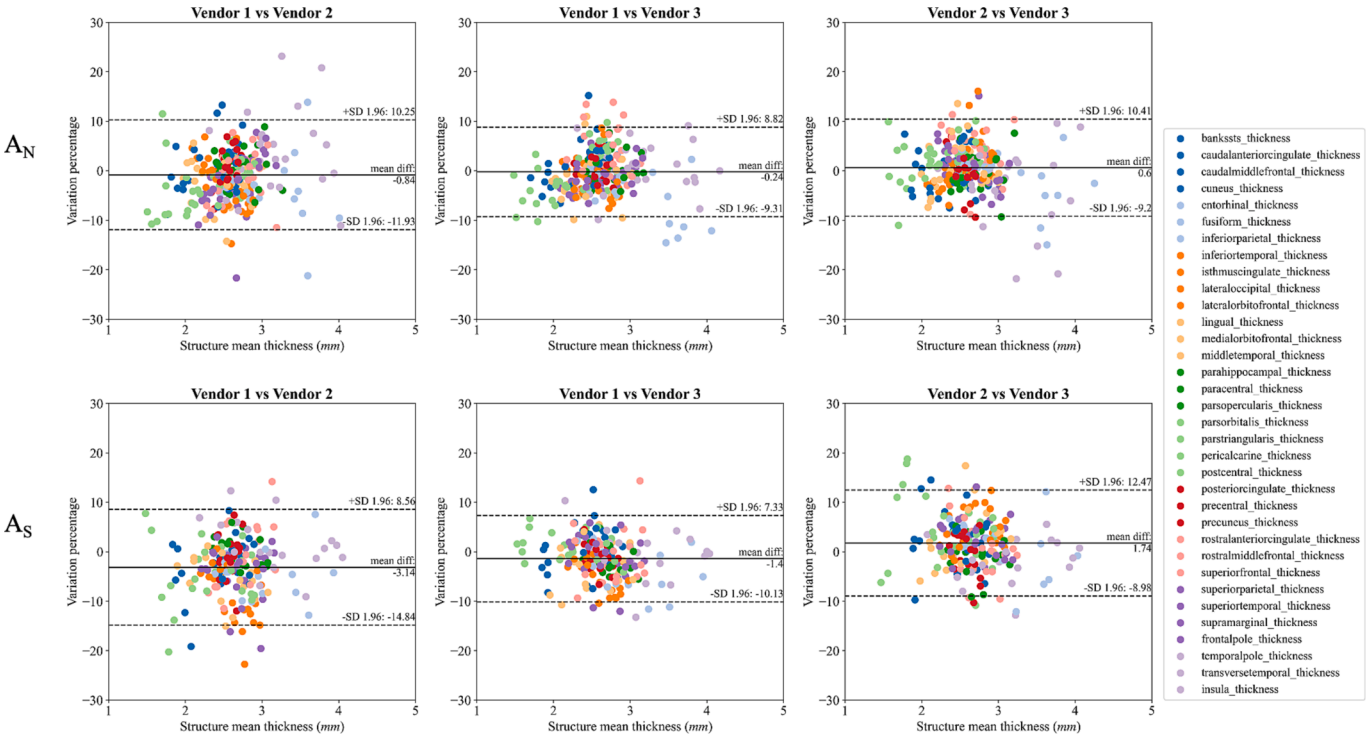

Fig. 7. Bland-Altman plots for the reproducibility assessment of the measurements of cortical thickness of brain parcels across different vendors (V1, V2, V3), considering the traveling brain data collected in A_N (first row, 4 subjects) and A_S (second row, 3 subjects). To simplify reading, the homologous structures in left and right hemispheres were represented with the same color.

data sets from about 5% to 8% depending on the considered structures. The highest values of variation, indicating lower levels of reproducibility, are related to nuclei accumbens, which are small and challenging structures to segment.

Regarding the measurements of cortical thickness, the standard deviation values range in both data sets from about 4% to 7% where the highest variations in the A_N traveling data set seem to be related to cortical regions with high thickness values such as the temporal regions (temporal pole, the transverse temporal cortex), the insula, and the entorhinal cortex.

Except for the finding on the difference between V1 and V2 of A_S in the thalamus, the specific areas that show statistically significant alterations are cortical regions: the inferior temporal gyrus, the lateralorbitofrontal gyrus, the postcentral gyrus and the superior parietal gyrus. The main contribution to the augmented variability in these regions may be related to the increased test–retest variability [35].

The order of magnitude of these intra-subject percent variations must be put in the context of normal aging or pathological alterations such as those related to neurodegenerative processes [1]. For example, the pattern of atrophy due to aging is threefold milder in normal aging than Alzheimer's Disease (AD) (5 vs. 18% in the medial temporal lobe [36,37]) and the annual rate of atrophy in these areas is significantly less pronounced (0.5 vs. 3% in aging vs. AD [38]).

Longitudinal studies on specific anatomical MRI biomarkers of the brain should then be designed according to these variability values and to the expected pathological effect in order to determine the correct experimental sample sizes.

In general, even though quality control measures remain strongly dependent on the scanner vendor even after the definition of the acquisition protocol, the agreement on the traveling brain anatomical measurements suggests that a good reproducibility can be achieved at inter-site level on the same subjects, with an overall variability (5–8%). This intra-subject variability, which is mainly due to the residual differences after image protocol definition, contributes to the measured intra-vendor (5–18%) and mean intra-site variability (9–17%) that were evaluated on different subjects and thus impacted by the inter-subject variability component too. However, the capability of a simple SVM classifier to identify the scanner vendor with an accuracy well above the chance level underlines the risk that multivariate approaches can be particularly sensitive to subtle image intensities changes that can be reflected in high-level anatomical measures.

6. Limitations

The global variability that we assessed in the multicentric experiment has many different sources: the test–retest variability [35,39,40], the inter-vendor variability, the inter-site variability, the inter-subject variability. By disaggregating the data by vendor, the intra-vendor variability was assessed in order to check whether systematic biases could be observed; by performing the traveling brain experiment the inter-vendor/inter-site variability was assessed. However, in all these cases we did not assess the test–retest variability which intrinsically contributes. Since the aim of our study was to assess the variability and reproducibility of morphometric measures derived from T1w images across different sites in a clinical setting after providing some Standard Operating Procedures (which is one of the most common scenarios in clinical research) we must be aware that the test–retest variability will always contribute to the global variability.

As discussed above, another limitation of this study is the not ideal harmonization of T1w sequences across vendors. The Standard Operating Procedures were defined by looking for a compromise between protocol matching and image acquisition in a clinical environment, imposing a uniform spatial resolution and similar time of acquisition. To minimize the variability across vendors, a better harmonization of parameters should be carried out, even if this could request the modification of advanced variables of sequences, not easily feasible in clinical setting.

As described in the method section, the study was designed by using one segmentation algorithm only. The choice was due to its large diffusion in usage and to its very well-known characterization in many

contexts. Different segmentation approaches could in principle have different impacts on the evaluation of the variability of anatomical measurements at both subcortical and cortical levels [41], being less or more prone to subtle signal variations in different brain areas. The results for both intra- and inter-site variability could be affected by the small numerosity of subjects collected in each site (mean and standard deviation of 5.5 ± 1.1 subjects per site), which can produce an overestimation of such variability. The numerosity of the two traveling brain experiments is also limited. Further studies should increment the intra-site sampling in order to reach a more robust statistical evaluation along with bigger traveling brain settings in different sites.

7. Conclusions

The work of the RIN – Neuroimaging Network allowed the acquisition in a multicentric framework of a normative dataset of cerebral T1-weighted MRI of young healthy subjects, by using Standard Operating Procedures. The analyses of the MRI derived measurements allowed the extraction of normative anatomical reference values together with their variability. The acquisitions with the same protocol on a dataset of traveling subjects allowed to disentangle the contribution of subject anatomical variability and the vendor impact. Although a good agreement was shown, the impact of the acquisition scanner on the MRI-derived anatomical measures is still not negligible and detectable through simple data mining approaches, particularly through multivariate classifiers.

8. The RIN Neuroimaging Network

Maria Grazia Bruzzone (Fondazione IRCCS Istituto Neurologico Carlo Besta), Claudia A. M. Gandini Wheeler-Kingshott (Fondazione IRCSS Istituto Neurologico Naz.le Mondino, UCL Queen Square Institute of Neurology, University of Pavia), Michela Tosetti (Fondazione IRCCS Stella Maris), Alberto Redolfi (IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli), Egidio D'Angelo (Fondazione IRCSS Istituto Neurologico Naz.le Mondino, University of Pavia), Gianluigi Forloni (Istituto di Ricerche Farmacologiche Mario Negri IRCCS), Raffaele Agati (IRCCS Istituto delle Scienze Neurologiche di Bologna), Marco Aiello (IRCCS SDN Istituto di Ricerca), Elisa Alberici (IRCCS Istituti Clinici Scientifici Maugeri), Carmelo Amato (Oasi Research Institute-IRCCS), Domenico Aquino (Fondazione IRCCS Istituto Neurologico Carlo Besta), Filippo Arrigoni (Istituto Scientifico, IRCCS E. Medea), Francesca Baglio (IRCCS Fondazione don Carlo Gnocchi onlus), Stefano Bastianello (Fondazione IRCSS Istituto Neurologico Naz.le Mondino), Laura Biagi (Fondazione IRCCS Stella Maris), Lilla Bonanno (IRCCS Centro Neurolesi Bonino Pulejo), Paolo Bosco (Fondazione IRCCS Stella Maris), Francesca Bottino (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Marco Bozzali (Fondazione IRCCS Santa Lucia), Chiara Carducci (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Irene Carne (IRCCS Istituti Clinici Scientifici Maugeri), Lorenzo Carnevale (IRCCS Neuromed), Antonella Castellano (IRCCS Ospedale San Raffaele), Carlo Cavaliere (IRCCS SDN Istituto di Ricerca), Mattia Colnaghi (Istituto Auxologico Italiano IRCCS), Giorgio Conte (Fondazione IRCCS Cà Granda Osp. Maggiore Policlinico), Mauro Costagli (University of Genova; Fondazione IRCCS Stella Maris), Silvia De Francesco (IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli), Greta Demichelis (Fondazione IRCCS Istituto Neurologico Carlo Besta), Valeria Elisa Contarino (Fondazione IRCCS Ca Granda Osp. Maggiore Policlinico), Andrea Falini (IRCCS Ospedale San Raffaele), Stefania Ferraro (Fondazione IRCCS Istituto Neurologico Carlo Besta), Giulio Ferrazzi (IRCCS Ospedale San Camillo), Lorenzo Figà Talamanca (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Cira Fundarò (IRCCS Istituti Clinici Scientifici Maugeri), Simona Gaudino (IRCCS Fondazione Policlinico Universitario Agostino Gemelli), Francesco Ghielmetti (Fondazione IRCCS Istituto Neurologico Carlo Besta), Ruben Gianeri (Fondazione IRCCS Istituto Neurologico Carlo Besta), Giovanni Giulietti (Fondazione IRCCS Santa Lucia), Marco Grimaldi (IRCCS Istituto Clinico Humanitas), Antonella Iadanza (IRCCS Ospedale San Raffaele), Marta Lancione (Fondazione IRCCS Stella Maris), Fabrizio Levrero (IRCCS Ospedale Policlinico San Martino). Raffaele Lodi (IRCCS Istituto delle Scienze Neurologiche di Bologna), Daniela Longo (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Giulia Lucignani (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Martina Lucignani (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Maria Luisa Malosio (IRCCS Istituto Clinico Humanitas), Vittorio Manzo (Istituto Auxologico Italiano, IRCCS), M. Marcella Laganà (IRCCS Fondazione don Carlo Gnocchi onlus), Silvia Marino (IRCCS Centro Neurolesi Bonino Pulejo), Jean Paul Medina (Fondazione IRCCS Istituto Neurologico Carlo Besta), Edoardo Micotti (Istituto di Ricerche Farmacologiche Mario Negri IRCCS), Claudia Morelli (Istituto Auxologico Italiano IRCCS), Alessio Moscato (IRCCS Istituti Clinici Scientifici Maugeri), Antonio Napolitano (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Anna Nigri (Fondazione IRCCS Istituto Neurologico Carlo Besta), Francesco Padelli (Fondazione IRCCS Istituto Neurologico Carlo Besta), Sara Palermo (Fondazione IRCCS Istituto Neurologico Carlo Besta), Fulvia Palesi (Fondazione IRCSS Istituto Neurologico Naz. le Mondino, University of Pavia), Patrizia Pantano (RCCS Neuromed), Chiara Parrillo (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Luigi Pavone (IRCCS Neuromed), Denis Peruzzo (Istituto Scientifico, IRCCS E. Medea), Nikolaos Petsas (IRCCS Neuromed), Alice Pirastru (IRCCS Fondazione don Carlo Gnocchi onlus), Letterio S. Politi (IRCCS Istituto Clinico Humanitas), Luca Roccatagliata (IRCCS Ospedale Policlinico San Martino), Elisa Rognone (Fondazione IRCSS Istituto Neurologico Naz.le Mondino), Andrea Rossi (Ospedale Pediatrico Istituto Giannina Gaslini, UniversitÀ di Genova), Maria Camilla Rossi-Espagnet (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Claudia Ruvolo (IRCCS Centro Neurolesi Bonino Pulejo), Marco Salvatore (IRCCS SDN Istituto di Ricerca), Giovanni Savini (IRCCS Istituto Clinico Humanitas), Fabrizio Tagliavini (Fondazione IRCCS Istituto Neurologico Carlo Besta), Emanuela Tagliente (IRCCS Istituto Ospedale Pediatrico Bambino Gesù), Claudia Testa (IRCCS Istituto delle Scienze Neurologiche di Bologna), Caterina Tonon (IRCCS Istituto delle Scienze Neurologiche di Bologna), Domenico Tortora (Ospedale Pediatrico Istituto Giannina Gaslini), Fabio Maria Triulzi (Fondazione IRCCS Cà Granda Osp. Maggiore Policlinico).

Funding

This study was funded by the Italian Minister of Health under the RC grant, the 5x1000 voluntary contributions to IRCCS Fondazione Stella Maris and under the following RIN projects: RRC-2016-2361095; RRC-2017-2364915; RRC-2018-2365796; RCR-2019-23669119_001 along with the contribution of the Ministry of Economy and Finance (CCR-2017-23669078).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2023.102577.

References

[1] Pini L, Pievani M, Bocchetta M, Altomare D, Bosco P, Cavedo E, et al. Brain atrophy in Alzheimer's Disease and aging. Ageing Res Rev 2016;30:25–48.
[2] Bosco P, Redolfi A, Bocchetta M, Ferrari C, Mega A, Galluzzi S, et al. The impact of automated hippocampal volumetry on diagnostic confidence in patients with suspected Alzheimer's disease: A European Alzheimer's Disease Consortium study. Alzheimer's Dement 2017;13(9):1013–23.

[3] Fennema-Notestine C, Hagler DJ, McEvoy LK, Fleisher AS, Wu EH, Karow DS, et al. Structural MRI biomarkers for preclinical and mild Alzheimer's disease. Hum Brain Mapp 2009;30(10):3238–53.

[4] Rohrer JD. Structural brain imaging in frontotemporal dementia. Biochim Biophys Acta 2012;1822:325–32. https://doi.org/10.1016/J.BBADIS.2011.07.014.

[5] Meyer S, Mueller K, Stuke K, Bisenius S, Diehl-Schmid J, Jessen F, et al. Predicting behavioral variant frontotemporal dementia with pattern classification in multi-center structural MRI data. NeuroImage Clin 2017;14:656–62.

[6] Ibarretxe-Bilbao N, Junque C, Marti MJ, Tolosa E. Brain structural MRI correlates of cognitive dysfunctions in Parkinson's disease. J Neurol Sci 2011;310:70–4. https://doi.org/10.1016/J.JNS.2011.07.054.

[7] Sarasso E, Agosta F, Piramide N, Filippi M. Progression of grey and white matter brain damage in Parkinson's disease: a critical review of structural MRI literature. J Neurol 2021;268:3144–79. https://doi.org/10.1007/S00415-020-09863-8.

[8] Whitwell JL, Weigand SD, Shiung MM, Boeve BF, Ferman TJ, Smith GE, et al. Focal atrophy in dementia with Lewy bodies on MRI: a distinct pattern from Alzheimer's disease. Brain A J Neurol 2007;130(3):708–19.

[9] Hanyu H, Shimizu S, Tanaka Y, Hirao K, Iwamoto T, Abe K. MR features of the substantia innominata and therapeutic implications in dementias. Neurobiol Aging 2007;28:548–54. https://doi.org/10.1016/J.NEUROBIOLAGING.2006.02.009.

[10] Wright IC, Rabe-Hesketh S, Woodruff PWR, David AS, Murray RM, Bullmore ET. Meta-analysis of regional brain volumes in schizophrenia. Am J Psychiatry 2000; 157:16–25. https://doi.org/10.1176/AJP.157.1.16.

[11] Lawrie SM, Abukmeil SS. Brain abnormality in schizophrenia. A systematic and quantitative review of volumetric magnetic resonance imaging studies. Br J Psychiatry 1998;172:110–20. https://doi.org/10.1192/BJP.172.2.110.

[12] Andreescu C, Butters MA, Begley A, Rajji T, Wu M, Meltzer CC, et al. Gray matter changes in late life depression—a structural MRI analysis. Neuropsychopharmacology 2008;33(11):2566–72.

[13] Amico F, Meisenzahl E, Koutsouleris N, Reiser M, Möller HJ, Frodl T. Structural MRI correlates for vulnerability and resilience to major depressive disorder. J Psychiatry Neurosci 2011;36:15. https://doi.org/10.1503/JPN.090186.

[14] Amaral DG, Schumann CM, Nordahl CW. Neuroanatomy of autism. Trends Neurosci 2008;31:137–45. https://doi.org/10.1016/J.TINS.2007.12.005.

[15] Ecker C. The neuroanatomy of autism spectrum disorder: an overview of structural neuroimaging findings and their translatability to the clinical setting. Autism 2017; 21:18–28. https://doi.org/10.1177/1362361315627136.

[16] Bosco P, Giuliano A, Delafield-Butt J, Muratori F, Calderoni S, Retico A. Brainstem enlargement in preschool children with autism: Results from an intermethod agreement study of segmentation algorithms. Hum Brain Mapp 2019;40:7–19. https://doi.org/10.1002/hbm.24351.

[17] Preston JL, Molfese PJ, Mencl WE, Frost SJ, Hoeft F, Fulbright RK, et al. Structural brain differences in school-age children with residual speech sound errors. Brain Lang 2014;128(1):25–33.

[18] Kadis DS, Goshulak D, Namasivayam A, Pukonen M, Kroll R, De Nil LF, et al. Cortical thickness in children receiving intensive therapy for idiopathic apraxia of speech. Brain Topogr 2014;27(2):240–7.

[19] Conti E, Retico A, Palumbo L, Spera G, Bosco P, Biagi L, et al. Autism spectrum disorder and childhood apraxia of speech: early language-related hallmarks across structural MRI study. J Pers Med 2020;10(4):275.

[20] Jovicich J, Barkhof F, Babiloni C, Herholz K, Mulert C, Berckel BNM, et al. Harmonization of neuroimaging biomarkers for neurodegenerative diseases: A survey in the imaging community of perceived barriers and suggested actions. Alzheimer's Dement (Amsterdam, Netherlands) 2019;11(1):69–73.

[21] Ferrari E, Bosco P, Calderoni S, Oliva P, Palumbo L, Spera G, et al. Dealing with confounders and outliers in classification medical studies: The Autism Spectrum Disorders case study. Artif Intell Med 2020;108:101926.

[22] Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27(4):685–91.

[23] Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, Renteria ME, et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav 2014;8(2):153–82.

[24] Nigri A, Ferraro S, Gandini Wheeler-Kingshott CAM, Tosetti M, Redolfi A, Forloni G, et al. Quantitative MRI harmonization to maximize clinical impact: the RIN–neuroimaging network. Front Neurol 2022;13. https://doi.org/10.3389/fneur.2022.855125.

[25] Lancione M, Bosco P, Costagli M, Nigri A, Aquino D, Carne I, et al. Multi-centre and multi-vendor reproducibility of a standardized protocol for quantitative susceptibility Mapping of the human brain at 3T. Phys Med 2022;103:37–45.

[26] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33(3):341–55.

[27] Fischl B. FreeSurfer. FreeSurfer Neuroimage 2012;62(2):774–81. https://doi.org/10.1016/j.neuroimage.2012.01.021.

[28] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31(3):968–80.

[29] Whitwell JL, Crum WR, Watt HC, Fox NC. Normalization of cerebral volumes by use of intracranial volume: Implications for longitudinal quantitative mr imaging. Am J Neuroradiol 2001;22:1483–9.

[30] Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ, et al. MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One 2017;12(9):e0184661.

[31] Magnotta VA, Friedman L. Measurement of signal-to-noise and contrast-to-noise in the fBIRN multicenter imaging study. J Digit Imaging 2006;19(2):140–7.

[32] Atkinson D, Hill DLG, Stoyle PNR, Summers PE, Keevil SF. Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. IEEE Trans Med Imaging 1997;16:903–10. https://doi.org/10.1109/42.650886.

[33] Gronenschild EHBM, Habets P, Jacobs HIL, Mengelers R, Rozendaal N, van Os J, et al. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS One 2012;7(6):e38234.

[34] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97. https://doi.org/10.1007/BF00994018.

[35] Knussmann GN, Anderson JS, Prigge MBD, Dean DC, Lange N, Bigler ED, et al. Test-retest reliability of FreeSurfer-derived volume, area and cortical thickness from MPRAGE and MP2RAGE brain MRI images. Neuroimage: Reports 2022;2: 100086. doi: 10.1016/J.YNIRP.2022.100086.

[36] Bakkour A, Morris JC, Dickerson BC. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. Neurology 2009;72:1048–55. https://doi.org/10.1212/01.wnl.0000340981.97664.2f.

[37] Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN, et al. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. Cereb Cortex 2009;19:497–510. doi: 10.1093/cercor/bhn113.

[38] Fjell AM, Westlye LT, Amlien I, Espeseth T, Reinvang I, Raz N, et al. Minute effects of sex on the aging brain: a multisample magnetic resonance imaging study of healthy aging and Alzheimer's disease. J Neurosci Off J Soc Neurosci 2009;29(27): 8774–83.

[39] Melzer TR, Keenan RJ, Leeper GJ, Kingston-Smith S, Felton SA, Green SK, et al. Test-retest reliability and sample size estimates after MRI scanner relocation. Neuroimage 2020;211:116608. doi: 10.1016/J.NEUROIMAGE.2020.116608.

[40] Maclaren J, Han Z, Vos SB, Fischbein N, Bammer R. Reliability of brain volume measurements: a test-retest dataset. Sci Data 2014:1. https://doi.org/10.1038/SDATA.2014.37.

[41] Palumbo L, Bosco P, Fantacci ME, Ferrari E, Oliva P, Spera G, et al. Evaluation of the intra- and inter-method agreement of brain MRI segmentation software packages: a comparison between SPM12 and FreeSurfer v6.0. Phys Medica 2019; 64:261–72.