Physica Medica 104 (2022) 93–100

Contents lists available at ScienceDirect

Physica Medica

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

MRI data quality assessment for the RIN - Neuroimaging Network using the ACR phantoms

Fulvia Palesi [a,b,1], Anna Nigri [c,1,*], Ruben Gianeri [c], Domenico Aquino [c], Alberto Redolfi [d], Laura Biagi [e], Irene Carne [f], Silvia De Francesco [d], Stefania Ferraro [g,c], Paola Martucci [a], Jean Paul Medina [c], Antonio Napolitano [h], Alice Pirastru [i], Francesca Baglio [i], Fabrizio Tagliavini [j], Maria Grazia Bruzzone [c], Michela Tosetti [e], Claudia A.M. Gandini Wheeler-Kingshott [a,b,k], the RIN – Neuroimaging Network[2]

[a] *Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy*
[b] *Brain Connectivity Center, IRCCS Mondino Foundation, Pavia, Italy*
[c] *U.O. Neuroradiologia, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy*
[d] *Laboratory of Neuroinformatics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy*
[e] *Medical Physics and MR Lab, Fondazione IRCCS Stella Maris, Pisa, Italy*
[f] *IRCCS Istituti Clinici Scientifici Maugeri, Pavia, Italy*
[g] *School of Life Science and Technology, MOE Key Laboratory for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China*
[h] *IRCCS Istituto Ospedale Pediatrico Bambino Gesù, Rome, Italy*
[i] *IRCCS Fondazione Don Carlo Gnocchi Onlus, Milan, Italy*
[j] *Direzione Scientifica, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy*
[k] *NMR Research Unit, Department of Neuroinflammation, Queen Square MS Centre, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, Gower Street, WC1E 6BT London, England, UK*

Abbreviations: qMRI, quantitative magnetic resonance imaging; ACR, American College of Radiology.

* Corresponding author at: at: U.O. Neuroradiologia, Fondazione IRCCS Istituto Neurologico Carlo Besta, Via Celoria 11, 20133 Milan, Italy.
E-mail address: anna.nigri@istituto-besta.it (A. Nigri).

[1] Equally contributing authors.

[2] **RIN - Neuroimaging Network: Collaborators Co-Authors.** 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 SYNLAB SDN), Elisa Alberici (IRCCS Istituti Clinici Scientifici Maugeri), Carmelo Amato (Oasi Research Institute-IRCCS), Filippo Arrigoni (Istituto Scientifico, IRCCS E. Medea), Stefano Bastianello (Fondazione IRCSS Istituto Neurologico Naz.le Mondino), 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), Lorenzo Carnevale (IRCCS Neuromed), Antonella Castellano (IRCCS Ospedale San Raffaele), Carlo Cavaliere (IRCCS SYNLAB SDN), Marta Cazzoli (IRCCS Fondazione don Carlo Gnocchi onlus), Mattia Colnaghi (Istituto Auxologico Italiano IRCCS), Giorgio Conte (Fondazione IRCCS Cà Granda Osp. Maggiore Policlinico), Mauro Costagli (University of Genova; Fondazione IRCCS Stella Maris), Greta Demichelis (Fondazione IRCCS Istituto Neurologico Carlo Besta), Valeria Elisa Contarino (Fondazione IRCCS Ca' Granda Osp. Maggiore Policlinico), Andrea Falini (IRCCS Ospedale San Raffaele), 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), 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), Susanna Lipari (IRCCS Fondazione don Carlo Gnocchi onlus), 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), Silvia Marino (IRCCS Centro Neurolesi Bonino Pulejo), Edoardo Micotti (Istituto di Ricerche Farmacologiche Mario Negri IRCCS), Claudia Morelli (Istituto Auxologico Italiano IRCCS), Alessio Moscato (IRCCS Istituti Clinici Scientifici Maugeri), Francesco Padelli (Fondazione IRCCS Istituto Neurologico Carlo Besta), 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), 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), 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).

https://doi.org/10.1016/j.ejmp.2022.10.008
Received 24 June 2022; Received in revised form 20 September 2022; Accepted 8 October 2022
Available online 12 November 2022

ARTICLE INFO

Keywords:
ACR
Quality control
Multisite

ABSTRACT

Purpose: Generating big-data is becoming imperative with the advent of machine learning. *RIN-Neuroimaging Network* addresses this need by developing harmonized protocols for multisite studies to identify quantitative MRI (qMRI) biomarkers for neurological diseases. In this context, image quality control (QC) is essential. Here, we present methods and results of how the RIN performs intra- and inter-site reproducibility of geometrical and image contrast parameters, demonstrating the relevance of such QC practice.

Methods: American College of Radiology (ACR) large and small phantoms were selected. Eighteen sites were equipped with a 3T scanner that differed by vendor, hardware/software versions, and receiver coils. The standard ACR protocol was optimized (in-plane voxel, post-processing filters, receiver bandwidth) and repeated monthly. Uniformity, ghosting, geometric accuracy, ellipse's ratio, slice thickness, and high-contrast detectability tests were performed using an automatic QC script.

Results: Measures were mostly within the ACR tolerance ranges for both T1- and T2-weighted acquisitions, for all scanners, regardless of vendor, coil, and signal transmission chain type. All measurements showed good reproducibility over time. Uniformity and slice thickness failed at some sites. Scanners that upgraded the signal transmission chain showed a decrease in geometric distortion along the slice encoding direction. Inter-vendor differences were observed in uniformity and geometric measurements along the slice encoding direction (i.e. ellipse's ratio).

Conclusions: Use of the ACR phantoms highlighted issues that triggered interventions to correct performance at some sites and to improve the longitudinal stability of the scanners. This is relevant for establishing precision levels for future multisite studies of qMRI biomarkers.

Introduction

The identification of early quantitative magnetic resonance imaging (qMRI) biomarkers is becoming a primary target for the diagnosis, prognosis, and understanding of mechanisms underlying neurological diseases. The main advantage of qMRI is the objectivity of the measures, unlike the classic approach based only on observation. Quantitative MRI - if clinically adopted - could indeed produce numerical features able to objectify and support the diagnostic and prognostic process.

In this context, "Big Data" collection has become the new frontier of innovation [1,2] because analysis of large amounts of data is fundamental to enhance the efficiency, quality, and prediction power of biomarkers based on qMRI [3,4]. In addition, the aggregation of different types of data, i.e. radiological, genetic, histopathological, and neurological, is paving the way for using advanced machine learning or deep learning techniques to stratify the population into classes not otherwise identifiable through a classical clinically-driven approach [3,5–8]. To this aim, in the last few years several neuroimaging networks [ESR/EIBALL (https://www.myesr.org/research/europeanimaging-biomarkers-alliance-eiball), Quantitative Imaging Biomarkers Alliance (QIBA) (https://www.rsna.org/research/quantitative-imaging-biomarkers-alliance)], composed of numerous sites around the world, have been created to define optimized protocols, collect and share big data [9–11]. Moreover, MRI has seen significant technological advances in both hardware and software [9,12], such as the introduction of digital transceivers in place of analog ones, new acquisition pulse sequences, and methods for image reconstruction [12,13] to accelerate data collection of MRI exams and improve quality, both in the clinical and research settings. It is to note, however, that these developments have contributed to promote a considerable heterogeneity among scanners, beyond the inter-vendor differences already present.

In the context of multisite projects for the collection of "Big data", the process of quality control (QC) becomes fundamental to assess the homogeneity and reproducibility of defined measures across all the scanners belonging to the same network [2,14]. Phantom-based QC methods allow the evaluation of the performance of different MRI scanners by comparing quantitative measures estimated from acquired images of ad-hoc phantoms with standardly defined tolerance ranges [9,15–19]. It is therefore possible to establish the accuracy and stability over time of the measures of interest in order to share data among sites of the network.

One of the most used phantoms for QC is the American College of Radiology (ACR) one, which has been designed and developed as part of the American College of Radiology accreditation program [18,20] for the assessment of image quality through geometric, contrast, and image intensity uniformity measurements. This phantom has been used in several multisite studies, which have demonstrated its effectiveness in assessing and monitoring image quality over time [21–23]. Previous multisite studies have reported that, generally, imaging metrics measured quantitatively fell within the ACR acceptance levels [21–24] with small to moderate intra- and inter-scanner variations, as gauged with the coefficient of variation, from 1 % up to 10 % depending on the metric [23,24]. However, it is important to report that software or hardware updates performed during an ACR longitudinal assessment can result in higher intra-scanner coefficients of variation in signal-to-noise ratio (SNR) (i.e. upgrade should reduce the intra-scanner coefficient of variation), underlying the importance of analyzing image quality over time [23].

The *RIN - Neuroimaging Network* (https://www.reteneuroscienze.it/en/progetti/neuroimaging/) is a national consortium that was born with the support of the Italian Ministry of Health to develop guidelines for the acquisition, processing, collecting, and sharing of large-scale multimodal qMRI data on healthy participants, patients, and animal models. The *RIN - Neuroimaging Network* is currently composed of twenty-three IRCCS Institutes distributed throughout Italy, which have been included based on their equipment and expertise. These can essentially be summarized as having a high field scanner (3T) or an ultra-high-field scanner (7T) for clinical or preclinical research, or known expertise in MRI data analysis and infrastructure management. Since the overarching goal of this network is to harmonize MRI protocols for translating them to the clinical setting, multiple tasks, exploiting both the technical and clinical skills of each site, have been proposed [25]. To achieve this challenging goal, the *RIN - Neuroimaging Network* also leverages high-level IT infrastructures such as NeuGRID [26], ARIANNA [27], and MIP (Medical Informatic Platform) [6].

One of the main goals of the *RIN - Neuroimaging Network* is to develop and share common MRI acquisition protocols and processing pipelines in the vision of future multisite clinical studies. In this context, the present work aimed to: 1) implement a QC routine across sites; 2) implement and optimize the acquisition protocol at each site, also creating an automatic analysis pipeline; 3) evaluate the accuracy and reproducibility of geometrical, contrast and uniformity measures across sites and vendors; 4) evaluate the stability over time of the obtained measures; 5) quantify the impact of a scanner upgrade on measures derived from the same phantom images following significant

technological advances.

Materials and methods

IRCCS sites

The *RIN - Neuroimaging Network* is composed of twenty-three IRCCS institutes distributed across the entire Italian territory. Eighteen of these IRCCS were suitable to be included in this work because they have a 3T scanner from different vendors, i.e. vendor 1 (V1, 9 scanners), vendor 2 (V2, 7 scanners), and vendor 3 (V3, 3 scanners). *Site15* used 2 scanners, one V1 and one V2. Further details of the scanner, software, and head-coil used at each site are reported in Table 1.

Phantom and acquisition protocol

The ACR phantom was used since it is a short cylinder of acrylic plastic filled with a solution of nickel chloride and sodium chloride, containing several structures specially designed to facilitate a variety of tests of scanner performance. Two sizes are available and can be chosen according to the receiver head-coil dimensions at each site: the ACR-large or ACR-small [18,20].

ACR phantom images started to be acquired and analyzed monthly from February 2018, at all sites. Starting from the standardized ACR protocol [18,20] that includes localizer, T1w, and T2w images, the acquisition protocol was adapted and implemented at three sites (*Site3* for V1, *Site10* for V2, and *Site17* for V3), one for each manufacturer, and distributed to the remaining sites. The protocol was modified as reported in the manual (https://zenodo.org/record/6320896), where particular attention has been paid to phantom setup and positioning. Briefly, for all sequences, post-processing filters were disabled and receiver bandwidth was set at 250 ± 20 Hz/px. Due to differences in scanner features, it was not possible to fix the echo time (TE) for the T2w images, thus TE values varied between 79 ms and 81 ms between sites. A further change to the suggested ACR protocol was implemented for sites with the ACR-small, where the acquisition matrix was modified (192 × 192 instead of 192 × 152) to obtain an in-plane isotropic voxel with a resolution of 0.625 mm (Supplementary Table 1).

ACR analysis

An automatic quality assurance pipeline was implemented by adapting the open-source automatic quality assurance (OSAQA) Matlab scripts available at http://jidisun.wix.com/osaqa-project [28].

First, all OSAQA Matlab scripts were converted and rewritten for the analysis of the ACR-small images, as the OSAQA scripts were specifically developed for analyzing the ACR-large images (i.e. all geometrical parameters had to be updated). Then, a change was implemented to allow the appropriate range of TE values (79 and 81 ms) for T2w images. Finally, the image thresholding routine implemented in the OSAQA scripts to assess geometrical accuracy was modified to avoid that background noise at the edges of the field of view compromised the results.

Five quality assurance tests from the ACR standard protocol were chosen to be performed both on T1w and T2w images, and standard ACR tolerance ranges were considered for comparison [18,20]: (1) geometric accuracy, which assesses the accuracy of the length measurement of several structures of the phantom [diameter in x-axis, right-left (RL); in y-axis, anterior-posterior (AP); diagonal with negative gradient (NG); diagonal with positive gradient (PG)]; (2) slice thickness, which evaluates the accuracy of slice selection; (3) high-contrast spatial resolution, which evaluates whether close individual bright dots can be resolved, (4) image intensity uniformity, and (5) ghost quantification test, which are both assessing image quality properties important for qMRI metrics.

A further test was implemented to evaluate geometrical distortions along the slice encoding direction. In detail, the test calculates the ratio between the area of the phantom in the slices used for geometric testing (first and third slices for ACR-small, first and fifth for ACR-large), approximating the phantom section to an ellipse, i.e. $area = \pi \cdot axis_1 \cdot axis_2$. This test was considered passed if the ellipse ratio was unitary, with a tolerance range set to ± 0.02. To identify the extreme acceptable values for ACR-large, the maximum acceptable length (192

Table 1
Description of scanner and equipment for each site of RIN – Neuroimaging Network. The symbol "→" year of installation" column indicates the scanners whose hardware was updated since the beginning of the project. Abbreviations: IRCCS = Scientific Institutes of Hospitalization and Care; T = tesla; ch = channels. (*) This scanner is a PET-MRI. (**) 25 ACR small scans were acquired at 1.5 T.

Site	Field (T)	Vendor	Transmission chain	Year of installation	Head coil	ACR type	# acquisitions (neuGRID)	# included scans
Site1	3T	V1	digital	2009	32 ch	Large	38	32
Site2	3T	V1	digital	2009	32 ch	Large	17	15
Site3	3T	V1	upgrade analog/digital	2009 → 2020	32 ch	Large	32	32
						Small	29	29
Site4	3T	V1	upgrade analog/digital	2011 → 2020	32 ch	Large	40	35
Site5	3T	V1	upgrade analog/digital	2009 → 2018	32 ch	Large	43	32
Site6	3T	V1	digital	2017	32 ch	Large	15	12
Site7	3T	V1	digital	2016	32 ch	Large	1	0
	3T					Small	3	0
Site8	3T	V1	digital	2016	32 ch	Large	47	35
Site9	3T	V2	analog	2009	8 ch	Large	10	8
Site10	3T	V2	analog	2012	32 ch	Small	40	37
Site11	3T	V2	digital	2011	32 ch	Small	27	17
Site12	3T	V2	digital	2017	20 ch	Large	36	28
	3T				64 ch	Small	35	34
Site13	3T	V2	digital	2018	64 ch	Small	31	31
Site14	3T	V2	digital	2018	20 ch	Large	9	0
					64 ch	Small	1	0
Site15	3T	V2	analog (*)	2011	12 ch	Large	41	23
	3T	V1	digital	2009	32 ch	Large	10	10
Site16	3T	V3	analog	2009	8 ch	Large	36	26
Site17	3T	V3	digital	2010	16 ch	Small	19	17
Site18	3T	V3	digital	2009	32 ch	Large	36	24
Site19		–	–		–	–	–	
Site20		acquiring 3T clinical scanner (**)						
Site21		–	–		–	–	–	
Site22		acquiring 3T clinical scanner						
Site23		acquiring 3T clinical scanner						

mm) for slice 1 and the minimum acceptable length (188 mm) for slice 5 were considered (and vice versa), determining a ratio of 1.02 (or 0.98). The same procedure has been applied for ACR-small.

For each site, all these QC measures were assessed at each time point to identify scanners with and without stable longitudinal image quality: scanners were considered longitudinally not stable if the same test failed at least at two consecutive time points. Only sites that acquired at least 8-time points were included in the longitudinal analysis presented in this work.

To create an automated and standardized pipeline for the analysis of phantom ACR acquisitions for all sites, the advanced services of the neuGRID platform (https://www.neugrid2.eu) were used. The platform provides a web-based interface and centralized database for the *RIN - Neuroimaging Network*. Once the data from each IRCCS hospital is transferred to the platform, a QC pipeline is automatically triggered to run scripts described previously and all metrics are automatically computed for each MRI scan. QC results are exported to a spreadsheet and archived along with DICOM images to ensure long-term performance reporting. An email summarizing the results of each analysis with tolerance ranges to evaluate acquisition performance is automatically sent to the uploader's mailbox. A detailed manual (https://zenodo.org/record/6320896) has been distributed to all users of the *RIN - Neuroimaging Network*. It contained detailed instructions on how to acquire and upload data to the centralized e-infrastructure for data exchange.

Statistical analysis

Statistical analysis was performed using R (https://www.R-project.org/) [29].

Mean and standard deviation over time were obtained at each site and for each vendor for geometric accuracy, slice thickness, ellipse ratio, image intensity uniformity, and ghosting measure. For the high-contrast spatial resolution test, the obtained data were qualitative ordinal data, so the mode over time was reported (i.e. ACR-large: 0.8, 0.9, 1 mm; ACR-small: 0.7, 0.8, 0.9 mm).

To assess the reproducibility between scanners of different vendors (i.e. inter-site), longitudinal scans within a site (i.e. intra-site), and before and after a scanner upgrade, absolute error (AE) was determined as a measure of deviation from the expected value for each measurement of each QC test, except high-contrast spatial resolution. The expected values for the standard ACR measurements were the real geometrical dimension for the lengths (ACR-large = 190 mm, ACR-small = 100 mm), the ratio between the elliptical areas equal to 1, and the slice thickness equal to 5 mm, whereas the ideal values for image intensity uniformity and ghosting were 100 % and 0, respectively. The AE index was expressed as the difference in absolute value between the mean value for each site ($\bar{x}_{\text{site}}$) and the expected values (x_{expected}). AE was reported as a percentage value, i.e. the percent absolute error (%AE):

$$\%AE = \left|\bar{x}_{site} - -x_{expected}\right| * 100 \,/ x_{expected}$$

For each measure, the maximum acceptable %AE was calculated using the tolerance range defined by the ACR recommendations [18,20] in the formula above

Results

Data were acquired on either or both the ACR phantoms from 18 sites and 19 scanners (*Site15* used 2 scanners, one of V1 and one of V2). However, two sites were excluded from analyses: *Site7* was excluded because it acquired <8-time points, while *Site14* was excluded for wrong acquisitions, due to an error in phantom positioning (Table 1).

In total, 16 sites (17 scanners) of the 18 that performed the QC were included in the statistical analysis: 11 scanners acquired images on the ACR- large, 4 scanners acquired images on the ACR- small, and 2 scanners acquired images on both phantoms (Table 1). A total of 477 scans (312 ACR-large, 165 ACR-small), distributed between vendors, were included in the statistical evaluation: 65 % V1, 19 % V2, 16 % V3 for ACR-large, and 18 % V1, 72 % V2, 10 % V3 for ACR-small. The presence of air bubbles and/or the setting of incorrect sequence parameters and/or incorrect positioning of the phantom led to the exclusion of 19 % of ACR acquisitions from the analysis.

Intra-scanner metrics evaluation

Tables 2 and 3, Supplementary Tables 2 and 3 reports mean value with relative standard deviation and %AE of all measurements obtained in the ACR-large and ACR-small for each site and grouped by vendor, for both T1w and T2w images.

13 (81.3 %) of the 16 analyzed sites passed all tests on all metrics calculated using both T1w and T2w scans while 3 sites failed some tests showing lower values than those expected. *Site3* failed the test of ellipse ratio (0.98 instead of 1) on ACR-large, *Site12* failed tests of ellipse ratio (0.98 instead of 1) and image intensity uniformity (80.88 % instead of 82 %) on ACR-large and slice thickness (4.06 mm instead of 4.30 mm) on ACR-small, and *Site16* failed image intensity uniformity (79.24 % instead of 82 %) test in ACR-large.

Intra-scanner longitudinal assessment

Longitudinal evaluation of scanner stability was described with %AE (Tables 2 and 3, Supplementary Tables 2 and 3) for both T1w and T2w images. T1w and T2w results were similar. As an example, longitudinal measures obtained from T1w images of ACR-large are shown in Supplementary Figs. (1 to 6).

As reported in the previous section for intra-scanner ACR measures, 13 of the 16 analyzed sites showed that all measures were in the tolerance range over time regardless of the used ACR phantom or image weighting, demonstrating good longitudinal reproducibility. In these sites, %AE was: ≤1 % for in-plane measurements of phantom lengths (i.e. RL, AP, NG, PG) (Supplementary Figs. 1 and 2), ≤1.78 % for ellipse ratio (Supplementary Figure 3), <4 % for slice thickness, <16.5 % for image intensity uniformity (Supplementary Figure 5), and <0.05 for ghosting (Supplementary Figure 6). Only the 3 sites that failed tests at a few time points presented a higher %AE than those expected. In particular, *Site3* and *Site12* showed a value close to the lower limit of the acceptable tolerance range for ellipse ratio test using ACR-large (Table 2, Supplementary Figure 3), while they passed the test using ACR-small.

High-contrast spatial resolution results showed that all scanners over time were able to distinguish two spots at 0.9 mm of distance in ACR-large and at 0.7 mm of distance in ACR-small, with one failing test at maximum per site (Table 3). Only *Site16*, *Site17*, and *Site18* showed performance worse than 0.9 mm in more than 5 random time points.

Inter-vendor comparisons

For the ACR-large, the ellipse ratio, which gives us an assessment of distortions along the slice encoding direction, showed inter-vendor variability (Fig. 1A, Table 2). In particular, V3 and V2 scanners showed comparable average ellipse ratios over time but with values close to 1, or ranging between 0.975 and 1 (with some lower outliers), respectively. Instead, V1 scanners showed greater variability with average ellipse ratios varied between 0.960 and 1.012.

Our results for slice thickness measures showed high intra- and inter-vendor reproducibility in both T1w and T2w acquisitions, as expressed by the small %AE that were comparable between vendors (T1w: 1.4 % V1, 2.7 % V2, 0.8 % V3; T2w: 1.0 % V1, 2.0 % V2, 3.8 % V3).

Image intensity uniformity was higher in V1 scanners demonstrating high values with the smallest %AE (T1w: 10.27 %; T2w: 10.55 %), averaging over time and inter-scanner, while the greatest variability was detected in V3 scanners that presented lower image intensity uniformity

Table 2
Mean value with relative standard deviation in brackets and %AE of the in-plane length in two slices, ratio between ellipses, and slice thickness calculated in ACR-large and ACR-small for each site and grouped by vendor for T1w images. Abbreviations: %AE = percent absolute error; RL = right-left; AP = anterior-posterior; NG = negative gradient; PG = positive gradient; Th = threshold.

Site	Vendor	ACR type	T1-weighted											
			RL-slice1 Length (mm)	AP-slice1 Length (mm)	RL-slice5 Length (mm)	AE (%) Th = 2 %	AP-slice5 Length (mm)	AE (%) Th = 2 %	NG-slice5 Length (mm)	PG-slice5 Length (mm)	Ellipse ratio	AE (%) Th = 2 %	Slice thickness (mm)	AE (%) Th = 14 %
Site1	V1	Large	188.16 (0.51)	188.41 (0.61)	189.78 (0.42)	0.12	189.63 (0.49)	0.20	190.53 (0.57)	190.28 (0.63)	0.99 (0.01)	1.49	5.01 (0.14)	0,16
Site2	V1	Large	188.93 (1.22)	189.13 (1.06)	189.87 (0.35)	0.07	190 (0.38)	0.00	190.6 (0.63)	190.33 (0.62)	0.99 (0.01)	0.94	5.34 (0.77)	6,81
Site3	V1	Large	187.84 (1.97)	188.09 (1.96)	189.91 (0.73)	0.05	189.88 (0.61)	0.07	190.59 (0.71)	190.44 (0.67)	**0.98 (0.02)**	**2.01**	5.01 (0.08)	0,23
Site4	V1	Large	188.6 (1.14)	188.97 (1.04)	189.71 (0.86)	0.15	189.77 (0.84)	0.12	190.46 (0.78)	190.14 (0.88)	0.99 (0.01)	1.00	5.17 (0.22)	3,34
Site5	V1	Large	189.25 (0.76)	189.16 (1.42)	190.28 (0.73)	0.15	190.19 (0.59)	0.10	190.84 (0.63)	191.25 (0.67)	0.99 (0.01)	1.08	5.03 (0.38)	0,56
Site6	V1	Large	189.57 (0.52)	189.33 (0.58)	189.73 (0.53)	0.14	190.23 (0.7)	0.12	190.8 (0.89)	191.03 (0.54)	0.99 (0)	0.55	5.01 (0.18)	0,12
Site8	V1	Large	188.49 (0.51)	188.86 (0.43)	188.86 (0.55)	0.60	188.51 (0.78)	0.78	189.31 (0.83)	189.26 (0.82)	1 (0.01)	0.01	5.05 (0.15)	1,07
Site15	V1	Large	190 (0)	190.3 (0.48)	190.4 (0.52)	0.21	190.1 (0.32)	0.05	191 (0)	191.5 (0.53)	1 (0)	0.10	4.96 (0.24)	0,88
Site9	V2	Large	188.88 (0.83)	189 (1.6)	190.5 (0.93)	0.15	190 (0)	0.00	191.25 (0.71)	190.13 (0.35)	0.99 (0.01)	1.38	5.04 (0.32)	0,78
Site12	V2	Large	188.73 (2.51)	188.38 (2.1)	190.46 (0.71)	0.29	190.69 (0.88)	0.34	191.73 (0.83)	191.73 (0.6)	**0.98 (0.02)**	**2.11**	4.77 (0.2)	4,69
Site15	V2	Large	190.17 (1.44)	189.7 (1.69)	191.35 (1.03)	0.71	191.96 (0.56)	1.03	192.7 (0.82)	191.3 (0.76)	0.98 (0.01)	1.78	4.89 (0.25)	2,11
Site16	V3	Large	190.19 (0.49)	190.35 (0.85)	190.35 (1.09)	0.18	189.96 (0.53)	0.02	190.5 (0.51)	190.96 (0.77)	1 (0.01)	0.12	4.79 (1.06)	4,25
Site18	V3	Large	188.67 (0.48)	189.58 (0.88)	189.71 (0.91)	0.15	188.67 (0.48)	0.70	189.83 (0.38)	190.38 (0.65)	1 (0.01)	0.06	5.15 (0.23)	2,98
All sites - V1		Large	188.85 (0.83)	189.03 (0.95)	189.82 (0.59)	0.14	189.79 (0.59)	0.17	190.52 (0.63)	190.53 (0.67)	0.99 (0.01)	1.01	5.07 (0.27)	1.38
All sites - V2		Large	189.26 (1.59)	189.03 (1.8)	190.77 (0.89)	0.46	190.88 (0.48)	0.61	191.89 (0.79)	191.05 (0.57)	0.98 (0.01)	1.87	4.9 (0.26)	2.70
All sites - V3		Large	189.43 (0.49)	189.96 (0.86)	190.03 (1)	0.02	189.31 (0.5)	0.35	190.17 (0.45)	190.67 (0.71)	1 (0.01)	0.03	4.97 (0.64)	0.78
			RL-slice1 Length (mm)	*AP-slice1 Length (mm)*	*RL-slice3 Length (mm)*	*AE (%) Th = 2 %*	*AP-slice3 Length (mm)*	*AE (%) Th = 2 %*	*NG-slice3 Length (mm)*	*PG-slice3 Length (mm)*	*Ellipse ratio*	*AE (%) Th = 2 %*	*Slice thickness (mm)*	*AE (%) Th = 14 %*
Site3	V1	Small	98.85 (0.86)	98.87 (1.03)	99.31 (0.58)	0.69	99.39 (0.59)	0.60	99.8 (0.58)	99.76 (0.56)	0.99 (0.01)	0.97	4.89 (0.18)	2,14
Site10	V2	Small	100.74 (0.43)	100.50 (0.62)	100.99 (0.45)	1.00	100.65 (0.53)	0.66	101.31 (0.35)	101.79 (0.42)	0.99 (<0.01)	0.40	5.01 (0.08)	0,24
Site11	V2	Small	100.47 (0.27)	100.33 (0.44)	100.66 (0.26)	0.66	100.55 (0.65)	0.55	101.02 (0.37)	101.54 (0.32)	0.99 (<0.01)	0.40	5.16 (0.18)	3,29
Site12	V2	Small	99.63 (0.34)	99.87 (0.52)	99.85 (0.34)	0.15	100.01 (0.47)	0.02	100.36 (0.31)	100.47 (0.34)	0.99 (<0.01)	0.37	**4.06 (0.57)**	**18,79**
Site13	V2	Small	100.30 (0.31)	100.00 (0.66)	100.34 (0.31)	0.34	99.97 (0.57)	0.02	100.7 (0.26)	100.64 (0.25)	0.99 (<0.01)	0.02	5.04 (0.15)	0,86
Site17	V3	Small	99.33 (0.15)	100.18 (0.41)	99.44 (0.18)	0.56	100.18 (0.27)	0.19	99.97 (0.28)	100.47 (0.24)	0.99 (<0.01)	0.10	5.11 (0.15)	2,24
All sites - V2		Small	100 (0.12)	100.14 (0.11)	100.16 (0.11)	0.45	100.2 (0.13)	0.28	100.59 (0.04)	100.84 (0.08)	0.99 (<0.01)	0.29	4.81 (0.22)	4.60

(around 81 %), failing the inferior limit of 82 %, with the highest %AE (T1w: 18.71 %; T2w: 19.28 %) (Fig. 1B, Table 3). *Site15* is equipped with an MRI-PET scanner that resulted in the V2 scanner with the highest image intensity uniformity (around 93 %).

Ghosting measures showed high intra- and inter-scanner reproducibility whatever the vendor.

Pre- and post-upgrade comparisons

Hardware upgrades impacted the amount of geometrical distortions along the slice encoding direction as demonstrated in the 3 V1 scanners that upgraded the signal transmission chain from analog to digital. Indeed, %AE of ellipse ratio is reduced in all sites ranging from 0.1 % (*Site5*) to 1.5 % (*Site3*) after the upgrade (Fig. 2, Table 4).

Discussion

Our work successfully implemented a harmonized QC procedure for MRI geometric distortions monitoring across multiple sites, with a centralized automated data analysis and reporting structure. This allowed the network to perform a seamless collection and analysis of phantom data and demonstrated the value of this activity to monitor MRI data quality across research hospitals. Overall, our data showed that more than 80 % of the 16 sites that took part in this study, regardless

Table 3
Mean value with relative standard deviation in brackets and %AE of the image intensity uniformity and ghosting measure and mode of the high-contrast spatial resolution calculated in ACR-large and ACR-small for each site and grouped by vendor for both T1w. Abbreviations: %AE = percent absolute error; LR = lower-right; UL = upper-left; Th = threshold.

Site	Vendor	ACR type	T1-weighted					
			High-contrast UL (mm)	High-contrast LR (mm)	Uniformity (%)	AE (%) Th = 18 %	Ghosting	AE (%) Th = 2.5 %
Site1	V1	Large	0.90	0.90	89.47 (1.4)	10.53	0.000	0.05
Site2	V1	Large	0.90	0.90	90.04 (1.23)	9.96	0.000	0.04
Site3	V1	Large	0.90	0.90	90.2 (1.4)	9.80	0.000	0.03
Site4	V1	Large	0.90	0.90	87.19 (2.9)	12.81	0.000	0.04
Site5	V1	Large	0.90	0.90	88.57 (1.29)	11.43	0.000	0.04
Site6	V1	Large	0.90	0.90	90.37 (1.78)	9.63	0.001	0.05
Site8	V1	Large	0.90	0.90	92.23 (1.57)	7.77	0.000	0.03
Site15	V1	Large	0.90	0.90	91.74 (0.99)	8.26	0.000	0.04
Site9	V2	Large	0.90	0.90	83.93 (5.47)	16.07	0.002	0.18
Site12	V2	Large	0.90	0.90	**80.88 (0.79)**	**19.12**	0.000	0.03
Site15	V2	Large	0.90	0.90	93.04 (0.66)	6.96	0.001	0.06
Site16	V3	Large	0.90	0.90	**79.24 (6.00)**	**20.76**	0.001	0.14
Site18	V3	Large	0.90	0.90	83.51 (1.44)	16.49	0.001	0.07
All sites - V1		Large	0.90	0.90	89.97 (1.57)	10.27	<0.001	0.04
All sites - V2		Large	0.90	0.90	85.95 (2.31)	13.19	0.001	0.07
All sites - V3		Large	0.90	0.90	**81.38 (3.72)**	**18.71**	0.001	0.11
Site3	V1	Small	0.70	0.70	93.54 (1.49)	6.45	<0.001	0.15
Site10	V2	Small	0.70	0.70	96.84 (0.55)	3.16	<0.001	0.10
Site11	V2	Small	0.70	0.70	95.17 (1.23)	4.82	<0.001	0.15
Site12	V2	Small	0.70	0.70	92.82 (1.46)	7.18	<0.001	0.18
Site13	V2	Small	0.70	0.70	96.07 (0.37)	3.92	<0.001	0.11
Site17	V3	Small	0.70	0.70	93.7 (0.65)	6.30	<0.001	0.45
All sites - V2		Small	0.70	0.70	94.86 (0.48)	4.74	<0.001	0.13

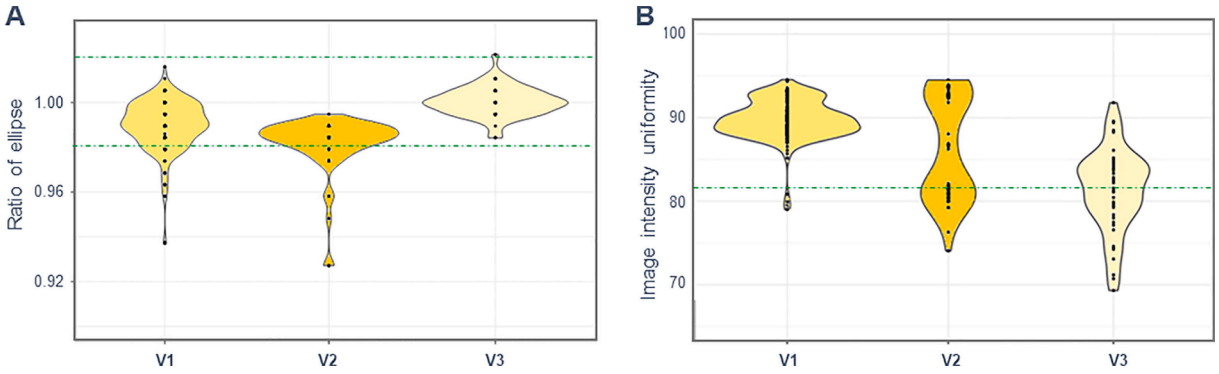

Fig. 1. ACR-large: all longitudinal data for each site were grouped according to vendor. On the left, violin plots represent inter-vendor variations of the ratio between the ellipses in T1-weighted images, which gives us an assessment of distortions along the slice encoding direction (panel A). On the right, violin plots show inter-vendor variations of uniformity measure (panel B) in T1-weighted images. Green lines represent limits of ACR tolerance ranges. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of vendor, coil, and signal transmission chain type, passed all QC tests. The remaining sites failed one or two tests over the 48 months of data collection, which in some cases required the scanner manufacturer's intervention to be rectified. The geometrical measures estimated through T1- and T2- weighted images were similar, according to the expected performance [24]. Image intensity uniformity showed an intra-vendor difference; indeed V3 scanners demonstrated a higher %AE than V1 scanners. On the other hand, V3 scanners showed the smallest intra- and inter-scanner variability of geometric measurements along the slice encoding direction. Scanners that upgraded the signal transmission chain reported a lower level of geometric distortions along the slice encoding direction measured with the ellipse ratio in the post-upgrade compared to pre-upgrade data.

Initially, a benchmarking of possible commercially available phantoms [9,17] for assessing scanner performance and their geometric accuracy was performed. The phantoms that were examined were the ACR phantom [18,20], the International Society for Magnetic Resonance in Medicine/National Institute of Standards and Technology (ISMRM/NIST) MRI system phantom [30], and the ADNI (Alzheimer's Disease Neuroimaging Initiative) phantom [16]. The ACR phantom was selected for its potential, i.e. for the comprehensive set of geometrical tests that can be performed with it, and for what was perceived to be a better cost-benefit ratio. However, it is to note that ACR is suitable for evaluating geometrical accuracy and gradient linearity assessment but it cannot be used for advanced micro-structural and functional evaluations.

During the protocol set-up and acquisition phase, some critical issues were detected that should be always taken into account, demonstrating that qMRI and QC protocols must be carefully acquired by a trained expert operator in order to provide stable and reproducible results. First, for a correct geometric estimation, it is crucial that the ACR phantom does not present any air bubbles larger than 1–1.5 cm, so distilled water should be periodically refilled as per instructions - this is not always followed. An equally important factor in determining image distortions is setting the receiver bandwidth and the reconstruction matrix as required so that the analysis software works accordingly [24]. To increase the intra- and inter-site reproducibility, an optimal range for the

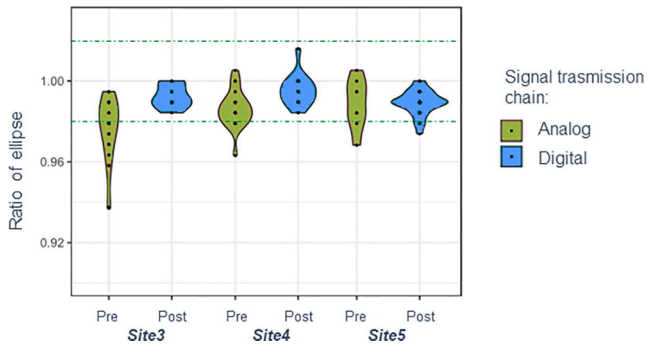

Fig. 2. ACR-large: the violin plots show the ratio between the ellipses in the 3 scanners which have upgraded the signal transmission chain from analog to digital. All longitudinal data of the 3 scanners were included and grouped according to pre- and post-upgrade. Abbreviations: pre = pre-upgrade; post = post-upgrade. Green lines represent limits of ACR tolerance ranges. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

radio frequency acquisition bandwidth parameter was identified to ensure consistent image quality while limiting distortions, e.g. a lower bandwidth increases image SNR but also geometric distortions [21,31]. In this framework, it is to note that these parameters must be manually set for every acquisition on GE scanners - i.e. the scanner does not store such changes from default values that are possible only using the "control variable" panel, which is available under a GE research agreement.

A further key issue is the positioning of the phantom inside the scanner, which must be performed by an expert operator, properly trained on the procedure. Indeed, 19 % of the ACR acquisitions were excluded from the analyses due to the presence of air bubbles and/or incorrect sequence parameters and/or inaccurate positioning of the phantom. Interestingly, these acquisitions were from earlier time points when acquisition compliance was low. The accuracy of the scans improved considerably by a central reporting of the issues and by the distribution of a detailed manual, updated during the first few trials as issues became apparent.

In this multisite study, we introduced a new measure of QC, the ellipse ratio between slices 1 and 5 for the ACR-large and slices 1 and 3 for the ACR-small, which was not included in the ACR recommendations. This measure allowed us to obtain an estimate of the geometric distortions along the slice encoding direction, in addition to the in-plane geometric distortions obtained by calculating the length of the ACR phantom diameter in slice 1 and slice 5 (i.e, for the ACR-large). By doing so, we were able to determine which sites acquired images with a significant deformation along the slice encoding direction and to take appropriate corrective measures, such as tuning the scanner gradients requiring the intervention of the manufacturer. In addition, our results showed that hardware upgrade impacted the amount of geometric distortions along the slice encoding direction, with a decrease of the %AE up to 1.5 %. This is consistent with evidence from a previous study, whereby variability in geometric accuracy measures decreased following hardware/software upgrades [23]. In-plane geometric accuracy measurements show %AEs of <1 %, consistent with previous studies, which identified similar variations on 1.5T and 3T scanners [22,24].

The evaluation of possible B0 inhomogeneities is fundamental for qMRI. In terms of contrast parameters, our results showed that the ghosting factor was orders of magnitude below the tolerance limit defined by ACR recommendations for all scanners, according to previous studies [22,24], whereas image intensity uniformity was more susceptible to intra- and inter-vendor variations, in addition to magnetic field strength (1.5T vs 3T) [22–24]. In particular, scanner age was affecting results, where the oldest scanner of the network exhibited the greatest intra- and inter-site variability, underlining the important progress introduced both in hardware and software over years. It is to note, however, that our results regarding inter-vendor variability should be considered only as a qualitative comparison, as the distribution of scanner's age and software versions was not comparable between vendors.

More than 95 % of the test values obtained with the ACR-small were in the tolerance range for all time points except for the measurement of slice thickness for a single site. From data acquired in this study, it emerges that the ACR-large is more sensitive to geometrical alterations and hence is more reliable as a QC phantom than the ACR-small. For example, *Site3* performed monitoring over time on both ACR phantoms with the same coil, demonstrating that QC evaluations performed using the ACR-large picked up possible alterations in geometric measurements hence becoming instrumental to implementing corrective actions, while QC tests on the ACR-small were insensitive to subtle distortions. Therefore, we recommend that sites with head coils that allow acquisitions with the ACR-large should use this phantom instead of the ACR-small. Moreover, the ACR-large occupies a more similar space to a subject's head inside the coil.

We, therefore, confirm, like previous studies, that the QC measures most prone to failure are image intensity uniformity and slice thickness [21–24], and we suggest that the ellipse ratio should be added as an additional measure to identify the need for gradient calibration.

Some considerations should be noted. The scanners participating in our study used head coils with different number of channels, whose source of variation was not considered in the intra- and inter-vendor evaluation. Moreover, it was not possible to perform with the chosen ACR phantoms a quantification of the variations of T1 and T2

Table 4
Pre- and post-upgrade QC evaluations on T1-weighted images. %AE of the geometrical distortions along the slice encoding direction, slice thickness, image intensity uniformity, and ghosting measures calculated in ACR-large and ACR-small for the 3 V1 scanners which have upgraded the signal transmission chain from analog to digital. Abbreviations: %AE = percent absolute error; Th = threshold.

Weight	*Code*	*Vendor*	*ACR type*	*pre-upgrade*				*post-upgrade*			
				Ellipse ratio AE (%) Th = 2 %	*Slice thickness AE (%) Th = 14 %*	*Uniformity AE (%) Th = 18 %*	*Ghosting AE (%) Th = 2.5 %*	*Ellipse ratio AE (%) Th = 2 %*	*Slice thickness AE (%) Th = 14 %*	*Uniformity AE (%) Th = 18 %*	*Ghosting AE (%) Th = 2.5 %*
T1	*Site3*	V1	Large	**2.36**	0.54	9.58	0.02	0.75	0.86	9.48	0.04
	Site4	V1	Large	1.37	3.93	12.47	0.05	0.45	2.44	11.32	0.02
	Site5	V1	Large	1.12	0.27	11.51	0.03	1.07	0.83	11.64	0.05
	All sites	- V1	Large	1.79	1.74	10.99	0.03	0.83	1.07	11.88	0.04
	Site3	V1	Small	1.08	2.24	6.85	0.15	0.31	1.55	4.01	0.16
T2	*Site3*	V1	Large	**2.28**	0.78	10.59	0.04	0.75	1.29	11.00	0.02
	Site4	V1	Large	1.30	0.10	13.33	0.10	0.22	2.61	11.58	0.08
	Site5	V1	Large	1.32	0.50	11.41	0.05	1.05	4.30	11.16	0.04
	All sites	- V1	Large	1.76	0.33	10.52	0.06	0.74	1.68	11.27	0.05
	Site3	V1	Small	1.18	3.06	6.81	0.11	0.16	3.95	4.25	0.05

measurements, which need more expensive dedicated phantoms. Finally, the automatic ACR software does not include the computation of SNR, which could be introduced in a further version.

Conclusion

Optimized ACR phantom protocols can be used to identify sites performing below acceptable image quality, take corrective actions, and determine intra- and inter-site accuracy levels by monitoring and correcting instrumental variations relevant to qMRI analysis. In addition, centralized automated analysis of collected data, with reporting, allowed each site to independently and timely monitor scanner performance over time, reducing operator-related variability. Technical expertise available at the *RIN - Neuroimaging Network* supported each site in understanding the source of specific problems and improved the quality of its protocols.

The *RIN - Neuroimaging Network* is the first large neuroimaging network in Italy and the present work demonstrates that good synergy among sites can improve the scientific impact of each participating site. The *RIN - Neuroimaging Network* is composed of many vendors, different hardware and software releases, and different types of coils. This cooperation has allowed us to harmonize the ACR MRI protocols on 3T scanners and to enhance the original ACR scanning procedure by reducing potential software and hardware confounding factors that would impact *in vivo* qMRI.

Funding

This project was funded by the Italian Minister of Health (RRC; RRC-2016-2361095; RRC-2017-2364915; RRC-2018-2365796; RCR-2019-23669119_001; RCR 2020-23670067; RCR 2022-23682285) and the Ministry of Economy and Finance (CCR-2017-23669078). The funding sources had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, and approval of the manuscript.

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.

Acknowledgments

This is a short text to acknowledge the contributions of specific colleagues and institutions that aided the efforts of the authors. We would also like to thank Italian Minister of Health and Ministry of Economy and Finance for the funding to support the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2022.10.008.

References

[1] Van Horn JD, Toga AW. Human neuroimaging as a 'Big Data' science. Brain Imaging Behav 2014;8:323–31. https://doi.org/10.1007/s11682-013-9255-y.
[2] Madan CR. Scan once, analyse many: using large open-access neuroimaging datasets to understand the brain. Neuroinformatics 2021;11:1–29. https://doi.org/10.1007/s12021-021-09519-6.
[3] Calhoun VD, Pearlson GD, Sui J. Data-driven approaches to neuroimaging biomarkers for neurological and psychiatric disorders: emerging approaches and examples. Curr Opin Neurol 2021;34:469–79. https://doi.org/10.1097/WCO.0000000000000967.
[4] Hubbard Cristinacce PL, Keaveney S, Aboagye EO, et al. Clinical translation of quantitative magnetic resonance imaging biomarkers – An overview and gap analysis of current practice. Phys Medica 2022;101:165–82. https://doi.org/10.1016/j.ejmp.2022.08.015.
[5] Archetti D, Ingala S, Venkatraghavan V, Wottschel V, Young AL, Bellio M, et al. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease. NeuroImage Clin 2019;24:101954. https://doi.org/10.1016/j.nicl.2019.101954.
[6] Redolfi A, De Francesco S, Palesi F, et al. Medical informatics platform (MIP): a pilot study across clinical Italian cohorts. *Front Neurol* 2020;11. https://doi.org/10.3389/fneur.2020.01021.
[7] Kwak K, Giovanello KS, Bozoki A, Styner M, Dayan E. Subtyping of mild cognitive impairment using a deep learning model based on brain atrophy patterns. Cell Reports Med 2021;2(12):100467.
[8] Saw SN, Ng KH. Current challenges of implementing artificial intelligence in medical imaging. Phys Medica 2022;100:12–7. https://doi.org/10.1016/j.ejmp.2022.06.003.
[9] Keenan KE, Biller JR, Delfino JG, Boss MA, Does MD, Evelhoch JL, et al. Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs. J Magn Reson Imaging 2019;49(7):e26–39.
[10] Helmer KG, Ambite JL, Ames J, Ananthakrishnan R, Burns G, Chervenak AL, et al. Enabling collaborative research using the Biomedical Informatics Research Network (BIRN). J Am Med Informatics Assoc 2011;18(4):416–22.
[11] Keator DB, Helmer K, Steffener J, Turner JA, Van Erp TGM, Gadde S, et al. Towards structured sharing of raw and derived neuroimaging data across existing resources. Neuroimage 2013;82:647–61.
[12] Keenan KE, Gimbutas Z, Dienstfrey A, Stupic KF. Assessing effects of scanner upgrades for clinical studies. J Magn Reson Imaging 2019;50(6):1948–54.
[13] Morelli JN, Runge VM, Ai F, Attenberger U, Vu L, Schmeets SH, et al. An image-based approach to understanding the physics of MR artifacts. RadioGraphics 2011;31(3):849–66.
[14] Duchesne S, Chouinard I, Potvin O, Fonov VS, Khademi A, Bartha R, et al. The Canadian dementia imaging protocol: harmonizing national cohorts. J Magn Reson Imaging 2019;49(2):456–65.
[15] Firbank MJ, Harrison RM, Williams ED, Coulthard A. Quality assurance for MRI: Practical experience. Br J Radiol 2000;73(868):376–83.
[16] Gunter JL, Bernstein MA, Borowski BJ, Ward CP, Britson PJ, Felmlee JP, et al. Measurement of MRI scanner performance with the ADNI phantom. Med Phys 2009;36(6Part1):2193–205.
[17] Keenan KE, Ainslie M, Barker AJ, Boss MA, Cecil KM, Charles C, et al. Quantitative magnetic resonance imaging phantoms: a review and the need for a system phantom. Magn Reson Med 2018;79(1):48–61.
[18] American College of Radiology. Phantom Test Guidance for Use of the Small MRI Phantom for the ACR MRI Accreditation Program. *Am Coll Radiol* 2018:1–29.
[19] De Deene Y, Wheatley M, Greig T, et al. A multi-modality medical imaging head and neck phantom: Part 1. Design and fabrication. Phys Medica 2022;96:166–78. https://doi.org/10.1016/j.ejmp.2022.02.010.
[20] American College of Radiology. Phantom Test Guidance for the ACR MRI Accreditation Program. *Am Coll Radiol* 2018:1–28. https://www.acraccreditation.org/-/media/ACRAccreditation/Documents/MRI/LargePhantomGuidance.pdf?la=en.
[21] Chen C-C, Wan Y-L, Wai Y-Y, Liu H-L. Quality assurance of clinical MRI scanners using ACR MRI phantom: Preliminary results. J Digit Imaging 2004;17(4):279–84.
[22] Mulkern RV, Forbes P, Dewey K, Osganian S, Clark M, Wong S, et al. Establishment and results of a magnetic resonance quality assurance program for the Pediatric brain Tumor consortium. Acad Radiol 2008;15(9):1099–110.
[23] Prohl AK, Scherrer B, Tomas-Fernandez X, Filip-Dhima R, Kapur K, Velasco-Annis C, et al. Reproducibility of structural and diffusion tensor imaging in the TACERN multi-center study. Front Integr Neurosci 2019;13. https://doi.org/10.3389/fnint.2019.00024.
[24] Ihalainen TM, Lönnroth NT, Peltonen JI, Uusi-Simola JK, Timonen MH, Kuusela LJ, et al. MRI quality assurance using the ACR phantom in a multi-unit imaging center. Acta Oncol (Madr) 2011;50(6):966–72.
[25] Nigri A, Ferraro S, Gandini Wheeler-Kingshott CAM, 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.
[26] Redolfi A, Manset D, Barkhof F, Wahlund L-O, Glatard T, Mangin J-F, et al. Head-to-head comparison of two popular cortical thickness extraction algorithms: a cross-sectional and longitudinal study. PLoS One 2015;10(3):e0117692. https://doi.org/10.1371/journal.pone.0117692.
[27] Retico A, Arezzini S, Bosco P, Calderoni S, Ciampa A, Coscetti S, et al. ARIANNA: a research environment for neuroimaging studies in autism spectrum disorders. Comput Biol Med 2017;87:1–7. https://doi.org/10.1016/j.compbiomed.2017.05.017.
[28] Sun J, Barnes M, Dowling J, Menk F, Stanwell P, Greer PB. An open source automatic quality assurance (OSAQA) tool for the ACR MRI phantom. Australas Phys Eng Sci Med 2015;38(1):39–46. https://doi.org/10.1007/s13246-014-0311-8.
[29] R: A language and environment for statistical computing. R Core Team (2021). R Foundation for Statistical Computing, Vienna, Austria.
[30] Russek S, MB, Jackson E, *et al.* Characterization of NIST/ISMRM MRI system phantom. In: *Proceedings of the 20th Annual Meeting of ISMRM2012, Melbourne, Australia.* 2012.
[31] Huand SY, Seethamraju RT, Patel P, *et al.* Body MR imaging: Artifacts. 2015:1–22.