Reproducibility and variability of quantitative magnetic resonance imaging markers in cerebral small vessel disease.

François De Guio, Eric Jouvent, Geert Jan Biessels, Sandra Black, Carol Brayne, Christopher Chen, Charlotte Cordonnier, Frank-Eric De Leeuw, Martin Dichgans, Fergus Doubal, et al.

Reproducibility and variability of quantitative MRI markers in cerebral small vessel disease

François De Guio[1,2], Eric Jouvent[1,2,3], Geert Jan Biessels[4], Sandra E. Black[5], Carol Brayne[6], Christopher Chen[7], Charlotte Cordonnier[8], Frank-Eric De Leeuw[9], Martin Dichgans[10,11], Fergus Doubal[12], Marco Duering[10], Carole Dufouil[13], Emrah Duzel[14], Franz Fazekas[15], Vladimir Hachinski[16], M. Arfan Ikram[17,18], Jennifer Linn[19], Paul M. Matthews[20], Bernard Mazoyer[21], Vincent Mok[22], Bo Norrving[23], John T. O'Brien[24], Leonardo Pantoni[25], Stefan Ropele[15], Perminder Sachdev[26], Reinhold Schmidt[15], Sudha Seshadri[27], Eric E. Smith[28], Luciano A. Sposato[16], Blossom Stephan[29], Richard H. Swartz[5], Christophe Tzourio[13], Mark van Buchem[30], Aad van der Lugt[17], Robert van Oostenbrugge[31], Meike W. Vernooij[17], Anand Viswanathan[32], David Werring[33], Frank Wollenweber[10], Joanna M. Wardlaw[12,34], Hugues Chabriat[1,2,3]

[1]University Paris Diderot, Sorbonne Paris Cité, UMRS 1161 INSERM, Paris, France; [2]DHU NeuroVasc Sorbonne Paris Cité, Paris, France; [3]AP-HP, Lariboisière Hosp, Department of Neurology, Paris, France; [4]Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands ;[5]Department of Medicine (Neurology), Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada ; [6]Department of Public Health and Primary Care, Cambridge University, Cambridge, UK ; [7]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore ; [8]University of Lille, CHU Lille, Inserm U1171, Lille, France ; [9]Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Department of Neurology, Nijmegen, The Netherlands ; [10]Institute for Stroke and Dementia Research, Klinikum der UniversitätMünchen, Ludwig-Maximilian-University (LMU), Munich, Germany ;[11]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany;[12]Department of Neuroimaging Sciences, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK ;[13]Inserm UMR897, University of Bordeaux, Bordeaux, France ;[14]Department of Cognitive Neurology and Dementia Research, University of Magdeburg, Magdeburg, Germany ; [15]Department of Neurology, Medical University of Graz, Graz, Austria ; [16]Department of Clinical Neurological Sciences, University of Western Ontario, London, Canada ; [17]Department of Radiology and department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands ;[18]Department of Neurology, Erasmus MC, University

Medical Center Rotterdam, Rotterdam, The Netherlands ; [19]Department of Neuroradiology, University Hospital Munich, Munich, Germany ; [20]Division of Brain Sciences, Department of Medicine, Imperial College London, London, UK ; [21]CNRS-CEA UMR5296, University of Bordeaux, Bordeaux, France ; [22]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China ; [23]Department of Clinical Sciences, Neurology, Lund University, Lund, Sweden;[24]Department of Psychiatry, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK ; [25]NEUROFARBA Department, University of Florence, Florence, Italy ; [26]Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, Australia; [27]Department of Neurology, Boston University School of Medicine, Boston, USA ; [28]Department of Clinical Neurosciences and Hotchkiss Brain Institute, University of Calgary, Calgary, Canada; [29]Institute of Health and Society, Newcastle University Institute of Ageing, Newcastle University, Newcastle, UK ; [30]Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands ; [31]Department of Neurology, Maastricht University Medical Center, Maastricht, The Netherlands ; [32]Department of Neurology, J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, USA; [33]Stroke Research Group, Department of Brain Repair and Rehabilitation, UCL, London, UK ; [34]Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE), University of Edinburgh, Edinburgh, UK

Corresponding author: Pr HuguesChabriat, Dept of Neurology, HopitalLariboisière, 2 rue AmbroiseParé, 75010 Paris, France. Email: hugues.chabriat@aphp.fr

This study was supported by grants from FondationAvenirand FondationLeducq (Transatlantic Network of Excellence on the Pathogenesis of Small Vessel Disease of the Brain, http://www.fondationleducq.org).
This work follows from a project funded by the Centres of Excellence in Neurodegeneration (COEN), ref MRC MR/J006971/1, “Standards for determining the vascular contribution to neurodegeneration” in which many of the authors participated.

Word count
Abstract: 170
Full text: 7893
Figures: 4
Table: 1

Abstract (200 words)

Brain imaging is essential for the diagnosis and characterization of cerebral small vessel disease (SVD). Several magnetic resonance imaging (MRI) markers have therefore emerged, providing new information on the diagnosis, progression and mechanisms of SVD. Yet, the reproducibility of these SVD markers has received little attention despite being widely used in cross-sectional and longitudinal studies. This review focuses on the main SVD-related markers on MRI including: white matter hyperintensities, lacunes, dilated perivascular spaces, microbleeds and brain volume. The aim is to summarise, for each marker, what is currently known about: 1) its reproducibility in studies with a scan-rescan procedure either in single or multi centre settings; 2) the acquisition-related sources of variability; and, 3) the techniques used to minimise this variability. Based on the results, we discuss technical and other challenges that need to be overcome in order for these markers to bereliablyused as outcome measures in future clinical trials. We also highlight the key points that need to be considered when designing multicentre MRI studiesof SVD.

Keywords: MRI, cerebral small vessel disease, marker, brain volume, atrophy, white matter hyperintensities, microbleeds, lacunes, perivascular spaces, reproducibility, repeatability, variability

1. Introduction

Cerebral small vessel disease (SVD) is a major source of cognitive impairment and is the second most common cause of dementia in older aged people.[1] SVD is typically sporadic and related to aging and vascular risk factors, but can also be caused by genetic conditions[2]. The most frequent sporadic forms of SVD are arteriolosclerosis and cerebral amyloid angiopathy (CAA). A significant proportion of patients with SVD have associated neurodegenerative diseases such as Alzheimer's Disease (AD), which can lead to a mixed dementia.[3] In therapeutic trials for dementias, surrogate imaging markers of both pathophysiological processes should ideally be used to account for the potential mixed origin.

Pathological features of SVD are heterogeneous and out of reach of current in vivo imaging techniques. Nonetheless, parenchymal lesions presumed to be caused by small vessel changes have been adopted as markers of SVD, giving neuroimaging a central role for diagnosis and characterization[4].The spectrum of brain magnetic resonance imaging (MRI) manifestations of SVD is wide, ranging from minor white matter hyperintensities (WMH) often seen in population-based studies to numerous subcortical lesions. Indeed, MRI using various sequences allows an evaluation of several imaging markers related to SVD, including recent small subcortical infarcts, WMH, lacunes, enlarged perivascular spaces (PVS), cerebral microbleeds (CMB), and brain atrophy, as summarized by an international working group defining neuroimaging standards (STRIVE) for research into SVD[5].

The uncertain or lower reproducibility of these MRI markers, especially across centres, is a major concern in cross-sectionaland longitudinal observationalstudies as well as inclinical trials. Indeed, a key challenge is to minimize the variability caused by technological and subject-related factors. Such variability may confound the detection of disease- or treatment-related MRI changes, thereby limiting the power to detect and follow the progression of imaging markers of SVD. From the patient being imaged to the final imaging marker of interest, which is quantified by an expertrater or an automatic image-processing algorithm, there are several sources of variability. These can includesubject-related physiological changes or instrument-related factors such as field strength, head coil, gradients and sequence parameters (see Fig. 1). Each can affect the reproducibility of the marker. Additional variation may be introduced at the image post-processing stage, depending on which software is used for inhomogeneities correction or tissue segmentation, for example. The definition of lesions and the interpretation of images by several experts may be other sources of variability. The measurement of each MRI marker therefore depends on several factors, the effects of which need to be minimized, corrected or taken into accountin a posteriori analyses for studies to be comparable.

The aimsof this review are to summarise current knowledge regarding the reproducibility and variability of quantitative imaging markers of SVD including WMH,lacunes, PVS, CMB and brain volume in both single centre and multicentre contexts. To highlight technical solutions for common problems and to identify the remaining challengesthat need to be overcome for such markers to be considered valid and reliable. Other SVD-related MRI markers, due to their acute manifestation (recent small subcortical infarct) or their so far non-quantitative assessment (cortical superficial siderosis, cortical microinfarcts) have not been examined in reproducibility studies to our knowledge and hence were not included in the literature search. Post-processing tools or expert ratings that also contribute to the global variability and accuracy of the final marker measurement are beyond the scope of this review. For each MRI marker of SVD we review: 1) its reproducibility in studies with a scan-rescan designin single or across multiple centres; 2) the acquisition-related sources of variability; and, 3) the techniques that can be used to minimise methodological limitations. We also discuss those factors that need to be considered to deal with image variability inherent in MRI markers in multicentre design studies such as clinical trials or longitudinal studies.

2. Methods

Search strategy and selection criteria

This review was designed to focus on the reproducibility and variability of well-defined MRI markers of SVD: WMH, lacunes, PVS, CMB andbrain volume.

For each biomarker, taking into account the vocabulary heterogeneity of each marker (see supplementary material), a PubMed search has been conducted up toAugust 2015 14th with the associated terms “MRI” AND (“variability” OR “reproducibility” OR “reliability” OR “consistency”). Papers published before 2002 were not considered to focus on recent MR technology. Reviews, conference abstracts and articles not written in english were excluded from analysis.Articles from our personal databases were also incorporated to the search. To be included in the reproducibility summary (Table 1), studies had to include subjects who underwent a scan-rescan procedure either at the same centre or/and at different centres. Filtering was firstly based on review of the title and abstract for relevance. Studies were not mandatorily related to SVD (e.g., multiple sclerosis or AD) to include the largest number of studies about variability and reliability for the MRI markers of interest.

Term and metrics definitions

As the development and implementation of quantitative imaging biomarkers has been hampered in part by inconsistent use of terminology, an interdisciplinary group of radiologists,

statisticians, physicists, and other researchers worked to develop a comprehensive terminology to serve as a foundation for quantitative imaging biomarker claims[6]. Definitions agreed are reproduced in supplementary material.

In order to simplify the manuscript reading, we will only use the terms “within-centre reproducibility” (strictly speaking, the repeatability) to report results from scan-rescan procedures in a single centre; and “between-centre reproducibility” to report results from multicentre studies.

Several metrics are used to measure variability orreproducibility across different studies. A brief description of each is included in supplementary material.

3. Results

3.1 Variation in the number of studies and in the evaluation of the reproducibility

Number of studies

There was large variation in the number of studies conducted for assessing the reproducibility and variability of each SVD MRI marker. Brain volume and brain atrophy measures have received the most attention (19 studies with within or/and between centre repeated scans of the same subjects)[7-22]. This is probably related to the significant imaging research efforts in AD community during the last decade in which cerebral atrophy was of primary concern. Particularly, the Alzheimer’s Disease Neuroimaging Initiative (ADNI) permitted the most important studies on this aspect, including repeated examinations of the same subject on the same or on different scanners [14, 17]. Only five studies were found that evaluated the within-centre or between-centre variability of WMH measures[13, 23-26](Table 1). Few studies examined the potential sources of variability of WMH volumes, despite a large number of image processing techniques attempting to quantify them accurately. Finally, there were only four studies on CMB[25, 27-29] and no studyon the variability of PVS and lacunes. This probably reflects the absence of consensus until recently about radiological terminology and definitions as well as difficulties to automatically quantify these markers[30]. Indeed, computational segmentation of lacunes and perivascular spaces has not yet been validated in a specific study. Overall, the reproducibility of measure of longitudinal change, i.e. the evolution of a given marker as a function of time, has only been evaluated for brain volume change (brain atrophy) in 4 studies[9, 17, 19, 22]and not for other markers.

Evaluation of reproducibility

Several studies with a scan-rescan procedure selected in this review included different measures of variability in volumetric measurements that were not consistent across studies: intra-class coefficients, coefficients of variation, standard deviations, mean absolute values of differences. Hence, most of these studies were not directly comparable (Table 1). This prevented a quantified comparison between the multiple factors that may influence the measurements. It would be valuable for the community to harmonize measures of reproducibilityand make sure that all important aspects of reproducibility are included.

3.2 Reproducibility and variability of MRI markers

3.2.1 White-matter hyperintensities (WMH)

Definition and interest in SVD

From a radiological perspective, WMH of presumed vascular origin have a higher signal intensity compared to the normal-appearing white matter on proton-density (PD)-weighted, T2-weighted and FLAIR (Fluid-attenuated inversion recovery) sequences and may appear iso- or hypointense on T1-weighted images[5]. It is widely accepted that WMH are important for clinical outcome, in terms of cognitive and functional impairment[5, 31]. Many qualitative and quantitative techniques have been used to measure WMH. Qualitative approaches are based on visual rating scales completed by well-trained readers[32-34]. Quantitative techniques involve image-processing algorithms to obtain volumetric measures or spatial distribution of WMH lesions[35-39]. In this review, we only focused on quantitative techniques, as they offer a continuous measure that is a more reliable and sensitive alternative to visual rating scales[40].

Within-centre reproducibility

Several studies found good to excellent reproducibility of WMH segmentation and volumetry for different computational image analysis methods using a scan-rescan procedure in a single site[13, 23, 26] (Table 1). For example, in the study recruiting the largest number of subjects, De Boer et al.[13] assessed the within-centre reproducibility of WMH segmentation by comparing computational segmentation in 30 subjects scanned twice within a short interval at 1.5T. They found with their best segmentation pipeline a mean (sd) WMH volume difference expressed in % of intracranial volume of 0.01% (± 0.05%) and a coefficient of variation less than 6%.

Between-centre reproducibility (effect of various vendors, coils)

No reportwas found about between-centre reproducibility of WMH measures in large cohorts evaluated in several centres and using the same magnetic field.

Magnetic field effect

It is not clear how the magnetic field strength impacts the measurement of WMH volume beyond the expected improved resolution at higher field. Indeed there are relatively few direct comparisons of different scanner field strengths in any situation[41]. In 10 patients with SVD, Theysohn et al. showed that FLAIR images acquired at 7T highlighted WMH known from 1.5T with comparable extent[25]. However, WMH were not adequately quantified in this study as neither number nor volume was assessed. On the contrary, in patients with multiple sclerosis, it has been reported that 3T scans showed a mean 10% higher WMH volume compared to 1.5T scans in total lesion volume based on PD-weighted images[24].

Sequence effect

WMH detection and automated quantification usually relies on the FLAIR sequence, which produces T2-weighted images with cerebrospinal fluid suppression. It has been recently shown that 3D FLAIR sequence parameters could be optimized to increase lesion detectionand identification in multiple sclerosis compared with default 3D FLAIR or 2D FLAIR (Fig. 2)[42]. This approach might also be used in SVD. Indirectly, this study showed that MR parameters (echo time, inversion time) influences the tissue contrast between the normal white matter and white matter lesions on which the algorithms to segment WMH heavily depend.

Some pipelines have used an interleaved PD/T2 acquisition for WMH segmentation, especially as this dual spin echo sequence is more sensitive to WMH in the deep basal ganglia and thalamus than FLAIR and also allows better detection of perivascular spaces[60,43].Multisequence image analysis approaches to quantify WMH, eg which combine FLAIR with T2* or T1 or T2 or several sequences together, are increasingly used and emphasise the need to obtain several key standard sequences when studying SVD.

Partial volume effects

Partial volume effect is the combination of signals from different tissue types in a single voxel whose image intensity is thus dependent on proportions of each tissue. Partial volume effects with clinically feasible resolutions complicate delineation of the lesion borders[44]. Particularly, smaller voxel sizes allow more accurate estimation and yield larger total lesion volumes, especially for patients with small lesions[45]. For example, in 10 patients with multiple sclerosis, lesion volumes were reported to be on average 9% greater with 3 mm compared to 5 mm slice thickness[45].

Other factors with some effects

The approach to post-processing for bias field (or B1 inhomogeneities) correction can have an important effect on WMH volume quantification according to the computational method used[46].

Reproducibility in longitudinal studies

No study was found to assess the reproducibility of any longitudinal measurement of WMH.

3.2.2 Lacunes and perivascular spaces

Definition and interest in SVD

From a radiological perspective, lacunes of presumed vascular origin are round or ovoid, subcortical, fluid-filled cavities of between 3 mm and 15 mm in diameter, consistent with a previous acute small subcortical infarct or haemorrhage in the territory of one perforating arteriole[5]. Number, size, shape and location are measures of interest for lacunes. To date, very few automatic image processing methods exist to segment lacunes, and most studies rely on semi-automatic computational approaches or manual segmentations by trained observers[37, 47-50].

PVS are fluid-filled spaces that follow the typical course of a small vessel as it goes through grey or white matter and have a signal intensity similar to CSF on all sequences[5,51]. Number, size, shape and location are measures of interest for PVS. Rating methods have been developed to describe these markers[52, 53]. To date, very few computational image processing methods exist to segment PVSthough some are in development[37, 54-56]. The first image processing methods for both lacunes and PVS detection were not fully robust and efficient, requiring time-consuming user manual intervention but new methods are more automated and much faster.

Within- or between-centre reproducibility

The literature is very sparse concerning the reproducibility of measurements describing lacunes or PVS[51].In 20 subjects scanned twice on a 1.5T scanner, one study assessed a good reproducibility for lacunar volume with the used dedicated software but the extent or numbers of lacunes per subject were not specified [26].

High-resolution images should be more sensitive for detecting lacunes and PVS because they are small. Therefore, their volume assessment may heavily depend on spatial resolution and geometrical inaccuracies induced by gradient nonlinearities.

Reproducibility in longitudinal studies

No study was found to assess the reproducibility of any longitudinal measurement for both lacunes and PVS.

3.2.3 Cerebral microbleeds

Definition and interest in SVD

From a radiological perspective, cerebral microbleeds (CMB) associated with SVD are small hypointense lesions on paramagnetic sensitive MR sequences such as T2*-weighted gradient-echo (GE) sequence, and are most frequently located in the cortico-subcortical junction, deep grey or white matter in the cerebral hemispheres, brainstem and cerebellum[5, 57]. CMB are associated with SVD and lobar CMB are often seen in patients in memory clinics who have CAA [5, 58-60]. Most studies suggest that they are associated with impaired cognitive function[61, 62], although their direct clinical impact remains uncertain. Several rating scales have been designed to minimize observer variation[63, 64] along with several post-processing algorithms for semi-automatic or automatic quantification[65-69]. We consider in this review all methods to study CMB as a few computational methods have been published and tested in reproducibility studies. Many MRI acquisition factors are expected to influence CMB detection as their size is a consequence of susceptibility effects that are highly dependent onMRI sequence parameters. In fact, hypointensities are typically larger than the physical size of the corresponding histopathological lesion (blooming effect)[70].

Within- or between-centre reproducibility

No study was found to assess CMB detection reproducibility through a scan-rescan procedure at the same centre or between centres with the same imaging protocol. However, several studies examined various factors involvedin CMB detection such as magnetic field strength or sequence parameters.

Magnetic field effect

Intravoxeldephasing due to susceptibility effects and associated signal loss in gradient-echo images increase with higher magnetic fields[71]. Several studies hence demonstrated an increased number of CMB at higher field strength[25, 27, 29] with an increase in contrast to noise ratio[27, 28] and CMB size[25, 28] as illustrated in Fig. 3.For example, mean increases from 1.5T to 3T were 48% for contrast to noise ratio and 12% for CMB size on 119 CMB detected in 4 subjects[70]. In addition, more CMB are found at higher field strength; Theysohn et al. found 101 CMB at 7T (mean size: 3.2 mm) against 33 at 1.5T (mean size: 1.4 mm) in 10 subjects[25].

Sequence effect

3D T2*-weighted gradient-echo images reveal significantly more CMB in more subjects compared to 2D T2*-weighted gradient-echo images, primarily because of higher spatial resolution[72]. In 200 older aged participants, Vernooij et al. reported 71 subjects with at least one microbleed (median number 2.5) using the 3D sequence compared to 42 subjects (median number 1.0) with the 2D sequence. Likewise, the susceptibility weighted imaging (SWI) technique, which uses phase images to enhance susceptibility-related contrasts in the resulting image, is particularly sensitive to paramagnetic and ferromagnetic substances[73]. CMB are better delineated on SWI images compared to classic T2*

images[25, 28, 74, 75] (Fig. 3). Another study also confirmed that SWI detected more CMB than conventional T2* images (prevalence of 40% vs. 23%, total number was 284 vs. 219 in 141 patients)[76]. To note, the differentiation between CMB and vessels might be more difficult with SWI compared to T2* images although this is a matter of debate [75, 77]. A last point is that susceptibility-weighted sequences are different according to vendors (e.g., 'SWI' for Siemens, 'SWAN' for General Electric, 'VenoBOLD' for Philips) and only the Siemens 'SWI' technique uses phase data. Therefore, the effect of vendor's susceptibility-weighted technique on CMB detection needs further assessment.

Sequence parameters effect

The echo time (TE) reflects the timescale on which dephasing occurs, with long echo times enabling more time for dephasing thereby enlarging the susceptibility effect. Thus, choice of TE directly impacts on the number and volume of microbleeds detected as well as the number of structures that may be mistaken for microbleeds (calcium or iron deposits, flow voids, etc.). Increasing TE improves the sensitivity to detect CMB but more artefacts are generated near air-tissues interfacescausing uncertainty about rating some lesions[78, 79].Otherwise, thin MR imaging sections compared with thicker sections were associated with a small decrease in CMB diameter and a large increase of contrast (CMB signal relative to tissue signal) and thus an improved CMB detection (Fig. 3)[28], due to a reduction of partial volume effects.

Reproducibility in longitudinal studies

No study was found to assess the reproducibility of any longitudinal measurement for CMB.

3.2.4 Brain volume

Definition and interest in SVD

Brain volumetry refers here to measures obtained through computational segmentation of high-resolution anatomical images. Global brain volume, intracranial volume (classically used for normalization purpose) and brain parenchymal fraction (BPF, brain volume divided by intracranial volume) were considered in this review as these are important markers in SVD [80, 81, 82 , 83-85]. For example, BPF has been associated with dementia rating scales and the modified Rankin score in CADASIL[81, 86], a monogenic SVD[2]. In sporadic forms, brain volume was notably associated with the presence and progression of lacunes[87]. Gray matter, white matter or hippocampus volumes are sometimes specifically assessed and can be markers of interest in SVD studies, but were not included in this review as these structures depend more on post-processing stages. However, no study amongst any of the following has assessed the influence of SVD markers, particularly WMH, on the performance of any brain volume measurement methods.

Within-centre reproducibility

Overall, several studies reported a satisfactory within-centre reproducibility of brain volumetric measurements at 1.5T[7, 10, 13-15, 17, 18] or 3T[10, 14, 15, 18] with the main brain image processing suites (FreeSurfer[88, 89], SPM[90], FSL[91],BrainVisa[92]). Table 1 presents per study the numbers of centres, subjects, rescans per subject and main results for reproducibility. For example, in a large cohort of 671 subjects who underwent two successive 3D T1 weighted scans, the median of the relative absolute difference in brain volume was found to be less than 1% [17]. However, the 90th percentile of the absolute value of the difference was reported at 5.11%, which is more than twice the value that would be expected for a Gaussian distribution. This highlights the importance of studying several statistical outcomes to infer the reproducibility beyond the median or mean value of the difference.Thus, little variability is added when measuring brain volume in a single centre without changing the conditions of acquisition, at least regarding commonly reported disease-related changes. It is worth noting that most studies with a scan-rescan design included small samples (all n< 50 except Cover et al.[17]) and recruited mainly healthy subjects (two studies with mild cognitive impairment (MCI) and AD patients[14, 17]and one study in multiple sclerosis[16]).

Between-centre reproducibility (effect of various vendors, magnetic field strengths, coils)

Effects of combining several MRI vendors, static magnetic field strengths or coil designs on brain volumetry can be assessed in studies where subjects are scanned several times in different centres (Table 1). As expected, compared to the within-centre variability, the between-centre volume variability was much higher at 1.5T[7, 8, 10, 11, 14, 15, 18, 65] and 3T[10, 12, 14, 15, 18]. In one of the largest multicentre studies with a scan-rescan procedure (a subset of ADNI database in which 172 subjects were rescanned once in 59 different MRI sites on 17 different scanner types), the effects of scanner hardware and imaging protocol on volume quantification were analyzed[14]. The most striking result was that the variance in volumetric measures was 10 times higher when subjects were imaged on different scanners compared to acquisitions on the same scanners. For example, for brain volume, the median (standard deviation) of the absolute within-subject variability was 0.3% (0.5%) vs. 4.8% (4.6%) for repeated scans on the same scanner vs. on different scanners. This was essentially explained by the scanner-dependent geometrical inaccuracies and by the differences in the gray to white matter tissue contrast from variation in the imaging sequences used and RF coils (while this contrast is obviously impacting gray matter and white matter volumes, it also influences the total brain volume computation[14]). Such geometric distortions originate from gradient field nonlinearity (in theory, a gradient system is supposed to produce an incremental magnetic field varying linearly with distance from isocenter) and the static field inhomogeneity, whilst the contrast between brain tissues is determined by a variety of sequence parameters. Both factors are dependent on the scanner hardware

(a combination of vendor, magnetic field and coil systems), which explained 30 to 50% of the variance of any volumetric measure in that study. In another study, 15 subjects underwent 4 scan sessions at 3 sites with two different vendors[10]. Interestingly, combining data of different vendors and/or field strength did not significantly change the standard deviation of the volume differences across conditions relative to the test–retest reproducibility within a fixed MRI system. However, the mean volume difference did change within one standard deviation from zero for intracranial volume.

Upgrades and repair

Only a few studies have examined the effects of MR systems upgrades and repairs on volumetric measurements. It appears that mixing volumes derived from data acquired across a major scanner upgrade (such as replacement of the magnet, gradient system and software) do not show significant changes, although volumes of some specific structures may show a slight scanner specific volume bias[10]. Similarly, consistent measures for brain structure volumes have been found before and after gradient coil replacement in a 3T system[93]. However, inadvertent differences in sequence parameters for example may be introduced during the upgrade (human error)[94]which can cause major variance between pre- and post-upgrade scans if not detected quickly. Scanner upgrade or replacement are almost impossible to avoid, and therefore centres should consider having longitudinal phantom and volunteer assessments for quality assurance and to calculate differences in brain volumetry across time that would allow conversion factors to be generated.

Sequence effect

Use of 3D inversion prepared T1-weighted fast gradient-echo sequences for brain tissue segmentation and volumetry is universal nowadays. It is widely used for its contrast properties and is a standard for most MRI vendors. It is therefore easy to implement in multicentre studies. Other sequences such as multi-echo fast low angle shot sequences have similar contrast properties and showed comparable volume reproducibility from Bland-Altman plot analysis than the MPRAGE sequence for example[10]. However, comparison between volumes derived from both sequence types demonstrated significant biases in the mean volume difference, probably due to a differing T_2* sensitivity[10].

Isotropic voxels are superior to non-isotropic voxels for accurateestimation of brain volume. In sequences with non-isotropic resolution, the choice of image orientation (sagittal, coronal or axial) naturally yields volumetric differences when comparing datasets acquired with different image orientations[95]. These differences may be due to partial volume averaging and susceptibility-induced geometric distortions. It is anticipated that these findings can be generalized to all MRI-derived metrics of brain volume.

Contrast effect

Variability of the tissue contrast, defined as the ratio between gray and white matter tissues image intensities on T1-weighted images, relates to the choice of MR sequence parameters such as echo time (TE), repetition time (TR), inversion time (TI) and flip angle (FA) along with the selection of field strength and coil. The contrast has been shown to have non-negligible influence on brain tissues segmentation and thus volumetry[7, 14]. For example, a 6% change in the average contrast led to a 2% change in the computed brain volume[14], which is of a similar size as the annual change in brain volume seen in the early stages of AD so could easily be interpreted as a disease effect.Notably, datasets with a higher contrast have a lower variability, but here the effect of different burdens of WMH have not been tested.

Subject-related factors

Physiological variability within the subject being imaged is a source of variability beyond the disease-related variability. It has even been suggested in a study of BPF variability in multiple sclerosis that the variability due to physiological and pathological causes is important and likely larger in magnitude than scan–rescan repositioning effects[16]. The authors found that the variance assigned to physiological fluctuations was nearly two times greater than the variance due to patient repositioning during scan-rescan. This implies that scan–rescan experiments might only provide a lower bound on the true error in repeated volumetric measurements from MRI exams. It is well known that the hydration status of an individual can affect the measurement of brain volume. Specifically, lack of fluid intake for 16 hours has been found to result in a 0.55% decrease in cerebral volume and re-hydration an increase of 0.72% in brain volume[96]. Inthe ADNI dataset and a large cohort of multiple sclerosis patients there was a significant effect of time-of-day on the BPF, with a greater brain volume in the morning[97]. Similarly, brain volume varies with plasma sodium[98]. These results suggest potential acquisition time and hydration biasesshould be randomized or statistically controlled to account for the day-to-day brainvolume fluctuations.

Factors with no or little effect

Correction for so-called B1 inhomogeneities (variations of image intensity due to the radiofrequency field) was not found to substantially change the reproducibility of whole-brain volumetric measurements[10], unlike its effect on WMH and focal infarcts[99]. Similarly, some studies found that greater acceleration factor in parallel imaging (this technique combines the signals of several coil elements in a phased array to reconstruct the image and finally improve the signal-to-noise ratio or reduce the scan time) had very little influence on test–retest accuracy[12, 100]. However, these studies tested the influence of acceleration factor in healthy controls. Results might be different if older aged subjects or patients move during the acquisition of the reference lines part of the parallel imaging

strategy. Finally, reproducibility of brain volumetric measures was also relatively unaffected by small differences in voxel geometry compared to the effects observed on measures of cortical thickness[12].

3.2.5 Brain atrophy

Definition and interest in SVD

Brain atrophy is the difference of total brain volume between two time points computed by a dedicated image processing software from high-resolution anatomical images.Quantitative measures derived from high-resolution anatomical images are considered over atrophy scales (not covered in the present review) as the quantitative measures have better sensitivity to change over time[101]. Surprisingly, very few studies reported brain atrophy measurement in SVD cohorts[102, 103].Several methods are available, such as SIENA, which estimates the percentage brain volume change of the same subject between two time points, by registering the two brain images together and resampling into the space halfway between the two[104, 105] andthe Boundary Shift Integral (BSI),that is an automated method that determines the total volume through which the boundaries of a given cerebral structure have moved[106].One alternative method to assess brain atrophy is by performing tissue segmentation at both time points, but this approach suffers from a higher variability given the added variability of each measurement[22].

Within-centre reproducibility

Within-centre reproducibility of quantitative image-processing methods for atrophy assessment have been shown to be relatively high in several studies (Table 1)[17,9, 19]. For example, a study using the ADNI dataset reported a 0.35% mean difference in 385 subjects acquired in more than 50 sites (105 healthy controls, 195 subjects with MCI and 85 with AD) undergoing two SIENA measurements 12 months apart[17]. Further, the 90th percentile of the difference was much larger than expected (1.33%) for a Gaussian distribution. These results should be compared to the annualized mean percentage brain volume change obtained in that study: -0.65% ± 0.82% for healthy controls, -1.15% ± 1.21% for MCI, -1.84% ± 1.33% for AD.

Between-centre reproducibility (effect of various vendors, magnetic field strengths, coils)

Some studies have examined the between-centre reproducibility of atrophy measurements using SIENA[17, 22, 107] or other image-processing techniques[22].It has been shown that even with scanner of the exact same model, scanner drift (a shift in the magnet's field) and inter-scanner variability may cancel out longitudinal brain volume change[44].It is worth noting that compared to voxel-based morphometry approaches, SIENA, which corrects for differences in imaging geometry using the outer skull surface for both time points, reduced the effects of scanner drift and inter-scanner variability on longitudinal morphometric results[107]. Similarly, other registration-based algorithms such as BSI

showbetter reproducibility compared to segmentation-based techniques such as Freesurfer or SIENAX[22]. However, segmentation-based methods using atlases can be improved by using fine registration methods andage-relevant templates[108].

Evaluation of coil effects on volumetric measurements showed that a change from a 12- to a 32-channel coil was responsible for a shift of 0.5% in the whole brain BSI[100], which appears unacceptable as it corresponds to the annual atrophy rate observed in healthy subjects[109].

Finally, analysing the back-to-back scans of the ADNI dataset in 200 pairs of images (118 subjects with 0-12 months pairs and 82 subjects with 6-24 months pairs), it has been reported than 3T was no more reproducible than 1.5T for the whole brain atrophy measure using SIENA[110].

Contrast effect

Atrophy measurement is strongly dependent on image contrast. Indeed, the error using BSI may exceed 100% if image contrast properties dramatically differ between the two scans in a measurement pair [111]. For example, in 10 volunteers, a mean brain volume change of -0.43% has been measured versus a +1.85% change when the second scan was acquired with a flip angle equal to 12° instead of 25° (the flip angle directly impacts the image contrast).

Noise effect

Noise is an unwanted random signal from various origins including: the subject's body, coils, electronics, nearby machinery, etc.The signal-to-noise ratio is often computed as the ratio of the average signal intensity over the standard deviation of the noise. Signal-to-noise ratio directly impacts BSI values as, for example, the position of the brain surface is seen to retract as noise in the second time point scan increases[111]. Thus, the average percent volume change progressively diverges from zero with increasing noise mimicking increased brain atrophy.

Subject-related factors

Motion willhave an impact on atrophy measurement. For example, it has been reported that, as the amplitude of the motion in the second scan increases, the mean absolute brain and ventricular BSI brain volume changes as well as the associated standard deviation increase progressively[111]. However, the different levels of motion have been simulated in this latter study and the severity may have been overestimated.

Subject's head positioning in the magnet also influences brain atrophy measures. Indeed, nonlinear gradient distortions associated with shifts in the long magnet's axis can significantly decrease the accuracy and precision of MRI-derived measures of brain atrophy assessed by SIENA[112]. These negative effects increase in magnitude with: (i) increases in the Z-distance (long magnet's axis) between the brains to be compared at two time points; and, (ii) increases in the Z-distance from magnet isocenter to the center of the pair of brains to be compared[112]. Typical gradient distortion field for a T1-

weighted acquisition is presented in Fig. 4. It shows the dramatic shifting experienced by voxels especially at the border of the brain and the need for accurate correction and repositioning of patients in longitudinal studies.

3.3 Techniques to manage variability

We have highlighted a large number of reasons why quantitative MRI markers of SVD may be variable beyond subject-related or disease-related factors. These sources of variability introduce significant barriers, and potential biases, to multicentre imaging studies of SVD. Fortunately, while adding variability to the measurements of markers is inherent to imaging studies, some techniques have been designed to reduce it.

First, an optimised MRI protocol with high contrast-to-noise ratio, minimal artefacts and good reproducibility must be determined. Hereto, a methodological approach has been proposed to target a protocol with the highest possible reproducibility based on the assessment of quality measures[21]. The principle is to acquire several datasets including: within-subject within-centre datasets, between-subjects between-centre datasets and between-subject within-centre datasets. By analysing the summary data quality and quantitative measures extracted with dedicated and validated software the desired optimal MRI protocol can be determined.

Second, to account for the problem of upgrades in longitudinal studies, Jovicich et al. advise planning a 'system upgrade calibration study' as part of the design, with subjects scanned shortly prior to and immediately after the upgrade so as to derive a correct estimation for potential systematic biases[10].

Third, several corrective techniques based on the MR system characteristics [113, 114] or using a dedicated phantom [112, 115, 116] have been developed to reduce the effects of gradient distortion on imaging reproducibility.The advantage of using a phantom travelling from one site to another is to apply a unified correction independently of the vendor's own corrections.Correction for gradient nonlinearity and intensity non-uniformity reduces the variance of longitudinal changes in brain volumes and improves the accuracy for detecting subtle volumetric changes[117]. For example, in 208 subjects, the raw mean percentage of brain volume change was -0.59% against -0.46% when images were corrected for gradient nonlinearity and intensity non-uniformity[117]. However, if the expected group difference far exceeds brain volume change related to geometric distortions, the extra time and expense associated to the phantom-based correction needs to be carefully evaluated.

Finally, the use of dedicated statistical methods is important in analysing multicentre MRI data. A simple *a posteriori* method to account for scanner hardware variability (combination of different

devices, coil types and field strengths) is to include the scanner hardware as a covariate in regression models [14]. Moreover, it has been shown that T1-weighted regional anatomical brain volume data can be reliably combined across 1.5T and 3T centres with the application of an appropriate correction procedure (regression-based correction function when data varied in a linear and systematic fashion)[118]. More sophisticated statistical methods such as linear mixed-effect models including protocol as well as individual MRI acquisition parameters have also emerged to more accurately model longitudinal changes (BPF and volume of WMH)[119]. Another statistical approach is based on the statement that classical inference testing is inappropriate because it is designed to detect differences and not to prove similarity. Equivalence testing has then been applied to determine if MRI-based measurements obtained under variable conditions can be pooled[120]. This could be generalized to all MRI-derived measurements. Otherwise, Moorhead et al. used scanner specific priors (probability mappings of tissue occupancy) in the segmentation pipeline to reduce tissue classification differences between scanners to improve statistical power. Interestingly, they found that the between scanner differences were not reduced to the level of within scanner variability, the ideal for scanner harmonisation[121].

Ultimately, central application of automatic quality assessment[94, 122]and image processing on raw data, using consistent techniques, can reduce the variability introduced by a posteriori image processing across centres.

3.4 Suggestions for future research

A between-centre variability study in a large group of subjects,representing a range of types and burdens of SVD lesions, to assess the effects of vendor and field strength (protocols as close as possible) on reproducibility of SVD lesions and brain volumes is missing. So are experiments examining the effects of changing coil, correcting for gradient distortion or field heterogeneities, upgrading the system, or varying classical parameters such as TE, TI or TR, on the final WMH volume. Finally, WMH appears to be a reliable marker in single centre studieswhere all acquisition and processing parameters are kept constant, but additional studies are missing to infer the true reproducibility of WMH measures in a multicentre context.

As for WMH, a multicentre study assessing the effect of scanner hardware (same magnetic field but different vendors and models) on CMB detection with a given optimized protocol would greatly add to the knowledge on the reliability of CMB detection. New data on methods of automated detection of CMB throughout different brain regions are needed. Also, the analysis of the potential effects of coil, and voxel size would be particularly useful. Finally, CMB can be used as a reliable marker in single

centre studies but additional efforts are needed to determine the sources of variability in multicentre data.

For lacunes and PVS, more automated and validated image processing algorithms are required, followed by assessment of reproducibility through both within and between centres acquisitions. In view of the likely interactions between acquisition protocols and performance of analysis methods, the assessment of all SVD features should be included in any study that assesses reliability of WMH and/or CMB and/or atrophy quantification.

4. Ten key points for assessing quantitative SVD markers in multicentre studies

Based on our review of the current literature regarding reproducibility and acquisition-related sources of variability of MRI markers in SVD and on personal experiences, we propose the following key points should be considered when designing or analysing a multicentre MRI study in SVD. Obviously, prior to acquisitions, the scientific question should guide which marker is the focus, although opportunities to assess all SVD markers should be obtained where possible. Optimising acquisition for contrast, signal-to-noise ratio and spatial resolution is particularly important when imaging elderly or cognitively impaired patient populations who will not tolerate prolonged acquisition times. Importantly, although it is essential to have accurate and reproducible measurements, the nature of the questions will be guiding the level of precision that is needed. Requirements might be different between interventional clinical trials with relatively short term duration and long-term population-based observational studies. For example, in some situations, the added noise due to multicentre variability will not alter a strong statistical association between one SVD-related MRI marker and one clinical outcome. Conversely, studying the quantitative evolution of one SVD-related marker during a short follow-up will require a very fine quantification that necessitates a variability caused by technological factors as small as possible. Hence we organized our recommendations as being a) advisable whatever the question guiding the study as they tackle effects being important and/or involve easy-to-follow techniques to manage variability, b) advisable in some conditions, i.e. when the expected effect of interest is small relatively to the unwanted noise caused by the multicentre setting.

a) strongly advisable and easy-to-follow recommendations

1. Automated quality checks for MRI sequence parameters are advisable (for each scan, comparison of the acquisition parameters with the intended protocol for each scanner). Also,

images should be monitored for motion and other artefacts such as magnetic susceptibility artefacts, aliasing artefacts or Gibbs artefacts. A minimum standard for quality control will have to be defined to exclude problematic data (for example, based on the severity of artefacts or from automatic measures of signal to noise ratio, background noise, etc.). Indeed, even if the computational method seems to run properly, artefacts can dramatically alter the data.

2. MRI gradient nonlinearities have to be corrected at least with corrections proposed in options by the manufacturer as they can affect dramatically measures of brain atrophy.
3. Subject's positioning in the scanner is another key aspect and strict instructions for MR technicians may be particularly useful to reproduce confidently the position of the head, from one subject to another, and from the first time point to the following time points, in longitudinal studies.
4. Considering the hardware, the same coil has to be used during an MR protocol both in cross-sectional and longitudinal studies. When upgrades or repairs are needed during the study period, a strategy should be in place to scan some participants twice, before and after the upgrade, to estimate the potential bias for a posteriori analyses. Otherwise, the imaging protocol and hardware should be kept as constant as possible over the whole time of the study.

b) recommendations to evaluate given the targeted trade-off between sensitivity and variability

5. Within-centre reproducibility is much higher than between-centre reproducibility. The need for a multicentre setup should then be carefully evaluated while taking into account specific aspects such as number of participants, subject transportation, data centralization, quality control and cost. For multicentre studies, the use of only one vendor or only some models can be an option.

6. In multicentre studies, protocol harmonization is a mandatory step to obtain similar sequences with acquisition parameters as close as possible. There is no way to get a perfect harmonization of MRI protocols between different vendors and models. Therefore, it is a trade-off between a high reproducibility of markers (few centres, same vendors) and a sufficient statistical power through the recruitment of a large number of subjects, generally in several centres with different scanner vendors and models. Specifying acceptable ranges for different acquisition parameters may help obtain satisfactory harmonization for large-scale clinical studies.

7. Spatial resolution is a key parameter in assessing correctly the variation of any MRI marker such as brain volumetry, longitudinal atrophy, volume of WMH or quantification of lacunes and PVS. Increasing spatial resolution decreases signal-to-noise ratio and increases acquisition duration. Sensitivity for quantifying small objects such as lacunes and PVS will be compromised if using voxels as big as the object itself or where there is a slice gap. Selection of optimal spatial resolution must therefore be targeted to the primary question.

8. When analysing multicentre MRI data, the use of statistical models to take into account the induced variability is recommended (adjusting on centre or using random effects models for example).

9. While some quantitative MRI markers of SVD are regularly assessed (brain atrophy,WMH volume), others are difficult to quantify given their small size in relation to conventional anatomical MRI (lacunes, PVS) or their strong dependence on MRI parameters (CMB). High-field MRI and detailed studies of these particular markers may be performed in studies nested within larger studies assessing more common markers in a clinical setting. For CMB characterization, SWI or quantitative susceptibility mapping are promising techniques. Since the total magnetic susceptibility of a CMB is an intrinsic physical property independent of imaging characteristics, mapping CMB by using quantitative susceptibility mapping may be an alternative approach more consistent over a wide spectrum of imaging parameters[79].

10. It is important for studies to acknowledge that variation is inevitable and they should detail the steps that have been made to minimise this variation.

Conclusion

There is a need to reduce image variability induced by technological choices (coil, MR sequence, MR parameters and options) as much as possible in large multicentre research studies and clinical trials using MRI in SVD. Indeed, the added variance due to the multicentre setup can be deleterious in various applications and should be minimised using multiple approaches. Protocol harmonization and a posteriori corrections can help in minimizing biases; minimisation on centre or scanner type may also be advisable. Even if several sources of variability have been previously evaluated, most studies have focused on brain volume measurements whilefewhave focused on WMH

quantification. Further, less is known concerning the sources of variability when quantifying lacunes, CMB and PVS. Additional studies are warranted to investigate the variability of these makers, which is a key aspect for future longitudinal or multicentre studies. Other potential SVD-related MRI markers derived from new and multiple techniques (such as diffusion tensor imaging, arterial spin labelling, magnetization transfer mapping, blood-brain barrier permeability imaging, BOLD functional MRI, cortical morphology, cortical microinfarcts, cortical superficial siderosis) will also have to address these important questions regarding theirreproducibility to facilitate large-scale, multicentre research.

Author contribution

FDG, EJ and HC designed the review. FDG collected and analysed the literature. FDG, EJ and HC wrote the initial draft of the review. JMW critically revised the manuscript. The manuscript was then sent to the imaging and small vessel disease experts involved in the COENgroup (network of Centres of Excellence in Neurodegeneration, http://www.coen.org). All authors of the present manuscript agreed to contribute and revised carefully the manuscript. Finally, all authors accepted the current version.

Disclosure/conflict of interest

The authors declare no conflict of interest.

Supplementary material for this paper can be found at

http://jcbfm.sagepub.com/content/by/supplemental-data

References

1. Román GC, Erkinjuntti T, Wallin A, Pantoni L, Chui HC. Subcortical ischaemic vascular dementia. Lancet Neurology 2002; 1(7): 426-436.
2. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserve E, Bousser M-G. Cadasil. Lancet Neurology 2009; 8(7): 643-653.
3. Selnes OA, Vinters HV. Vascular cognitive impairment. Nature Clinical Practice Neurology 2006; 2(10): 538-547.
4. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. Lancet neurology 2010; 9(7): 689-701.

5. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. The Lancet Neurology 2013; 12(8): 822-838.

6. Kessler LG, Barnhart HX, Buckler AJ, Choudhury KR, Kondratovich MV, Toledano A et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Statistical methods in medical research 2015; 24(1): 9-26.

7. Schnack HG, van Haren NEM, Pol HEH, Picchioni M, Weisbrod M, Sauer H et al. Reliability of brain volumes from multicenter MRI acquisition: A calibration study. Human Brain Mapping 2004; 22(4): 312-320.

8. Ewers M, Teipel SJ, Dietrich O, Schönberg SO, Jessen F, Heun R et al. Multicenter assessment of reliability of cranial MRI. Neurobiology of Aging 2006; 27(8): 1051-1059.

9. Smith SM, Rao A, De Stefano N, Jenkinson M, Schott JM, Matthews PM et al. Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: Cross-validation of BSI, SIENA and SIENAX. NeuroImage 2007; 36(4): 1200-1206.

10. Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B et al. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 2009; 46(1): 177-192.

11. Reig S, Sanchez-Gonzalez J, Arango C, Castro J, Gonzalez-Pinto A, Ortuno F et al. Assessment of the Increase in Variability When Combining Volumetric Data From Different Scanners. Human Brain Mapping 2009; 30(2): 355-368.

12. Wonderlick JS, Ziegler DA, Hosseini-Varnamkhasti P, Locascio JJ, Bakkour A, van der Kouwe A et al. Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. NeuroImage 2009; 44(4): 1324-33.

13. de Boer R, Vrooman HA, Ikram MA, Vernooij MW, Breteler MMB, van der Lugt A et al. Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods. Neuroimage 2010; 51(3): 1047-1056.

14. Kruggel F, Turner J, Muftuler LT, Alzheimers Dis Neuroimaging I. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 2010; 49(3): 2123-2133.

15. Huppertz H-J, Kröll-Seger J, Klöppel S, Ganz RE, Kassubek J. Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. NeuroImage 2010; 49(3): 2216-2224.

16. Sampat MP, Healy BC, Meier DS, Dell'Oglio E, Liguori M, Guttmann CR. Disease modeling in multiple sclerosis: assessment and quantification of sources of variability in brain parenchymal fraction measurements. Neuroimage 2010; 52(4): 1367-73.

17. Cover KS, van Schijndel RA, van Dijk BW, Redolfi A, Knol DL, Frisoni GB et al. Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans. Psychiatry Research-Neuroimaging 2011; 193(3): 182-190.

18. Shokouhi M, Barnes A, Suckling J, Moorhead TW, Brennan D, Job D et al. Assessment of the impact of the scanner-related factors on brain morphometry analysis with Brainvisa. BMC Med Imaging 2011; 11: 23.

19. de Bresser J, Portegies MP, Leemans A, Biessels GJ, Kappelle LJ, Viergever MA. A comparison of MR based segmentation methods for measuring brain atrophy progression. Neuroimage 2011; 54(2): 760-768.

20. Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IA, Farrell JA et al. Multi-parametric neuroimaging reproducibility: a 3-T resource study. NeuroImage 2011; 54(4): 2854-66.

21. Chalavi S, Simmons A, Dijkstra H, Barker GJ, Reinders A. Quantitative and qualitative assessment of structural magnetic resonance imaging data in a two-center study. Bmc Medical Imaging 2012; 12.

22. Durand-Dubief F, Belaroussi B, Armspach JP, Dufour M, Roggerone S, Vukusic S et al. Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. AJNR. American journal of neuroradiology 2012; 33(10): 1918-24.

23. Wei X, Warfield SK, Zou KH, Wu Y, Li X, Guimond A et al. Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy. Journal of magnetic resonance imaging : JMRI 2002; 15(2): 203-9.

24. Sicotte NL, Voskuhl RR, Bouvier S, Klutch R, Cohen MS, Mazziotta JC. Comparison of multiple sclerosis lesions at 1.5 and 3.0 Tesla. Investigative Radiology 2003; 38(7): 423-427.

25. Theysohn JM, Kraff O, Maderwald S, Barth M, Ladd SC, Forsting M et al. 7 Tesla MRI of Microbleeds and White Matter Lesions as Seen in Vascular Dementia. Journal of Magnetic Resonance Imaging 2011; 33(4): 782-791.

26. Ramirez J, Scott CJ, Black SE. A short-term scan-rescan reliability test measuring brain tissue and subcortical hyperintensity volumetrics obtained using the lesion explorer structural MRI processing pipeline. Brain topography 2013; 26(1): 35-8.

27. Stehling C, Wersching H, Kloska SP, Kirchhof P, Ring J, Nassenstein I et al. Detection of asymptomatic cerebral microbleeds: A comparative study at 1.5 and 3.0 T. Academic Radiology 2008; 15(7): 895-900.

28. Nandigam RNK, Viswanathan A, Delgado P, Skehan ME, Smith EE, Rosand J et al. MR Imaging Detection of Cerebral Microbleeds: Effect of Susceptibility-Weighted Imaging, Section Thickness, and Field Strength. American Journal of Neuroradiology 2009; 30(2): 338-343.

29. Conijn MMA, Geerlings MI, Biessels GJ, Takahara T, Witkamp TD, Zwanenburg JJM et al. Cerebral Microbleeds on MR Imaging: Comparison between 1.5 and 7T. American Journal of Neuroradiology 2011; 32(6): 1043-1049.

30. Potter GM, Marlborough FJ, Wardlaw JM. Wide variation in definition, detection, and description of lacunar lesions on imaging. Stroke 2011; 42(2): 359-66.

31. Prins ND, Scheltens P. White matter hyperintensities, cognitive impairment and dementia: an update. Nat Rev Neurol 2015; 11(3): 157-65.

32. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJR. American journal of roentgenology 1987; 149(2): 351-6.

33. Schmidt R, Fazekas F, Kleinert G, Offenbacher H, Gindl K, Payer F et al. Magnetic resonance imaging signal hyperintensities in the deep and subcortical white matter. A comparative study between stroke patients and normal volunteers. Archives of neurology 1992; 49(8): 825-7.

34. Scheltens P, Barkhof F, Leys D, Pruvo JP, Nauta JJP, Vermersch P et al. A SEMIQUANTITATIVE RATING-SCALE FOR THE ASSESSMENT OF SIGNAL HYPERINTENSITIES ON MAGNETIC-RESONANCE-IMAGING. Journal of the neurological sciences 1993; 114(1): 7-12.

35. Maillard P, Delcroix N, Crivello F, Dufouil C, Gicquel S, Joliot M et al. An automated procedure for the assessment of white matter hyperintensities by multispectral (T1, T2, PD) MRI and an evaluation of its between-centre reproducibility based on two large community databases. Neuroradiology 2008; 50(1): 31-42.

36. Samaille T, Fillon L, Cuingnet R, Jouvent E, Chabriat H, Dormont D et al. Contrast-Based Fully Automatic Segmentation of White Matter Hyperintensities: Method and Validation. Plos One 2012; 7(11).

37. Ramirez J, Gibson E, Quddus A, Lobaugh NJ, Feinstein A, Levine B et al. Lesion Explorer: A comprehensive segmentation and parcellation package to obtain regional volumetrics for subcortical hyperintensities and intracranial tissue. Neuroimage 2011; 54(2): 963-973.

38. Yoshita M, Fletcher E, DeCarli C. Current concepts of analysis of cerebral white matter hyperintensities on magnetic resonance imaging. Topics in magnetic resonance imaging : TMRI 2005; 16(6): 399-407.

39. Dyrby TB, Rostrup E, Baare WF, van Straaten EC, Barkhof F, Vrenken H et al. Segmentation of age-related white matter changes in a clinical multi-center study. Neuroimage 2008; 41(2): 335-45.

40. van den Heuvel DM, ten Dam VH, de Craen AJ, Admiraal-Behloul F, van Es AC, Palm WM et al. Measuring longitudinal white matter changes: comparison of a visual rating scale with a volumetric measurement. AJNR. American journal of neuroradiology 2006; 27(4): 875-8.

41. Wardlaw JM, Brindle W, Casado AM, Shuler K, Henderson M, Thomas B et al. A systematic review of the utility of 1.5 versus 3 Tesla magnetic resonance brain imaging in clinical practice and research. European radiology 2012; 22(11): 2295-303.

42. Polak P, Magnano C, Zivadinov R, Poloni G. 3D FLAIRED: 3D fluid attenuated inversion recovery for enhanced detection of lesions in multiple sclerosis. Magnetic Resonance in Medicine 2012; 68(3): 874-881.

43. Ramirez J, Berezuk C, McNeely AA, Scott CJ, Gao F, Black SE. Visible Virchow-Robin spaces on magnetic resonance imaging of Alzheimer's disease patients and normal elderly from the Sunnybrook Dementia Study. Journal of Alzheimer's disease : JAD 2015; 43(2): 415-24.

44. Garcia-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med Image Anal 2013; 17(1): 1-18.

45. Filippi M, Horsfield MA, Campi A, Mammi S, Pereira C, Comi G. RESOLUTION-DEPENDENT ESTIMATES OF LESION VOLUMES IN MAGNETIC-RESONANCE-IMAGING STUDIES OF THE BRAIN IN MULTIPLE-SCLEROSIS. Annals of Neurology 1995; 38(5): 749-754.

46. Valdes Hernandez MC, Ghandour D, Gonzales-Castro V, Munoz-Maniega S, Armitage PAA, Wang X et al. On the computational assessment of white matter hyperintensity progression: difficulties in method selection and bias field correction performance on images with significant white matter pathology. Neuroradiology in press.

47. Wang Y, Catindig JA, Hilal S, Soon HW, Ting E, Wong TY et al. Multi-stage segmentation of white matter hyperintensity, cortical and lacunar infarcts. NeuroImage 2012; 60(4): 2379-2388.

48. Hervé D, Godin O, Dufouil C, Viswanathan A, Jouvent E, Pachaï C et al. Three-dimensional MRI analysis of individual volume of Lacunes in CADASIL. Stroke; a journal of cerebral circulation 2009; 40(1): 124-128.

49. Viswanathan A, Guichard JP, Gschwendtner A, Buffon F, Cumurcuic R, Boutron C et al. Blood pressure and haemoglobin A1c are associated with microhaemorrhage in CADASIL: a two-centre cohort study. Brain : a journal of neurology 2006; 129(Pt 9): 2375-83.

50. Benjamin P, Lawrence AJ, Lambert C, Patel B, Chung AW, MacKinnon AD et al. Strategic lacunes and their relationship to cognitive impairment in cerebral small vessel disease. NeuroImage. Clinical 2014; 4: 828-37.

51. Hernandez Mdel C, Piper RJ, Wang X, Deary IJ, Wardlaw JM. Towards the automatic computational assessment of enlarged perivascular spaces on brain magnetic resonance images: a systematic review. Journal of magnetic resonance imaging : JMRI 2013; 38(4): 774-85.

52. Adams HHH, Cavalieri M, Verhaaren BFJ, Bos D, van der Lugt A, Enzinger C et al. Rating Method for Dilated Virchow-Robin Spaces on Magnetic Resonance Imaging. Stroke 2013; 44(6): 1732-+.

53. Potter GM, Chappell FM, Morris Z, Wardlaw JM. Cerebral perivascular spaces visible on magnetic resonance imaging: development of a qualitative rating scale and its observer reliability. Cerebrovascular diseases (Basel, Switzerland) 2015; 39(3-4): 224-31.

54. Wang X, Hernandez MD, Doubal F, Chappell FM, Piper RJ, Deary IJ et al. Development and initial evaluation of a semi-automatic approach to assess perivascular spaces on conventional magnetic resonance images. Journal of neuroscience methods 2015.

55. Zong X, Park SH, Shen D, Lin W. Visualization of perivascular spaces in the human brain at 7T: Sequence optimization and morphology characterization. NeuroImage 2015.

56. Berezuk C, Ramirez J, Gao F, Scott CJ, Huroy M, Swartz RH et al. Virchow-Robin Spaces: Correlations with Polysomnography-Derived Sleep Parameters. Sleep 2015; 38(6): 853-8.

57. Greenberg SM, Vernooij MW, Cordonnier C, Viswanathan A, Al-Shahi Salman R, Warach S et al. Cerebral microbleeds: a guide to detection and interpretation. Lancet neurology 2009; 8(2): 165-174.

58. de Laat KF, van den Berg HA, van Norden AG, Gons RA, Olde Rikkert MG, de Leeuw FE. Microbleeds are independently related to gait disturbances in elderly individuals with cerebral small vessel disease. Stroke 2011; 42(2): 494-7.

59. van Norden AG, van Uden IW, de Laat KF, Gons RA, Kessels RP, van Dijk EJ et al. Cerebral microbleeds are related to subjective cognitive failures: the RUN DMC study. Neurobiol Aging 2013; 34(9): 2225-30.

60. Pettersen JA, Sathiyamoorthy G, Gao FQ, Szilagyi G, Nadkarni NK, St George-Hyslop P et al. Microbleed topography, leukoaraiosis, and cognition in probable Alzheimer disease from the Sunnybrook dementia study. Archives of neurology 2008; 65(6): 790-5.

61. Poels MM, Ikram MA, van der Lugt A, Hofman A, Niessen WJ, Krestin GP et al. Cerebral microbleeds are associated with worse cognitive function: the Rotterdam Scan Study. Neurology 2012; 78(5): 326-33.

62. Gregoire SM, Scheffler G, Jager HR, Yousry TA, Brown MM, Kallis C et al. Strictly lobar microbleeds are associated with executive impairment in patients with ischemic stroke or transient ischemic attack. Stroke; a journal of cerebral circulation 2013; 44(5): 1267-72.

63. Cordonnier C, Potter GM, Jackson CA, Doubal F, Keir S, Sudlow CLM et al. Improving Interrater Agreement About Brain Microbleeds Development of the Brain Observer MicroBleed Scale (BOMBS). Stroke 2009; 40(1): 94-99.

64. Gregoire SM, Chaudhary UJ, Brown MM, Yousry TA, Kallis C, Jager HR et al. The Microbleed Anatomical Rating Scale (MARS) Reliability of a tool to map brain microbleeds. Neurology 2009; 73(21): 1759-1766.

65. Barnes SRS, Haacke EM, Ayaz M, Boikov AS, Kirsch W, Kido D. Semiautomated detection of cerebral microbleeds in magnetic resonance images. Magnetic Resonance Imaging 2011; 29(6): 844-852.

66. Seghier ML, Kolanko MA, Leff AP, Jager HR, Gregoire SM, Werring DJ. Microbleed Detection Using Automated Segmentation (MIDAS): A New Method Applicable to Standard Clinical MR Images. Plos One 2011; 6(3).

67. Kuijf HJ, de Bresser J, Geerlings MI, Conijn MMA, Viergever MA, Biessels GJ et al. Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform. Neuroimage 2012; 59(3): 2266-2273.

68. Kuijf HJ, Brundel M, de Bresser J, van Veluw SJ, Heringa SM, Viergever MA et al. Semi-Automated Detection of Cerebral Microbleeds on 3.0 T MR Images. PLoS One 2013; 8(6): e66610.

69. Fazlollahi A, Meriaudeau F, Giancardo L, Villemagne VL, Rowe CC, Yates P et al. Computer-aided detection of cerebral microbleeds in susceptibility-weighted imaging. Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society 2015.

70. Schrag M, McAuley G, Pomakian J, Jiffry A, Tung S, Mueller C et al. Correlation of hypointensities in susceptibility-weighted images to tissue histology in dementia patients with cerebral amyloid angiopathy: a postmortem MRI study. Acta Neuropathologica 2010; 119(3): 291-302.

71. De Guio F, Benoit-Cattin H, Davenel A. Signal decay due to susceptibility-induced intravoxel dephasing on multiple air-filled cylinders: MRI simulations and experiments. Magma (New York, N.Y.) 2008; 21(4): 261-71.

72. Vernooij MW, Ikram MA, Wielopolski PA, Krestin GP, Breteler MMB, van der Lugt A. Cerebral microbleeds: Accelerated 3D T2*-weighted GRE MR imaging versus conventional 2D T2*-weighted GRE MR imaging for detection. Radiology 2008; 248(1): 272-277.

73. Haacke EM, Xu YB, Cheng YCN, Reichenbach JR. Susceptibility weighted imaging (SWI). Magnetic Resonance in Medicine 2004; 52(3): 612-618.

74. Guo LF, Wang G, Zhu XY, Liu C, Cui L. Comparison of ESWAN, SWI-SPGR, and 2D T2*-weighted GRE sequence for depicting cerebral microbleeds. Clinical neuroradiology 2013; 23(2): 121-7.

75. Cheng AL, Batool S, McCreary CR, Lauzon ML, Frayne R, Goyal M et al. Susceptibility-weighted imaging is more reliable than T2*-weighted gradient-recalled echo MRI for detecting microbleeds. Stroke; a journal of cerebral circulation 2013; 44(10): 2782-6.

76. Goos JDC, van der Flier WM, Knol DL, Pouwels PJW, Scheltens P, Barkhof F et al. Clinical Relevance of Improved Microbleed Detection by Susceptibility-Weighted Magnetic Resonance Imaging. Stroke 2011; 42(7): 1894-1900.

77. Charidimou A, Jager HR, Werring DJ. Cerebral microbleed detection and mapping: principles, methodological aspects and rationale in vascular dementia. Experimental gerontology 2012; 47(11): 843-52.

78. Gregoire SM, Werring DJ, Chaudhary UJ, Thornton JS, Brown MM, Yousry TA et al. Choice of echo time on GRE T2*-weighted MRI influences the classification of brain microbleeds. Clinical Radiology 2010; 65(5): 391-394.

79. Liu T, Surapaneni K, Lou M, Cheng LQ, Spincemaille P, Wang Y. Cerebral Microbleeds: Burden Assessment by Using Quantitative Susceptibility Mapping. Radiology 2012; 262(1): 269-278.

80. Peters N, Holtmannspötter M, Opherk C, Gschwendtner A, Herzog J, Sämann P et al. Brain volume changes in CADASIL: a serial MRI study in pure subcortical ischemic vascular disease. Neurology 2006; 66(10): 1517-1522.

81. Jouvent E, Viswanathan A, Mangin J-F, O' sullivan M, Guichard J-P, Gschwendtner A et al. Brain atrophy is related to lacunar lesions and tissue microstructural changes in CADASIL. Stroke; a journal of cerebral circulation 2007; 38(6): 1786-1790.

82. Viswanathan A, Godin O, Jouvent E, O'Sullivan M, Gschwendtner A, Peters N et al. Impact of MRI markers in subcortical vascular dementia: a multi-modal analysis in CADASIL. Neurobiology of Aging 2010; 31(9): 1629-1636.

83. Lawrence AJ, Patel B, Morris RG, MacKinnon AD, Rich PM, Barrick TR et al. Mechanisms of cognitive impairment in cerebral small vessel disease: multimodal MRI results from the St George's cognition and neuroimaging in stroke (SCANS) study. PloS one 2013; 8(4): e61014.

84. Jokinen H, Lipsanen J, Schmidt R, Fazekas F, Gouw AA, van der Flier WM et al. Brain atrophy accelerates cognitive decline in cerebral small vessel disease: the LADIS study. Neurology 2012; 78(22): 1785-92.

85. Jouvent E, Viswanathan A, Chabriat H. Cerebral atrophy in cerebrovascular disorders. Journal of neuroimaging : official journal of the American Society of Neuroimaging 2010; 20(3): 213-218.

86. O'sullivan M, Jouvent E, Saemann PG, Mangin J-F, Viswanathan A, Gschwendtner A et al. Measurement of brain atrophy in subcortical vascular disease: a comparison of different approaches and the impact of ischaemic lesions. NeuroImage 2008; 43(2): 312-320.

87. Kloppenborg RP, Nederkoorn PJ, Grool AM, Vincken KL, Mali WP, Vermeulen M et al. Cerebral small-vessel disease and progression of brain atrophy: the SMART-MR study. Neurology 2012; 79(20): 2029-36.

88. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis - I. Segmentation and surface reconstruction. Neuroimage 1999; 9(2): 179-194.

89. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis - II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 1999; 9(2): 195-207.

90. Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005; 26(3): 839-851.

91. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004; 23: S208-S219.

92. Mangin J-F, Rivière D, Cachia A, Duchesnay E, Cointepas Y, Papadopoulos-Orfanos D et al. A framework to study the cortical folding patterns. NeuroImage 2004; 23 Suppl 1: S129-38.

93. Huang L, Wang X, Baliki MN, Wang L, Apkarian AV, Parrish TB. Reproducibility of structural, resting-state BOLD and DTI data between identical scanners. PLoS One 2012; 7(10): e47684.

94. Gedamu EL, Collins DL, Arnold DL. Automated quality control of brain MR images. Journal of magnetic resonance imaging : JMRI 2008; 28(2): 308-19.

95. Patwardhan AJ, Eliez S, Warsofsky IS, Glover GH, White CD, Giedd JN et al. Effects of image orientation on the comparability of pediatric brain volumes using three-dimensional MR data. J Comput Assist Tomogr 2001; 25(3): 452-7.

96. Duning T, Kloska S, Steinstrater O, Kugel H, Heindel W, Knecht S. Dehydration confounds the assessment of brain atrophy. Neurology 2005; 64(3): 548-50.

97. Nakamura K, Brown RA, Narayanan S, Collins DL, Arnold DL. Diurnal fluctuations in brain volume: Statistical analyses of MRI from large populations. Neuroimage 2015.

98. Heye AK, Thrippleton MJ, Chappell FM, Valdes Hernandez MD, Armitage PA, Makin SD et al. Blood pressure and sodium: association with MRI markers in cerebral small vessel disease. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism 2015.

99. Wardlaw JM, Valdes Hernandez MC, Munoz-Maniega S. What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. Journal of the American Heart Association 2015; 4(6): 001140.

100. Krueger G, Granziera C, Jack CR, Gunter JL, Littmann A, Mortamet B et al. Effects of MRI scan acceleration on brain volume measurement consistency. Journal of Magnetic Resonance Imaging 2012; 36(5): 1234-1240.

101. Ross DE, Ochs AL, DeSmit ME, Seabaugh JM, Havranek MD. Man Versus Machine Part 2: Comparison of Radiologists' Interpretations and NeuroQuant Measures of Brain Asymmetry and Progressive Atrophy in Patients With Traumatic Brain Injury. The Journal of neuropsychiatry and clinical neurosciences 2015; 27(2): 147-52.

102. Ritchie SJ, Dickie DA, Cox SR, Valdes Hernandez MdC, Corley J, Royle NA et al. Brain volumetric changes and cognitive ageing during the eighth decade of life.

103. Peters N, Holtmannspotter M, Opherk C, Gschwendtner A, Herzog J, Samann P et al. Brain volume changes in CADASIL: a serial MRI study in pure subcortical ischemic vascular disease. Neurology 2006; 66(10): 1517-22.

104. Smith SM, de Stefano N, Jenkinson M, Matthews PM. Normalized accurate measurement of longitudinal brain change. Journal of computer assisted tomography 2001; 25(3): 466-475.

105. Smith SM, Zhang YY, Jenkinson M, Chen J, Matthews PM, Federico A et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 2002; 17(1): 479-489.

106. Freeborough PA, Fox NC. The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI. Ieee Transactions on Medical Imaging 1997; 16(5): 623-629.

107. Takao H, Hayashi N, Ohtomo K. Effect of Scanner in Longitudinal Studies of Brain Volume Changes. Journal of Magnetic Resonance Imaging 2011; 34(2): 438-444.

108. Dickie DA, Job DE, Gonzalez DR, Shenkin SD, Wardlaw JM. Use of brain MRI atlases to determine boundaries of age-related pathology: the importance of statistical method. PloS one 2015; 10(5): e0127939.

109. O'Brien JT, Paling S, Barber R, Williams ED, Ballard C, McKeith IG et al. Progressive brain atrophy on serial MRI in dementia with Lewy bodies, AD, and vascular dementia. Neurology 2001; 56(10): 1386-8.

110. Cover KS, van Schijndel RA, Popescu V, van Dijk BW, Redolfi A, Knol DL et al. The SIENA/FSL whole brain atrophy algorithm is no more reproducible at 3T than 1.5 T for Alzheimer's disease. Psychiatry research 2014; 224(1): 14-21.

111. Preboske GM, Gunter JL, Ward CP, Jack CR. Common MRI acquisition non-idealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. Neuroimage 2006; 30(4): 1196-1202.

112. Caramanos Z, Fonov VS, Francis SJ, Narayanan S, Pike GB, Collins DL et al. Gradient distortions in MRI: Characterizing and correcting for their effects on SIENA-generated measures of brain volume change. Neuroimage 2010; 49(2): 1601-1611.

113. Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R et al. Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. Neuroimage 2006; 30(2): 436-443.

114. Tadic T, Jaffray DA, Stanescu T. Harmonic analysis for the characterization and correction of geometric distortion in MRI. Medical physics 2014; 41(11): 112303.

115. Maikusa N, Yamashita F, Tanaka K, Abe O, Kawaguchi A, Kabasawa H et al. Improved volumetric measurement of brain structure with a distortion correction procedure using an ADNI phantom. Medical physics 2013; 40(6): 062303.

116. Davids M, Zollner FG, Ruttorf M, Nees F, Flor H, Schumann G et al. Fully-automated quality assurance in multi-center studies using MRI phantom measurements. Magn Reson Imaging 2014; 32(6): 771-80.

117. Takao H, Abe O, Hayashi N, Kabasawa H, Ohtomo K. Effects of Gradient Non-Linearity Correction and Intensity Non-Uniformity Correction in Longitudinal Studies Using Structural Image Evaluation Using Normalization of Atrophy (SIENA). Journal of Magnetic Resonance Imaging 2010; 32(2): 489-492.

118. Pfefferbaum A, Rohlfing T, Rosenbloom MJ, Sullivan EV. Combining atlas-based parcellation of regional brain data acquired across scanners at 1.5 T and 3.0 T field strengths. Neuroimage 2012; 60(2): 940-951.

119. Chua AS, Egorova S, Anderson MC, Polgar-Turcsanyi M, Chitnis T, Weiner HL et al. Handling changes in MRI acquisition parameters in modeling whole brain lesion volume and atrophy data in multiple sclerosis subjects: Comparison of linear mixed-effect models. NeuroImage. Clinical 2015; 8: 606-10.

120. Pardoe HR, Cutter GR, Alter R, Hiess RK, Semmelroch M, Parker D et al. Pooling Morphometric Estimates: A Statistical Equivalence Approach. Journal of neuroimaging : official journal of the American Society of Neuroimaging 2015.

121. Moorhead TW, Gountouna VE, Job DE, McIntosh AM, Romaniuk L, Lymer GK et al. Prospective multi-centre Voxel Based Morphometry study employing scanner specific segmentations: procedure development using CaliBrain structural MRI data. BMC Med Imaging 2009; 9: 8.

122. Mortamet B, Bernstein MA, Jack CR, Jr., Gunter JL, Ward C, Britson PJ et al. Automatic quality assessment in structural brain magnetic resonance imaging. Magnetic resonance in medicine 2009; 62(2): 365-72.

Titles and legends to figures

Figure 1 title: Schematic view of the different sources of variability in a multicentre MRI study (adapted from www.imaios.com)

Figure 2 title: Enhanced lesion detection in 3D FLAIR sequence (courtesy of Polak et al.[42])

Figure 2 legend: Sagittal slices from (**a**, **d**) 3D FLAIRED (a FLAIR sequence optimized in order to enhance detection of lesions); (**b**, **e**) GE 3D FLAIR; and (**c**, **f**) 2D FLAIR. Note increased lesion detection in the frontal lobe (**1**) deep white matter (**2**) and juxtacortical (**3**) regions with 3D FLAIRED. Improved deep white matter lesion (**4**) detection and resolution in 3D FLAIRED compared to other sequences. Cerebellar lesion (**5**) detected in 3D FLAIRED and GE 3D FLAIR, but uncertainly discerned in 2D FLAIR.

Figure 3title: Effect of slice thickness, sequence and magnetic field on CMB detection (courtesy of Nandigam et al.[28])

Figure 3 legend: The pairs of images illustrate comparisons of thick-section GRE (A) versus thin-section SWI (B); thin-section GRE (C) versus thin-section SWI (D); thick-section GRE (E) versus thin-section GRE (F) (all preceding images at 1.5T); and SWI at 1.5T (G) versus SWI at 3T (H). The black arrows in (A) and (B) illustrate a CMB prospectively counted on both sequences, whereas lesions denoted by white arrows were initially identified only on the SWI image. The black arrows in the remaining images highlight lesions on the paired images for comparison.

Figure 4 title: Typical gradient distortion field in a T1-weighted acquisition (courtesy of Caramanos et al. [112])

Figure 4 legend: Sagittal, coronal, and axial views of a phantom-based gradient distortion field that was determined based on a T1-weighted acquisition. The color scale represents the distance (in millimeters) that a voxel moves because of gradient distortion between its “real” location and its “apparent” location on such an MRI scan that has not been corrected for the gradient distortion.

Table 1. Summary of reproducibility studies (within-centre or/and between centre scan-rescan procedure)

Study			Between centre		Within centre					
First author	Yearpublished	Marker of interest	Nb subjectsrescanned	Nb of MRI sites	Nb subjectsrescanned	Nb rescans per subject	Magneticfield	Analysistool	Between-centre reproducibility	Within-centre reproducibility
Brain volumetry										
Schnack	2004	ICV, BV, GMV, WMV	6	4	5	2	1.5T, 1T	own segmentation	BV: ICC = 0.98	BV: ICC = 1
Ewers	2006	GMV, WMV, CSFV	1	10	NA	NA	1.5T	SPM2	CoV=5% for all volumes	NA
Jovicich	2009	ICV	15	3	15	2	1.5T, 3T	Freesurfer	NA	Reproducibilityerror=2.56%
Reig	2009	ICV, GMV, WMV, CSFV	5	5	NA	NA	1.5T	SPM-based	ICV: CoV=1.9%, ICC=0.99	NA
Wonderlick	2009	BV, GMV, WMV	11	2	NA	NA	3T	Freesurfer	BV: ICC=0.99	NA
De Boer	2010	BV, GMV, WMV, CSFV	NA	NA	30	2	1.5T	FSL, SPM5, kNN	NA	BV: CoV=0.39% with FSL, CoV=0.72% with SPM5
Huppertz	2010	BV, GMV, WMV, CSFV, ICV	1	6	1	3	1.5T, 3T	SPM5	BV: CoV=3.78%; ICV: CoV=2.97%	BV: CoV=0.5%; ICV: CoV=0.6%
Kruggel	2010	ICV, BV, GMV, WMV	172	59	41	2	1.5T, 3T	Fantasm	BV: median RAVD=4.76%, 90^{th} percentile=12.34%; ICV: median RAVD=1.66%, 90th percentile=3.92%	BV: median RAVD=0.30%, 90^{th} percentile RAVD=1.09%; ICV: median RAVD=0.29%, 90th percentile RAVD=1.27%
Sampat	2010	BPF	NA	NA	18	2	1.5T	own segmentation	NA	BPF: CoV=0.46%
Cover	2011	Normalizedbrain volume	NA	NA	671	2	1.5T	SIENAX	NA	BV: median RAVD=0.96%, 90th percentile RAVD=5.11%
Shokouhi (1.5T)	2011	ICV, GMV, WMV, CSFV	13	3	13	2	1.5T	BrainVISA	BV: reliability[a]=0.97	BV: reliability=0.99
Shokouhi (3T)	2011	ICV, GMV, WMV, CSFV	11	5	11	2	3T	BrainVISA	BV: reliability=0.88	BV: reliability=0.96
Landman	2011	GMW, WMV	NA	NA	21	2	3T	Java Imaging Science Toolkit	NA	GMV: CoV=2.6%; WMV: CoV=1.5%
Chalavi	2012	GMW, WMV	3	2	3	2	3T	Freesurfer, SPM	NA	NA
Huang	2012	WMV, CSFV	6	2	NA	NA	3T	FSL	WMV: ICC=0.62; CSFV: ICC=0.87	NA
Pfefferbaum	2012	Supratentorial volume	114	2	NA	NA	1.5T, 3T	Freesurfer	BV: ICC=0.81	NA

Brainatrophy										
Smith	2007	PBVC	NA	NA	68	2	1.5T	SIENA, BSI	NA	SIENA: median RAVD=0.16%; BSI: median RAVD=0.17%
De Bresser	2011	PBVC	NA	NA	10	2	1.5T	SIENA, SPM8, kNN	NA	coefficient of repeatability[b]=0.92%for SIENA, 2.64% for SPM8, 0.31% for kNN
Cover	2011	PBVC	NA	NA	385	2	1.5T	SIENA	NA	PBVC: median RAVD=0.35%, 90th percentile RAVD=1.37%
Durand-dubief	2012	Atrophy	9	2	NA	NA	1.5T	7 segmentation algorithms	SEM: 1.77 for Freesurfer, 0.98 for BSI, 0.85 for SIENA	NA
White-matterhypenintensities										
Wei	2002	WMH volume	NA	NA	20	2	1.5T	3 pipelines: PVEC, TDS, TDS+	NA	CoV range: 2.6 - 7.5% according to pipeline
Sicotte	2003	WMH volume	25	2	NA	NA	1.5T, 3T	threshold-based segmentation	median RAVD: 12.5%	NA
De Boer	2010	WMH volume	NA	NA	30	2	1.5T	auto trainedkNN	NA	CoV: 5.87%
Theysohn	2011	WMH grading	20	2	NA	NA	1.5T, 7T		75% of subjects with equal grading at 7T compared to 1.5T, 13% superior and 13% inferior	NA
Ramirez	2013	WMH volume (total subcortical, deep white, periventricular)	NA	NA	20	2	1.5T	Lesion Explorer	NA	ICC=1; mean AVD in 10 older subjects=213mm^3
Cerebral microbleeds										
Stehling	2008	CMBnumber	25	2	NA	NA	1.5T, 3T		more CMB at 3T vs 1.5T	NA
Nandigam	2009	CMB contrast index and diameter	4	2	NA	NA	1.5T, 3T		smaller contrast index and diameter at 1.5T vs 3T	NA
Conijn	2011	CMBnumber	34	2	NA	NA	1.5T, 7T		more CMB at 7T vs 1.5T	NA
Theysohn	2011	CMBnumber	20	2	NA	NA	1.5T, 7T		more CMB at 7T vs 1.5T	NA

ICV: intracranial volume, BV: brain volume (GMV + WMV), GMV: gray matter volume, WMV: white matter volume, CSFV: cerebrospinal volume, BPF: Brain parenchymal fraction, WMH: white matter hyperintensities, CMB: cerebral microbleeds, ICC: Intraclass correlation coefficient, CoV: coefficient of variation, NA: not available, Reproducibility error: The group reproducibility error for each structure is derived averaging the reproducibility errors across subjects, where for each subject the error is estimated as the absolute test–retest volume percent change relative to the mean test–retest volume, kNN: k-Nearest-Neighbors segmentation method, RAVD: relative absolute value of the difference, PBVC: percent brain volume change, [a]reliability in this study has been computed as the ratio of the variance excluding the contribution from the factor (within-centre or between-centre) to the total variance, [b]coefficient of repeatability: 1.96 times the standard deviation of the differences between two measurements (Bland-Altman), SEM: standard error of measurement, AVD: absolute value of the difference.