NeuroImage 142 (2016) 188–197

Contents lists available at ScienceDirect

NeuroImage

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

Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis

Viola Biberacher [a,b,*,1], Paul Schmidt [b,c,1], Anisha Keshavan [d], Christine C. Boucard [a,b], Ruthger Righart [a,b], Philipp Sämann [e], Christine Preibisch [f], Daniel Fröbel [f], Lilian Aly [a], Bernhard Hemmer [a,g], Claus Zimmer [f], Roland G. Henry [d], Mark Mühlau [a,b]

[a] Neurology, Technische Universität München, Ismaninger Str. 22, 81675 Munich, Germany
[b] TUM-Neuroimaging Center, Technische Universität München, Munich, Germany
[c] Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 Munich, Germany
[d] Neurology, University of California San Francisco, 675 Nelson Rising Lane, San Francisco, CA 94158, United States
[e] Neuroimaging Core Unit, Max Planck Institute of Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany
[f] Neuroradiology, Technische Universität München, Ismaninger Str. 22, 81675 Munich, Germany
[g] Munich Cluster for Systems Neurology (SyNergy), Feodor-Lynen-Str. 17, 81377 Munich, Germany

ARTICLE INFO

Article history:
Received 8 January 2016
Accepted 14 July 2016
Available online 16 July 2016

Keywords:
Multiple sclerosis
Magnetic resonance imaging
Scanner-related variability

ABSTRACT

Brain volumetric measurements in multiple sclerosis (MS) reflect not only disease-specific processes but also other sources of variability. The latter has to be considered especially in multicenter and longitudinal studies. Here, we compare data generated by three different 3-Tesla magnetic resonance scanners (Philips Achieva; Siemens Verio; GE Signa MR750). We scanned two patients diagnosed with relapsing remitting MS six times per scanner within three weeks (T1w and FLAIR, 3D). We assessed T2-hyperintense lesions by an automated lesion segmentation tool and determined volumes of grey matter (GM), white matter (WM) and whole brain (GM + WM) from the lesion-filled T1-weighted images using voxel-based morphometry (SPM8/VBM8) and SIENAX (FSL). We measured cortical thickness using FreeSurfer from both, lesion-filled and original T1-weighted images. We quantified brain volume changes with SIENA.

In both patients, we found significant differences in total lesion volume, global brain tissue volumes and cortical thickness measures between the scanners. Morphometric measures varied remarkably between repeated scans at each scanner, independent of the brain imaging software tool used.

We conclude that for cross-sectional multicenter studies, the effect of different scanners has to be taken into account. For longitudinal monocentric studies, the expected effect size should exceed the size of false positive findings observed in this study. Assuming a physiological loss of brain volume of about 0.3% per year in healthy adult subjects (Good et al., 2001), which may double in MS (De Stefano et al., 2010; De Stefano et al., 2015), with current tools reliable estimation of brain atrophy in individual patients is only possible over periods of several years.

Introduction

Magnetic resonance imaging (MRI) can detect minimal changes of the brain structure. However, uncertainty of the measurement interferes with these subtle structural changes. This is particularly important in multicenter and longitudinal studies.

Sources of variance include:

1) Scanner-related factors, which comprise differences between the measurements acquired with one scanner (Intrascanner variability) as well as differences between scanners (Interscanner variability) (Droby et al., 2015; Han et al., 2006; Huppertz et al., 2010; Jovicich et al., 2009; Shokouhi et al., 2011).

2) Factors related to the scanning procedure as subject positioning (Caramanos et al., 2010) and subject motion (Reuter et al., 2015).

3) Physiological factors like hydration status (Duning et al., 2005; Kempton et al., 2011; Kempton et al., 2009; Walters et al., 2001), diurnal fluctuations of brain volume (Nakamura et al., 2015) or brain volume fluctuations dependent on the menstrual cycle (Hagemann et al., 2011).

4) Factors related to data-processing: Manual brain MRI analysis is time-consuming and might be biased by intra- and inter-rater variability. Therefore, various tools for automated assessment of brain MRI have been developed. However, differences between the software tools have

* Corresponding author at: Department of Neurology, Technische Universität München, Ismaningerstr. 22, D-81675 Munich, Germany.
E-mail address: viola.biberacher@tum.de (V. Biberacher).
[1] Contributed equally to the manuscript.

http://dx.doi.org/10.1016/j.neuroimage.2016.07.035
1053-8119/

been described as well (Derakhshan et al., 2010; Durand-Dubief et al., 2012; Jovicich et al., 2009; Shokouhi et al., 2011).

In healthy controls (HC), various studies demonstrated significant differences in brain tissue volumes between scanners (Moorhead et al., 2009; Shokouhi et al., 2011; Suckling et al., 2012). Regional variability was especially high in deep grey matter (GM), the internal capsule, the cingulate gyrus and the cerebral cortex (Droby et al., 2015; Huppertz et al., 2010; Jovicich et al., 2009; Moorhead et al., 2009; Schnack et al., 2010; Suckling et al., 2012).

Multiple sclerosis (MS) is a chronic immune-mediated disease of the central nervous system. The pathological hallmarks are focal white matter (WM) lesions in the brain and spinal cord. However, GM atrophy seems to contribute to clinical deficits in very early stages (Geurts and Barkhof, 2008). The course of MS and the response to immunomodulatory treatment is highly variable (Compston and Coles, 2008). Therefore, analysis of large datasets is required to understand its pathophysiology and improve the treatment for the individual patient. This often requires pooling of data from several MS centers. Longitudinal MRI studies are important to monitor subclinical disease activity and response to immunomodulatory treatment. Despite the broad practice of multicenter and longitudinal MRI studies in MS, intrascanner in relation to interscanner variability has not yet been investigated in MS patients. In addition to the sources of variance listed above, disease-related effects beyond the object of investigation have to be considered: MS lesions and diffuse brain tissue changes disturb automated image analysis (Chard et al., 2002) and add additional variance (Sampat et al., 2010). The disease might cause partially opposing effects on measurements with inflammation leading to volume increase as well as neurodegeneration and demyelination leading to volume decrease. Immunomodulatory treatment can produce “pseudoatrophy” by reducing inflammation, but also decelerates atrophy rates on the long term (Zivadinov et al., 2008). Finally, scan quality might decrease to a certain degree with increasing disability (e.g. motion artefacts). Therefore, it is likely that variability of MRI measurement is higher in MS patients than in HC.

In our study, we aimed to quantify the interscanner and intrascanner variability of brain MRI that has to be expected in multicenter and longitudinal MRI studies in MS. As classical scan-rescan experiments do not cover all sources of variance and therefore underestimate the true error, we chose an interval of several days between the scans at one scanner. We hereby compromised between covering all sources of variability with a minimal risk of the appearance of new MS lesions or true atrophy during the study period. As precise adaption of MRI sequences across centers has proven to be unpractical over the last years, we refrained from such an attempt. Instead, we decided to compare repetitive scans of three 3-Tesla scanners adhering to the standards of the German Competence Network Multiple Sclerosis (http://www.kompetenznetz-multiplesklerose.de/en). We analyzed the images by several commonly applied automated brain imaging software tools in two independent laboratories (University of California San Francisco = UCSF and Technische Universität München = TUM). We focused on two main MRI-parameters in MS, namely the volume and distribution of T2-hyperintense WM lesions as well as the volume of the brain and its subcompartments. Both represent commonly used secondary endpoints in pivotal clinical trials in MS with longitudinal and multicenter design (Sormani et al., 2014).

Materials and methods

Subjects

This study was performed in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans and was approved by the local ethics committee. Inclusion criteria were a diagnosis of MS (Polman et al., 2011), evidence of at least 10 MS lesions in a preceding brain MRI performed in regular clinical routine, and a good general health condition to undergo three MRI scans in 1 day. Exclusion criteria were contraindications for an MRI scan. We first analyzed intra- and interscanner variability in one MS patient (MS1). During the revision process, we repeated our experiment with a second MS patient (MS2). Both patients were recruited from the outpatient MS clinic of the Department of Neurology of Technische Universität München. Written informed consent was obtained. MS1 was female, 29 years old, had a relapsing remitting disease course (RRMS), a disease duration of 5 years and an EDSS of 1.0. She was under therapy with intramuscular interferon beta 1a once a week (every Tuesday evening) for 14 months and had no relapse in the last 12 months. MS2 was female, 24 years old, had RRMS, a disease duration of 5.5 years and an EDSS of 2. She was under therapy with natalizumab since 26 months and had no relapse in the last 27 months. Natalizumab was given 5 days before the first MRI scan and 6 days after the last MRI scan. EDSS was performed before and after the study and remained unchanged during the study period. Both patients did not experience a relapse during the study period.

Magnetic resonance imaging

We compared three different 3-Tesla MRI scanner: Philips Achieva, Siemens Verio (both Klinikum rechts der Isar, Technische Universität München), and General Electric (GE) Signa MR750 (Max Planck Institute of Psychiatry München). Within three weeks, the subjects were scanned six times at all three scanners with an interval of several days between the scans. With one exception (replication analysis, 04-15/16-2016), patients were scanned at all three scanners in 1 day. The scan times and order of scans were pseudo-randomized. Scans of MS1 were acquired in October 2013 (10-07-2013, 10-10-2013, 10-14-2013, 10-17-2013, 10-21-2013, 10-24-2013). Scans of MS2 were acquired in April 2016 (04-04-2016, 04-07-2016, 04-11-2016, 04-15-2016 (GE)/04-16-2016 (Siemens/Philips), 04-18-2016, 04-21-2016. Additionally, we performed repositioning experiments (4 consecutive scans) with MS2 at the GE scanner (04-15-2016). The scanning protocol included a 3D GRE T1-weighted sequence and a 3D FLAIR sequence that had already been established by the local staff for use in clinical routine (Table 1). Patient positioning was done by different technicians according to the guidelines in clinical routine (Frankfort horizontal plane). Two scans of MS1 were excluded from final data analysis due to scanning protocol violations (Siemens 2013-10-21 and Philips 2013-10-24).

Data processing

Lesion segmentation and lesion filling

T2-hyperintense lesions were segmented from FLAIR and T1-weighted images by a lesion growth algorithm (Schmidt et al., 2012) as implemented in the lesion segmentation tool (LST) toolbox version 1.2.3 (www.statistical-modelling.de/lst.html) for SPM8 (http://www.fil.ion.ucl.ac.uk/spm). For MR images of MS1, lesions were segmented using different initial thresholds (κ) for the lesion belief map (Fig. S1). Then, the optimal threshold was determined separately for each scanner by visual inspection by three different experienced raters in a two-step procedure. First, one rater set κ for one scanner being blinded for the images of the other scanners. Second, the three raters cross-checked their choices to avoid the introduction of a rater-bias to scanner-specific κ. This second step led to only minimal changes in the scanner-specific κ values. The same κ values were used for lesion segmentation of MR images of MS2.

Lesions were filled with normal appearing WM intensities in the T1-weighted images by LST (Ceccarelli et al., 2012; Chard et al., 2010; Sdika and Pelletier, 2009).

Table 1
Scanning protocols.

	GE		Philips		Siemens	
	T1w (BRAVO)	FLAIR	T1w (MPRAGE)	FLAIR	T1w (MPRAGE)	FLAIR
Orientation	Sagittal	Sagittal	Sagittal	Axial	Sagittal	Sagittal
Number of slices	160	160	170	144	176	160
Slice thickness (mm)	1	1	1	1.5	1	1
Field of view (mm)	256 × 256	256 × 256	240 × 240	230 × 185	270 × 270	250 × 250
Voxel size (mm)	1 × 1 × 1	1 × 1 × 1	1 × 1 × 1	1 × 1 × 1.5	1.1 × 1.1 × 1	1 × 1 × 1
Repetition time (ms)	8.2	7500	9	10,000	<9	6000
Echo time (ms)	3.2	118	4	140	2.45	395
Inversion time (ms)	450	2173	1000	2750	900	2100
Shot interval (ms)	n.a.	n.a.	3000	n.a.	1900	n.a.

GE, general electric.

Estimation of global brain tissue volumes

Global volumes of GM, WM and whole brain were estimated from the lesion filled T1-weighted images of MS1 by VBM8's longitudinal pipeline (TUM) and FSL's SIENAX pipeline (UCSF).

The longitudinal pipeline of VBM8 proceeds as follows: After an initial realignment, the mean of the images of all time points is calculated and used as a reference image in a subsequent realignment step. The coregistered images are then corrected for signal inhomogeneities (correction of bias-field). Next, spatial normalization parameters are estimated from the mean image. These normalization parameters are applied to the segmentations of the bias-corrected images. After segmentation, this pipeline gives the volumes of the three main tissue compartments (CSF, GM, WM) from which we calculated the global brain volume (GM + WM).

SIENAX starts by extracting brain and skull images from the single whole-head input data by FSL's brain extraction tool (BET) (Smith, 2002). However, BET produced considerable misclassification of brain tissue as described before (Boesen et al., 2004; Fennema-Notestine et al., 2006; Leung et al., 2011; Popescu et al., 2012; Vrenken et al., 2013). Therefore, we used a modified workflow: BET was disabled in the original pipeline and SIENAX was provided with skull-stripped images obtained by FreeSurfer (see below), skull masks from BET, and original input images. Details of the modified workflow are available online: http://github.com/akeshavan/munich/tree/gh-pages/scripts. SIENAX was then applied as usual: the brain image is affine-registered to MNI152 space using the skull image to determine the registration scaling (Jenkinson et al., 2002; Jenkinson and Smith, 2001); the resulting volumetric scaling factor is primarily used to normalise for head size. Next, tissue-type segmentation with partial volume estimation is carried out (Zhang et al., 2001) in order to calculate total volumes of brain tissues (including separate estimates of volumes of GM and WM).

In the replication analysis, volumetric brain measures were determined only by VBM8 as described above.

Brain volume changes

Two-time-point percentage brain volume change (PBVC) was estimated with SIENA (Smith et al., 2001; Smith et al., 2002a), part of FSL (Smith et al., 2004). Although WM lesions do not seem to disturb registration-based methods such as SIENA (Battaglini et al., 2012), we used the lesion-filled images for better comparability with the results of SIENAX and VBM/SPM. In MS1, the two institutions involved in the project (UCSF and TUM), performed SIENA analysis independently. SIENA starts by extracting brain and skull images from the two-time-point whole-head input data (Smith, 2002). Similar to SIENAX, results of brain extraction was not satisfactory. Therefore, a similar modification of the pipeline as described above was used: BET was disabled; instead, members of both institutions followed their preferences: the TUM group used the PVE labels from VBM8 for brain segmentation, while the UCSF group used FreeSurfer's aparc + aseg segmentation after dilating it by one voxel. The pipeline was then followed as usual: The two brain images are aligned to each other (Jenkinson et al., 2002; Jenkinson and Smith, 2001) using the skull images to constrain the registration scaling; both brain images are resampled into the space halfway between the two. Next, tissue-type segmentation is carried out (Zhang et al., 2001) in order to find brain/non-brain edge points, and then perpendicular edge displacement (between the two time points) is estimated at these edge points. Finally, the mean edge displacement is converted into a (global) estimate of PBVC between the two time points.

In the replication analysis, SIENA's PBVC was assessed by TUM only as described above.

We additionally estimated brain volume change from the global brain volumes derived with SPM/VBM by calculating the PBVC between all combinations of time points in both patients.

Cortical thickness and its changes

In MS1, analysis with FreeSurfer was performed independently by UCSF and TUM using FreeSurfer 5.3. As lesion filling may influence cortical thickness, analysis was performed once from the lesion-filled (UCSF) and once from the original (TUM) T1-weighted images. The following steps were performed: brain extraction, intensity normalization, automated tissue segmentation, surface-based analysis of cortical thickness, generation of the WM-surface and pial surface, surface topology correction, automated whole brain segmentation and spherical interindividual surface alignment (Dale et al., 1999; Fischl and Dale, 2000). In FreeSurfer, cortical thickness is measured at each vertex of the surface as the shortest distance between the WM and cortical pial surface and varies between 2 and 5 mm across the cerebral cortex. To estimate alterations of cortical thickness across different time points, the data were processed in the longitudinal stream. The procedure is optimized for longitudinal data using an unbiased within-subject template. The longitudinal stream offers significantly increased statistical power compared with cross-sectional processing (Reuter et al., 2012). The change in cortical thickness was calculated from all combinations of time points. In the replication analysis, cortical thickness was determined by TUM only from the original T1-weighted images.

Voxel-wise analysis of grey matter variance

To analyse the degree of interscanner variability across the brain, we performed voxel-wise analysis of GM variance in both patients. Normalized GM maps were estimated by VBM8 (Ashburner, 2007; Bezzola et al., 2011) (http://dbm.neuro.uni-jena.de/vbm8) from the lesion-filled T1-weighted images. The resulting GM images were modulated (Good et al., 2001) accounting only for nonlinear volume changes (Bezzola et al., 2011). Finally, GM images were smoothed with a Gaussian kernel of 8 mm full width at half maximum (default option of SPM8

and VBM8). For all scanners, standard deviation (SD) was calculated in a voxel-wise manner from the smoothed GM maps.

Statistical analysis

For data analysis, the R language for statistical computing (R-Core-Team, 2015) was used. For each scanner, mean, SD and coefficient of variation (CV) were calculated for global volumes and cortical thickness as well as median absolute error, SD, minimum and maximum for PBVC and cortical thickness change. We assessed the effect of scanner on lesion volume, brain volume, brain volume change and cortical thickness by one-way Analyses of variance (ANOVA). Model assumptions were checked by residual plots. When scanners showed different variances, Welch's test for equal means in a one-way layout (Welch, 1951) was used instead of ANOVA. For pairwise comparison of scanners, Tukey's range test (Tukey, 1949) was applied in order to correct for an increase of the Type I error.

We investigated the effect of scan time on brain volume and cortical thickness by a simple regression analysis without and with including the scanner as a factor.

Results

Lesion segmentation

MS1 had periventricular, juxtacortial and infratentorial lesions as well as many small lesions scattered over the whole brain (Fig. 1A). Lesion volume differed significantly between the scanners (all p values <0.0001) and was highest for GE (Fig. 1C left panel and Table 2). This was due to both, bigger lesion size and a higher detection rate for the small, scattered lesions. However, the rate of false positive lesions in the midline was also higher for GE. Analysis of the lesion volume over time suggested a weekly fluctuation of lesion volumes consistent for all scanners (Fig. 1C left panel). Under the assumption of independent lesion loads between different scanners, the probability of observing the same order for the first four scans for all scanners is 0.0017. There are $4! = 24$ possible arrangements for the first four lesion loads; the probability that scanner 2 produces exactly the same order as scanner 1 is 1/24; the same holds for scanner 3; thus, for independent scanners, the probability that lesion loads of scanner 2 and 3 show the same arrangement as scanner 1 is $(1/24) * (1/24) = 0.0017$. We therefore assume that the weekly fluctuation of lesion volume in MS1 is more likely to mirror a biological rather than a technical effect.

MS 2 also showed periventricular, juxtacortical and infratentorial lesions (Fig. 1B). Again, lesion volume differed significantly between the scanners (all p values <0.0001), but was highest in Philips (Fig. 1C, right panel and Table 2). Analysis of lesion volume over time did not show any systematic trend consistent for all scanners (Fig. 1C, right panel). Intrascanner variability of MS lesion volume was considerably lower in the repositioning analysis than over the longer period of three weeks (SD 0.05 ml versus 0.45 ml).

Intra- and interscanner variability of brain volume measures

We assessed global brain tissue volumes with SPM/VBM and SIENAX in MS1 (Fig. 2). Mean, SD and CV for all scanners are given in Table 2. With both methods, we found differences in brain volumes between the scanners. The difference between Philips and the two other scanners was significant for SIENAX ($p < 0.001$ for both comparisons). For SPM/VBM, only the difference between Philips and GE was significant ($p = 0.006$). No significant effect of scan time on brain volume was observed. GM volume differed remarkably between scans and scanners (Fig. 2B; Table 2), while WM volume (Fig. 2C; Table 2) differed between scanners but less within scanners.

Replication analysis (MS2) was performed with SPM/VBM (Table 2 and Fig. S2). Again, we found significant differences in brain volume between the scanners (Philips – GE, $p < 0.001$; Siemens – Philips, $p < 0.001$). No significant effect of scan time on brain volume was found.

Intrascanner variability of volumetric measures was higher over the longer period of 3 weeks than in the repositioning experiment (Table 2, Fig. S2).

Intra- and interscanner variability of cortical thickness measures

In MS1, we assessed cortical thickness twice (Fig. 2D), from lesion-filled (left panel) and original (right panel) T1-weighted images. Mean, SD and CV are given in Table 2. All scanners differed significantly from each other with respect to cortical thickness (all p values <0.02). No significant effect of scan time on cortical thickness was observed. With regard to mean values and variability measures, lesion filling did not introduce a systematic bias. In MS2, cortical thickness was determined from the original T1-weighted images (Table 2 and Fig. S2). Cortical thickness differed significantly between Philips and GE ($p = 0.02$). Intrascanner variability of cortical thickness was lower in the repositioning experiment than over the longer period of 3 weeks (SD 0.01 mm versus 0.04 mm).

Percentage change of brain volume and cortical thickness

Fig. 3A shows PBVC for MS1 between all consecutive pairs of scans at one scanner assessed with SIENA by UCSF (top panel) and TUM (bottom panel). We did not observe a consistent trend across scanners, i.e. we saw random fluctuations of brain volume with opposing trends between scanners. We calculated PBVC between all possible pairs of scans. Median absolute error, SD, minimum and maximums are given in Table 3. We observed PBVC between −4.45% and +4.84% (UCSF). Fluctuations were smaller with SIENA analysis performed by TUM, probably due to different methods of skull stripping (PBVC between −1.38% and 2.27%). We found significant effect of scanner on PBVC (UCSF, $p = 0.01$; TUM, $p = 0.003$).

PBVC calculated from brain volume of all consecutive pairs of scans assessed with SPM/VBM showed similar random fluctuations (Fig. 3 B), as did percentage change of cortical thickness of consecutive scans assessed by FreeSurfer (Fig. 3C). Combining all possible pairs of scans (Table 3) showed PBVC (SPM/VBM) between −2.05% and +2.18% and percentage change of cortical thickness between −3.86% and +4.02% (lesions filled) and between −3.23% and 2.7% (lesions not filled), respectively.

In MS2 (Table 3 and Fig. S3), analysis was performed by TUM only. SIENA's PBVC fluctuated between −0.71% and 0.81% No significant effect of scanner on SIENA's PBVC was observed. PBVC calculated from the brain measures in SPM/VBM fluctuated between −2.12 and 3.82, percentage change in cortical thickness fluctuated between −3.91% and 4.02%. Intrascanner variability of SIENA was lower in the repositioning experiment than over the longer period of three weeks (MAE 0.11 versus 0.36).

Comparison of brain imaging software tools

With all brain imaging software tools, we observed random fluctuations of measurements in MS1 (Fig. 3), with opposing trends between different scanners and methods. For SIENA, we noticed higher variance in the analysis performed by UCSF than by TUM, probably due to different methods of skull stripping. Replication analysis with MS2 showed similar results (Fig. S2 and Table 3). Reliability of PBVC seemed to be better when assessed with SIENA than when calculated from the brain volume values derived by SPM/VBM.

Voxel-wise analysis of grey matter variance

Fig. 4 shows voxel-wise analysis of GM variance expressed by SD for MS1. For GE, we found the highest variance in the basal ganglia, near the base of the skull and in the occipital lobe. With Philips, variance was

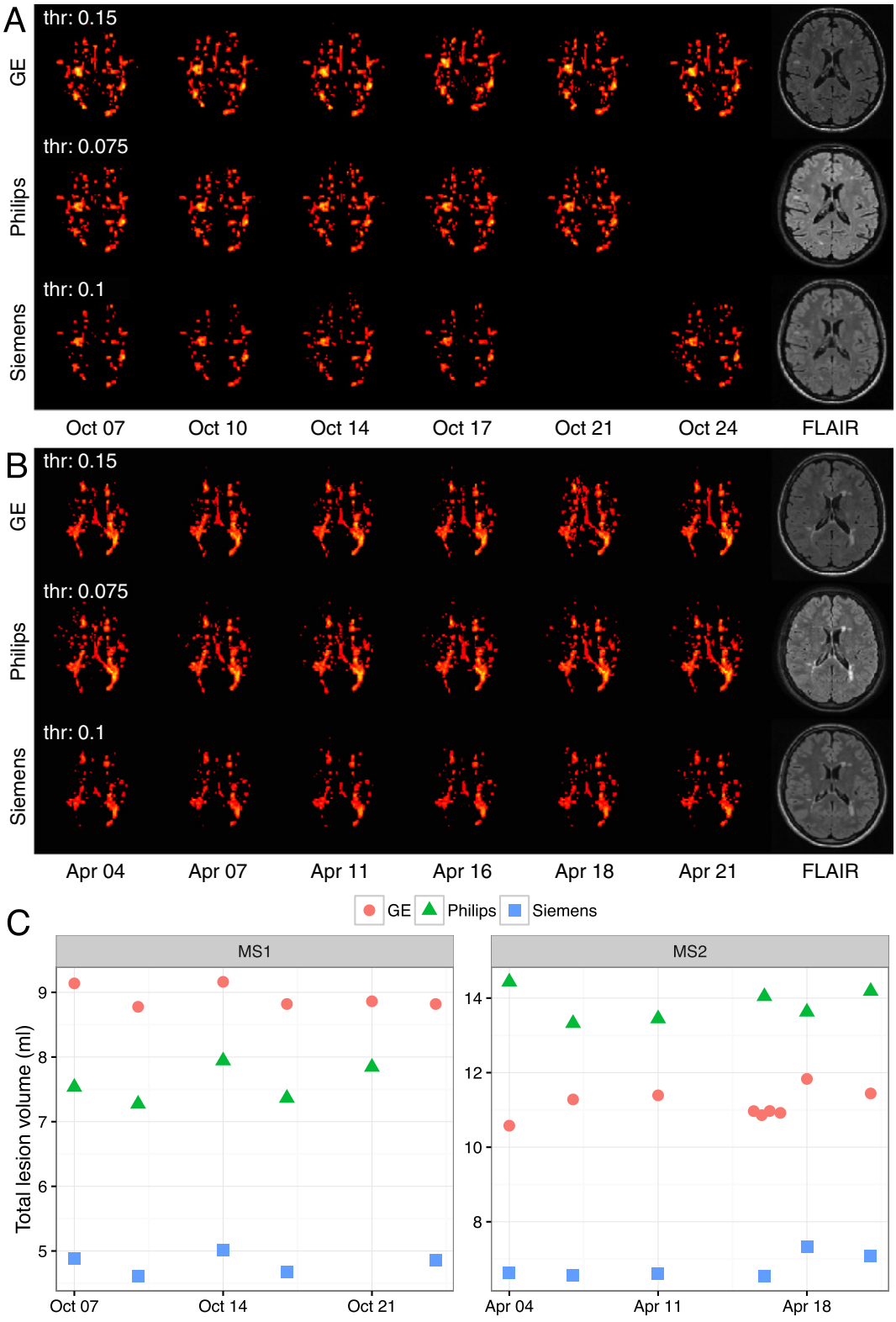

Fig. 1. White matter lesions (MS1 and MS2). A + B) Voxel-wise means along axial slices of lesion maps for each time point and scanner are depicted for MS1 (A) and MS2 (B). One slice of the normalized FLAIR image (t1) is shown on the right site. C) Time course of total lesion volume is shown for each scanner in MS1 (left panel) and MS2 (right panel). Note, the four red dots at April 15 (MS2) mark the repositioning experiment. (A/C left panel) Missing values are due to violations of the scanning protocol (Philips, MS1 10-24-2013; Siemens, MS1 10-21-2013).

highest around the venous sinuses and the anterior base of the skull. Siemens showed the most widespread distribution with highest variance in the basal ganglia, around the venous sinuses, near the base of the skull and in the whole, but particularly the occipital cortex.

Voxel-wise analysis of GM variance in MS2 showed similar predilection areas for local GM variance, although in this subject with the GE scanner variance was more widespread scanner than in MS1 (Fig. S4).

Table 2
Morphometric measures (MS1 and MS2).

	GE			Philips			Siemens		
	Mean	SD	CV (%)	Mean	SD	CV (%)	Mean	SD	CV (%)
	MS1 (3 week interval)								
WM lesion volume (ml)	8.93	0.17	1.94	7.59	0.29	3.86	4.81	0.16	3.39
Brain volume SIENAX (ml)	925.62	3.70	0.40	859.99	5.98	0.70	921.28	7.75	0.84
GM SIENAX (ml)	509.30	4.01	0.79	464.10	5.12	1.10	479.80	5.45	1.14
WM SIENAX (ml)	416.30	6.39	1.54	395.90	1.57	0.40	441.50	11.00	2.49
Brain volume SPM (ml)	938.38	3.59	0.38	921.09	9.77	1.06	930.63	8.52	0.92
GM SPM/VBM (ml)	541.49	4.62	0.85	543.16	9.38	1.73	526.41	8.02	1.52
WM SPM/VBM (ml)	396.90	3.20	0.81	377.93	0.95	0.25	404.22	1.18	0.29
Cortical thickness (mm) FS, lesions filled	2.38	0.02	0.82	2.24	0.03	1.32	2.31	0.04	1.64
Cortical thickness (mm) FS, lesions not filled	2.36	0.02	0.85	2.23	0.02	0.90	2.31	0.03	1.30
	MS2 (3 week interval)								
WM lesion volume (ml)	11.23	0.45	3.98	13.84	0.44	3.2	6.79	0.33	4.91
Brain volume SPM (ml)	1058.4	14.98	1.42	1097.19	5.24	0.48	1046.25	8.82	0.84
GM SPM/VBM (ml)	622.06	12.72	2.04	656.25	5.07	0.77	602.96	7.24	1.2
WM SPM/VBM (ml)	436.35	3.58	0.82	440.94	1.89	0.43	443.29	3.92	0.88
Cortical thickness (mm) FS, lesions not filled	2.51	0.04	1.46	2.46	0.02	0.8	2.49	0.03	1.37
	MS2 (repositioning)								
WM lesion volume (ml)	10.93	0.05	0.48	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Brain volume SPM (ml)	1067.81	4.29	0.40	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
GM SPM/VBM (ml)	628.16	4.02	0.64	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
WM SPM/VBM (ml)	439.65	3.24	0.74	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Cortical thickness (mm) FS, lesions not filled	2.52	0.01	0.56	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

MS, multiple sclerosis; WM, white matter; GM, grey matter; FS, FreeSurfer; ml, milliliter; SD, standard deviation; CV, coefficient of variation; GE, general electric.

Discussion

In this study, we quantified interscanner variability in relation to intrascanner variability of brain MRI in two patients with MS. We compared three different 3-Tesla scanners using several commonly used brain imaging software tools. We found significant scanner effects on MS lesion volume, brain tissue volume and cortical thickness measures. We found high intrascanner variability of brain tissue volumes and cortical thickness measures within three weeks that exceeded variance of simple repositioning experiments.

The effect of different scanners on MS lesion volume observed in this study, is higher than described before (Filippi et al., 1999). We found the major difference between Siemens and the two other scanners, indicating that interscanner variability of MS lesion volume might be largely dependent on the scanners and scanning protocols. The resolution of the T1-weighted images in Siemens ($1.1 \times 1.1 \times 1.0$) was different from the two other scanners ($1.0 \times 1.0 \times 1.0$), which might explain smaller lesion volumes with this scanner. However, all protocols were optimized for brain MRI in MS patients at this particular scanner and comply with the standards of the 'German Competence Network Multiple Sclerosis'. We refrained from adapting the protocols to each other in order to include all sources of interscanner variability that has to be expected in multicenter studies in MS.

When we displayed the lesion volume over time in MS1 (Fig. 1C, left panel), we observed a weekly fluctuation of MS lesion volume that was consistent for all scanners, which could not be found in MS2 (Fig. 1C,

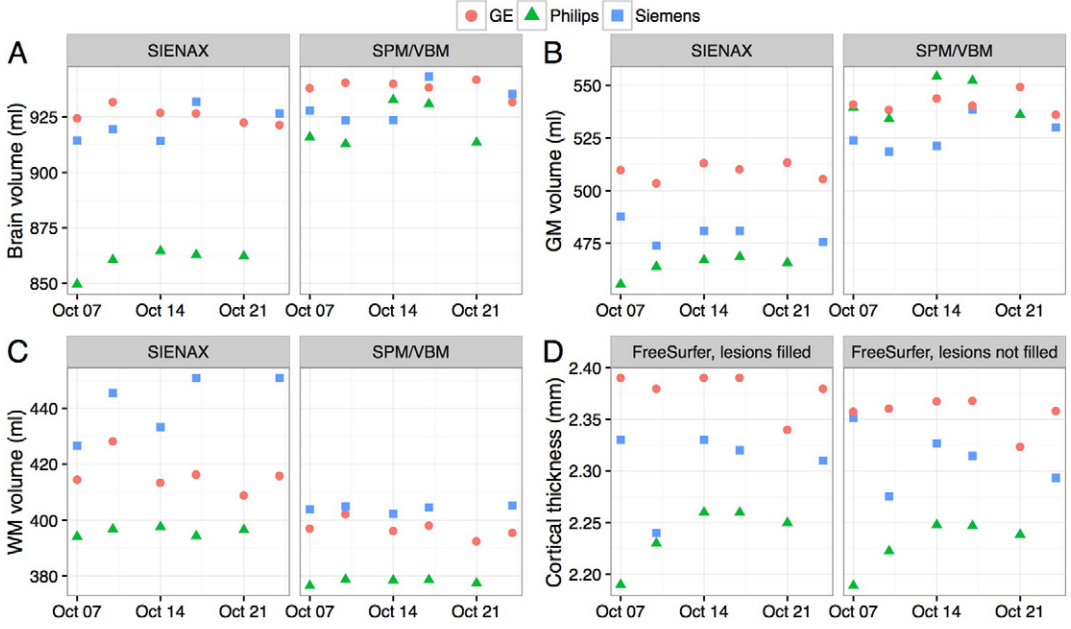

Fig. 2. Volumetric measures (MS1). Total brain volume (A), grey matter (GM) volume (B) and white matter (WM) volume (C) derived with SIENAX (left panels) and SPM/VBM (right panels) are shown for each scanner and time point in milliliter (ml). Cortical thickness assessed with FreeSurfer (D) from the lesion filled (left panel) and original (right panel) T1-weighted images is given for each scanner and time point in millimeter (mm).

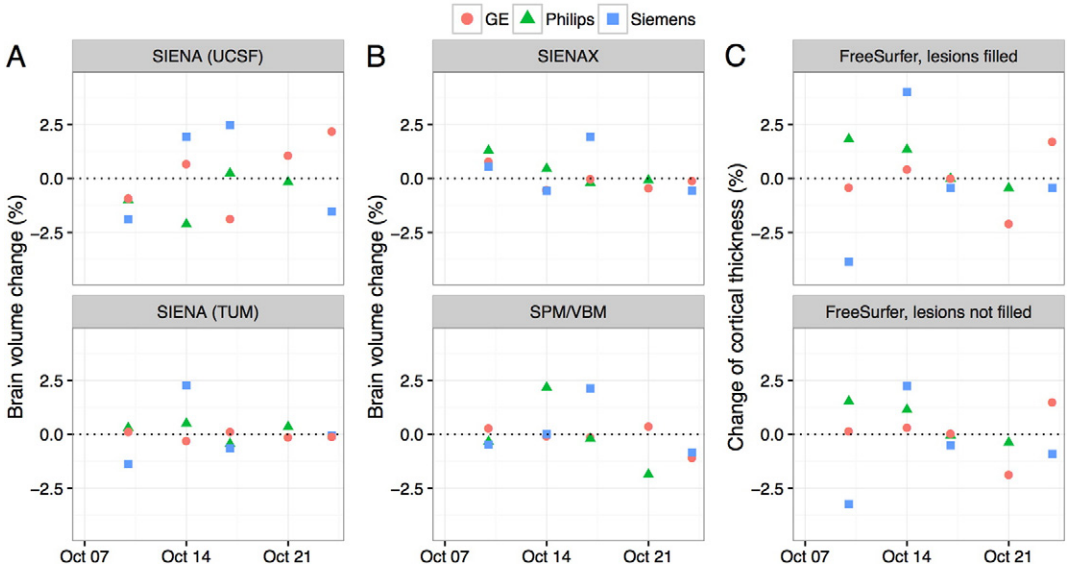

Fig. 3. Brain volume and cortical thickness change (MS1). A) Percentage brain volume change (PBVC) between all pairs of consecutive scans assessed with SIENA by UCSF (top panel) and TUM (bottom panel) is shown. B) PBVC calculated from brain volumes assessed with SIENAX (top panel) and SPM/VBM (bottom panel) is shown for all pairs of consecutive scans. C) Cortical thickness changes between all pairs of consecutive scans assessed with FreeSurfer from the filled (top panel) and non-filled (bottom panel) T1-weighted images are depicted.

right panel). We can only speculate on the nature of this phenomenon, but it might be related to the weekly injections of interferon beta 1a that the patient administered every Tuesday evening. Monthly fluctuations of MS lesion volumes have been described (McFarland et al., 1996; Stone et al., 1995), but to our knowledge, MRI studies observing MS lesion volume within shorter intervals (Absinta et al., 2015; Guttmann et al., 2015; Hannoun et al., 2015) did not address this phenomenon.

We found considerable intrascanner variability of morphometric measures, which was slightly higher than in one similar study in HCs (Huppertz et al., 2010). As an example, CV of brain volume determined with SPM/VBM ranged between 0.38% and 1.42%, while Huppertz et al. described values between 0.1% and 0.78%. This might be explained by disease-related effects. Yet, one study investigating intrascanner variability of brain parenchymal fraction (BPF) in untreated MS patients (Sampat et al., 2010) reported even lower CVs in MS patients (0.3% - 0.6%) than in HCs (0.46%).

In contrast to one previous study (Nakamura et al., 2015), we did not observe a significant effect of scan time on brain volume. However, the reported effect is very small (<0.5% of brain volume) and taking into account the error of measurement detected in our study, it seems impossible to reliably measure this effect in a single subject.

Median absolute error of SIENA's PBVC within three weeks was higher than described in the validation study of the method in HC (Smith et al., 2002b). However, one previous study in MS patients, found SD of PBVC comparable to our results (Durand-Dubief et al., 2012), indicating that disease-related effects might introduce additional variance. Unlike the former two studies (Durand-Dubief et al., 2012; Smith et al., 2002b), we did not run SIENA with the default parameters, because we observed severe misclassifications of brain tissue with BET in our data as had been described before (Boesen et al., 2004; Fennema-Notestine et al., 2006; Leung et al., 2011; Popescu et al., 2012; Vrenken et al., 2013). After editing of the SIENA workflow, accuracy of the measurement was much better, especially when PVE labels from VBM8 were used for brain segmentation (Table 3 and Fig. 3).

The benefit of lesion filling seems to be clear for segmentation-based methods as VBM/SPM and SIENAX (Battaglini et al., 2012; Ceccarelli et al., 2012; Chard et al., 2010; Sdika and Pelletier, 2009), but we did not find a clear recommendation for lesion filling when estimating cortical thickness. While one study found only small effects of lesion filling

Table 3
Brain volume and cortical thickness change (MS1 and MS2).

	GE				Philips				Siemens			
	MAE (%)	SD (%)	Min (%)	Max (%)	MAE (%)	SD (%)	Min (%)	Max (%)	MAE (%)	SD (%)	Min (%)	Max (%)
	MS 1											
PBVC SIENA UCSF	1.73	1.94	−1.96	4.84	2.44	1.57	−4.45	0.24	1.58	1.94	−1.89	3.77
PBVC SIENA TUM	0.13	0.14	−0.38	0.13	0.41	0.39	−0.44	0.82	0.82	1.18	−1.38	2.27
PBVC SPM/VBM	0.29	0.52	−1.08	0.42	1.73	1.55	−2.05	2.18	1.06	1.12	−0.84	2.14
Cortical thickness change FS, lesions filled	0.42	1.38	−2.09	1.71	1.35	1.08	−0.44	1.83	0.86	3.23	−3.86	4.02
Cortical thickness change FS, lesions not filled	0.42	0.97	−1.88	1.48	1.13	1.18	−0.42	2.70	1.50	1.75	−3.23	2.24
	MS 2 (3 week interval)											
PBVC SIENA TUM	0.36	0.45	−0.71	0.81	0.18	0.29	−0.61	0.23	0.33	0.39	−0.69	0.5
PBVC SPM/VBM	1.57	2	−2.12	3.82	0.5	0.69	−1.29	0.82	1.01	1.23	−1.72	2.26
Cortical thickness change FS, lesions not filled	1.69	2.15	−2.68	4.02	0.9	1.15	−1.84	2.01	0.85	1.99	−3.91	3.41
	MS 2 (repositioning)											
PBVC SIENA TUM	0.11	0.16	−0.23	0.23	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
PBVC SPM/VBM	0.45	0.54	−0.89	0.59	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Cortical thickness change FS, lesions not filled	0.85	0.56	−1.07	0.2	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

PBVC, percentage brain volume change; University of California San Francisco; TUM, Technische Universität München; FS, FreeSurfer; GE, general electric; MAE, median absolute error; SD, standard deviation; Min, minimum; Max, maximum, GE, general electric.

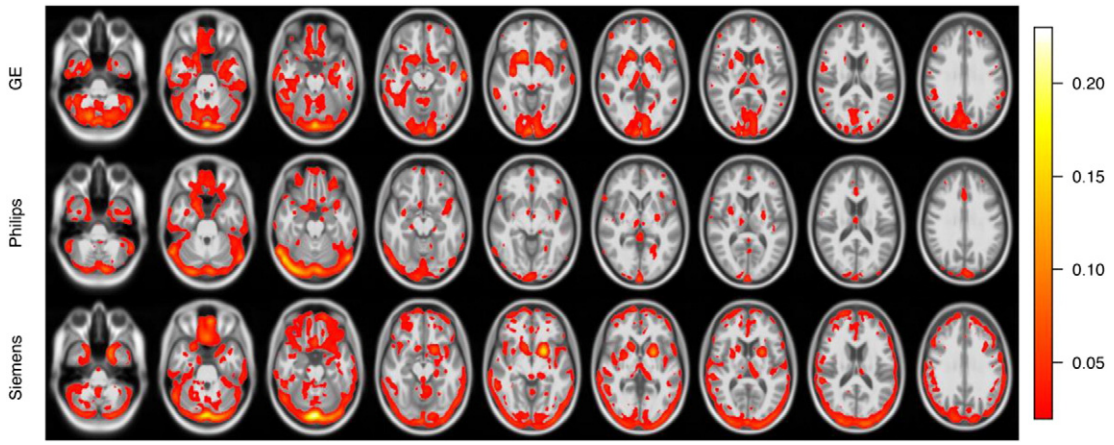

Fig. 4. Voxel-wise analysis of grey matter variance. Voxel-wise standard deviation of all grey matter images is displayed for each scanner. Only values >0.02 are shown.

on regional and global cortical thickness measures (Govindarajan et al., 2015), others described improvement of accuracy of cortical thickness measurement by lesion filling (Magon et al., 2014; Tillema et al., 2015). However, lesion filling might also produce artificial thinning of the cortex in case of juxta-cortical lesions (Tillema et al., 2015). Following the preferences of the two laboratories, we performed the analysis twice, from the lesion-filled (UCSF) and original (TUM) T1-weighted images, which did not result in systematic differences of cortical thickness measures in this particular subject. However, this is only an observation in a single subject and larger studies in patients with high load of juxta-cortical lesions are needed to better address this issue.

The small subject number is one major limitation of our study. Our subjects were very similar concerning lesion load, disease stage and age, so that our results cannot simply be extrapolated to all MS patients. We noticed that our patients were very experienced with MR scanning, probably due to many scans in clinical routine they had undergone before. They positioned themselves accurately in the correct position. Variability with more disabled MS patients might be even higher due to more pronounced alterations of the brain structure (MS lesions, diffuse white matter damage, atrophy). Unfortunately, we were not able to find a more disabled patient who was willing to undergo this strenuous protocol. Further, the scanners with the protocols established constitute only one example of the respective brands. Whether our findings also translate to other scanners of respective brands remains to be studied.

Another drawback of our study design is that we are not able to disentangle the different sources of variance. For example, scanner effects cannot be differentiated from protocol effects. However, this was not the purpose of our study. Our aim was to include all sources of variance across and within scanners as it must be expected in clinical routine as well as in longitudinal and multicenter studies in MS patients.

In conclusion, the overall degree of variance across three different standardized 3 Tesla protocols necessitates modelling of each scanner, which has been suggested recently to account for numerous different protocols including different field strengths (Jones et al., 2013). Most importantly, the overall degree of variance within one scanner challenges the idea of individual atrophy measurement from MRI scans in clinical routine by current methods. In any case, the effect size has to exceed the intrascanner variability. Assuming a physiological loss of brain volume of about 0.3% per year in healthy adult subjects (Good et al., 2001), which may double in MS (De Stefano et al., 2010; De Stefano et al., 2015), even with the brain imaging tool with highest accuracy, reliable estimation of brain atrophy in individual patients seems only possible over periods of at least 5 years.

Finally, the degree of variance across scanners is largely independent of the brain imaging software tool, but optimized settings are required, especially for SIENA. As we were only able to test three of the commonly used brain imaging software tools, we decided to make the MRI images publicly available as Supplemental material to this article to enable further analysis.

Disclosure statement

Viola Biberacher has received travel expenses for attending meetings from TEVA Pharma GmbH.

Paul Schmidt has nothing to disclose.

Anisha Keshavan has nothing to disclose.

Christine C. Boucard has nothing to disclose.

Ruthger Righart has received research support from the Hertie Foundation.

Philipp Sämann has nothing to disclose.

Christine Preibisch has nothing to disclose.

Daniel Fröbel has nothing to disclose.

Lilian Aly has nothing to disclose.

Bernhard Hemmer has served on scientific advisory boards for Roche, Novartis, Bayer Schering, Merck Serono, Biogen Idec, GSK, Chugai Pharmaceuticals, Micromet, Genentech and Genzyme Corporation; has received speaker honoraria from Bayer Schering, Novartis, Biogen Idec, Merck Serono, Roche, and Teva Pharmaceutical Industries Ltd.; and has received research support from Biogen Idec, Bayer Schering, Merck Serono, Five prime, Metanomics, Chugai Pharmaceuticals, and Novartis. He has filed a patent for the detection of antibodies and T cells against KIR4.1 in a subpopulation of MS patients and genetic determinants of neutralizing antibodies to interferon-beta.

Claus Zimmer has nothing to disclose.

Roland G. Henry has nothing to disclose.

Mark Mühlau has received research support from Merck Serono and Novartis; he has received travel expenses for attending meetings from Bayer, and Merck Serono; he has received honoraria for lecturing from Merck Serono; he has received Investigator fees for a Phase III clinical study from Biogen Idec.

Funding

This work was funded by the 'Hertie Foundation' (grand P1140092 'Myelin mapping in MS') and supported by the 'German Competence Network Multiple Sclerosis' (German Ministry for Research and Education grand 01GI1307B).

Acknowledgment

We want to thank our patients who were willing to undergo this strenuous protocol and allowed us to acquire this unique dataset.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2016.07.035.

References

Absinta, M., Nair, G., Sati, P., Cortese, I.C., Filippi, M., Reich, D.S., 2015. Direct MRI detection of impending plaque development in multiple sclerosis. Neurol. Neuroimmunol. Neuroinflamm. 2, e145.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113.

Battaglini, M., Jenkinson, M., De Stefano, N., 2012. Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum. Brain Mapp. 33, 2062–2071.

Bezzola, L., Merillat, S., Gaser, C., Jancke, L., 2011. Training-induced neural plasticity in golf novices. J. Neurosci. 31, 12444–12448.

Boesen, K., Rehm, K., Schaper, K., Stoltzner, S., Woods, R., Luders, E., Rottenberg, D., 2004. Quantitative comparison of four brain extraction algorithms. NeuroImage 22, 1255–1261.

Caramanos, Z., Fonov, V.S., Francis, S.J., Narayanan, S., Pike, G.B., Collins, D.L., Arnold, D.L., 2010. Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. NeuroImage 49, 1601–1611.

Ceccarelli, A., Jackson, J.S., Tauhid, S., Arora, A., Gorky, J., Dell'Oglio, E., Bakshi, A., Chitnis, T., Khoury, S.J., Weiner, H.L., Guttmann, C.R.G., Bakshi, R., Neema, M., 2012. The impact of lesion in-painting and registration methods on voxel-based morphometry in detecting regional cerebral gray matter atrophy in multiple sclerosis. Am. J. Neuroradiol. 33, 1579–1585.

Chard, D.T., Parker, G.J., Griffin, C.M., Thompson, A.J., Miller, D.H., 2002. The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology. J. Magn. Reson. Imaging 15, 259–267.

Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A., 2010. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J. Magn. Reson. Imaging 32, 223–228.

Compston, A., Coles, A., 2008. Multiple sclerosis. Lancet 372, 1502–1517.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis — I. Segmentation and surface reconstruction. NeuroImage 9, 179–194.

De Stefano, N., Giorgio, A., Battaglini, M., Rovaris, M., Sormani, M.P., Barkhof, F., Korteweg, T., Enzinger, C., Fazekas, F., Calabrese, M., Dinacci, D., Tedeschi, G., Gass, A., Montalban, X., Rovira, A., Thompson, A., Comi, G., Miller, D.H., Filippi, M., 2010. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. Neurology 74, 1868–1876.

De Stefano, N., Stromillo, M.L., Giorgio, A., Bartolozzi, M.L., Battaglini, M., Baldini, M., Portaccio, E., Amato, M.P., Sormani, M.P., 2015. Establishing pathological cut-offs of brain atrophy rates in multiple sclerosis. J. Neurol. Neurosurg. Psychiatry.

Derakhshan, M., Caramanos, Z., Giacomini, P.S., Narayanan, S., Maranzano, J., Francis, S.J., Arnold, D.L., Collins, D.L., 2010. Evaluation of automated techniques for the quantification of grey matter atrophy in patients with multiple sclerosis. NeuroImage 52, 1261–1267.

Droby, A., Lukas, C., Schaenzer, A., Spiwoks-Becker, I., Giorgio, A., Gold, R., De Stefano, N., Kugel, H., Deppe, M., Wiendl, H., Meuth, S.G., Acker, T., Zipp, F., Deichmann, R., 2015. A human post-mortem brain model for the standardization of multi-centre MRI studies. NeuroImage 110, 11–21.

Duning, T., Kloska, S., Steinstrater, O., Kugel, H., Heindel, W., Knecht, S., 2005. Dehydration confounds the assessment of brain atrophy. Neurology 64, 548–550.

Durand-Dubief, F., Belaroussi, B., Armspach, J.P., Dufour, M., Roggerone, S., Vukusic, S., Hannoun, S., Sappey-Marinier, D., Confavreux, C., Cotton, F., 2012. Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. Am. J. Neuroradiol. 33, 1918–1924.

Fennema-Notestine, C., Ozyurt, I.B., Clark, C.P., Morris, S., Bischoff-Grethe, A., Bondi, M.W., Jernigan, T.L., Fischl, B., Segonne, F., Shattuck, D.W., Leahy, R.M., Rex, D.E., Toga, A.W., Zou, K.H., Brown, G.G., 2006. Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction, and slice location. Hum. Brain Mapp. 27, 99–113.

Filippi, M., Rocca, M.A., Gasperini, C., Sormani, M.P., Bastianello, S., Horsfield, M.A., Pozzilli, C., Comi, G., 1999. Interscanner variation in brain MR lesion load measurements in multiple sclerosis using conventional spin-echo, rapid relaxation-enhanced, and fast-FLAIR sequences. Am. J. Neuroradiol. 20, 133–137.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. U. S. A. 97, 11050–11055.

Geurts, J.J., Barkhof, F., 2008. Grey matter pathology in multiple sclerosis. Lancet Neurol. 7, 841–851.

Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14, 21–36.

Govindarajan, K.A., Datta, S., Hasan, K.M., Choi, S., Rahbar, M.H., Cofield, S.S., Cutter, G.R., Lublin, F.D., Wolinsky, J.S., Narayana, P.A., 2015. Effect of in-painting on cortical thickness measurements in multiple sclerosis: a large cohort study. Hum. Brain Mapp. 36, 3749–3760.

Guttmann, C.R., Rousset, M., Roch, J.A., Hannoun, S., Durand-Dubief, F., Belaroussi, B., Cavallari, M., Rabilloud, M., Sappey-Marinier, D., Vukusic, S., Cotton, F., 2015. Multiple sclerosis lesion formation and early evolution revisited: a weekly high-resolution magnetic resonance imaging study. Mult. Scler.

Hagemann, G., Ugur, T., Schleussner, E., Mentzel, H.J., Fitzek, C., Witte, O.W., Gaser, C., 2011. Changes in brain size during the menstrual cycle. PLoS One 6, 7.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage 32, 180–194.

Hannoun, S., Roch, J.A., Durand-Dubief, F., Vukusic, S., Sappey-Marinier, D., Guttmann, C.R., Cotton, F., 2015. Weekly multimodal MRI follow-up of two multiple sclerosis active lesions presenting a transient decrease in ADC. Brain Behav. 5, e00307.

Huppertz, H.-J., Kroell-Seger, J., Kloeppel, S., Ganz, R.E., Kassubek, J., 2010. Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. NeuroImage 49, 2216–2224.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841.

Jones, B.C., Nair, G., Shea, C.D., Crainiceanu, C.M., Cortese, I.C., Reich, D.S., 2013. Quantification of multiple-sclerosis-related brain atrophy in two heterogeneous MRI datasets using mixed-effects modeling. Neuroimage Clin. 3, 171–179.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., Fischl, B., 2009. 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 46, 177–192.

Kempton, M.J., Ettinger, U., Schmechtig, A., Winter, E.M., Smith, L., McMorris, T., Wilkinson, I.D., Williams, S.C.R., Smith, M.S., 2009. Effects of acute dehydration on brain morphology in healthy humans. Hum. Brain Mapp. 30, 291–298.

Kempton, M.J., Ettinger, U., Foster, R., Williams, S.C.R., Calvert, G.A., Hampshire, A., Zelaya, F.O., O'Gorman, R.L., McMorris, T., Owen, A.M., Smith, M.S., 2011. Dehydration affects brain structure and function in healthy adolescents. Hum. Brain Mapp. 32, 71–79.

Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., Ourselin, S., 2011. Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. NeuroImage 55, 1091–1108.

Magon, S., Gaetano, L., Chakravarty, M.M., Lerch, J.P., Naegelin, Y., Stippich, C., Kappos, L., Radue, E.-W., Sprenger, T., 2014. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. BMC Neurosci. 15.

McFarland, H.F., Stone, L.A., Calabresi, P.A., Maloni, H., Bash, C.N., Frank, J.A., 1996. MRI studies of multiple sclerosis: implications for the natural history of the disease and for monitoring effectiveness of experimental therapies. Mult. Scler. 2, 198–205.

Moorhead, T.W.J., Gountouna, V.-E., Job, D.E., McIntosh, A.M., Romaniuk, L., Lymer, G.K.S., Whalley, H.C., Waiter, G.D., Brennan, D., Ahearn, T.S., Cavanagh, J., Condon, B., Steele, J.D., Wardlaw, J.M., Lawrie, S.M., 2009. Prospective multi-centre voxel based morphometry study employing scanner specific segmentations: procedure development using CaliBrain structural MRI data. BMC Med. Imaging 9, 8.

Nakamura, K., Brown, R.A., Narayanan, S., Collins, D.L., Arnold, D.L., Alzheimer's Dis, N., 2015. Diurnal fluctuations in brain volume: statistical analyses of MRI from large populations. NeuroImage 118, 126–132.

Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann. Neurol. 69, 292–302.

Popescu, V., Battaglini, M., Hoogstrate, W.S., Verfaillie, S.C.J., Sluimer, I.C., van Schijndel, R.A., van Dijk, B.W., Cover, K.S., Knol, D.L., Jenkinson, M., Barkhof, F., de Stefano, N., Vrenken, H., Grp, M.S., 2012. Optimizing parameter choice for FSL-brain extraction tool (BET) on 3D T1 images in multiple sclerosis. NeuroImage 61, 1484–1494.

R-Core-Team, 2015. R: A Language and Environment for Statistical Computing, Version 3.2.2. R Foundation for Statistical Computing, Vienna, Austria.

Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. NeuroImage 61, 1402–1418.

Reuter, M., Tisdall, M.D., Qureshi, A., Buckner, R.L., van der Kouwe, A.J.W., Fischl, B., 2015. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. NeuroImage 107, 107–115.

Sampat, M.P., Healy, B.C., Meier, D.S., Dell'Oglio, E., Liguori, M., Guttmann, C.R.G., 2010. Disease modeling in multiple sclerosis: assessment and quantification of sources of variability in brain parenchymal fraction measurements. NeuroImage 52, 1367–1373.

Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V.J., Zimmer, C., Hemmer, B., Muhlau, M., 2012. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. NeuroImage 59, 3774–3783.

Schnack, H.G., van Haren, N.E.M., Brouwer, R.M., van Baal, G.C.M., Picchioni, M., Weisbrod, M., Sauer, H., Cannon, T.D., Huttunen, M., Lepage, C., Collins, D.L., Evans, A., Murray, R.M., Kahn, R.S., Pol, H.E.H., 2010. Mapping reliability in multicenter MRI: voxel-based morphometry and cortical thickness. Hum. Brain Mapp. 31, 1967–1982.

Sdika, M., Pelletier, D., 2009. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. Hum. Brain Mapp. 30, 1060–1067.

Shokouhi, M., Barnes, A., Suckling, J., Moorhead, T.W.J., Brennan, D., Job, D., Lymer, K., Dazzan, P., Marques, T.R., MacKay, C., McKie, S., Williams, S.C.R., Lawrie, S.M., Deakin, B., Williams, S.R., Condon, B., 2011. Assessment of the impact of the scanner-related factors on brain morphometry analysis with Brainvisa. BMC Med. Imaging 11.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.
Smith, S.M., De Stefano, N., Jenkinson, M., Matthews, P.M., 2001. Normalized accurate measurement of longitudinal brain change. J. Comput. Assist. Tomogr. 25, 466–475.
Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002a. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17, 479–489.
Smith, S.M., Zhang, Y.Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002b. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17, 479–489.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (Suppl. 1), S208–S219.
Sormani, M.P., Arnold, D.L., De Stefano, N., 2014. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. Ann. Neurol. 75, 43–49.
Stone, L.A., Albert, P.S., Smith, M.E., DeCarli, C., Armstrong, M.R., McFarlin, D.E., Frank, J.A., McFarland, H.F., 1995. Changes in the amount of diseased white matter over time in patients with relapsing-remitting multiple sclerosis. Neurology 45, 1808–1814.
Suckling, J., Barnes, A., Job, D., Brennan, D., Lymer, K., Dazzan, P., Marques, T.R., MacKay, C., McKie, S., Williams, S.R., Williams, S.C.R., Deakin, B., Lawrie, S., 2012. The neuro/PsyGRID calibration experiment: identifying sources of variance and bias in multicenter MRI studies. Hum. Brain Mapp. 33, 373–386.
Tillema, J.M., Hulst, H.E., Rocca, M.A., Vrenken, H., Steenwijk, M.D., Damjanovic, D., Enzinger, C., Ropele, S., Tedeschi, G., Gallo, A., Ciccarelli, O., Rovira, A., Montalban, X., de Stefano, N., Stromillo, M.L., Filippi, M., Barkhof, F., 2015. Regional cortical thinning in multiple sclerosis and its relation with cognitive impairment: A multicenter study. Mult. Scler.
Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5, 99–114.
Vrenken, H., Jenkinson, M., Horsfield, M.A., Battaglini, M., van Schijndel, R.A., Rostrup, E., Geurts, J.J.G., Fisher, E., Zijdenbos, A., Ashburner, J., Miller, D.H., Filippi, M., Fazekas, F., Rovaris, M., Rovira, A., Barkhof, F., de Stefano, N., Grp, M.S., 2013. Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. J. Neurol. 260, 2458–2471.
Walters, R.J., Fox, N.C., Crum, W.R., Taube, D., Thomas, D.J., 2001. Haemodialysis and cerebral oedema. Nephron 87, 143–147.
Welch, B.L., 1951. On the comparison of several mean values — an alternative approach. Biometrika 38, 330–336.
Zhang, Y.Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.
Zivadinov, R., Reder, A.T., Filippi, M., Minagar, A., Stueve, O., Lassmann, H., Racke, M.K., Dwyer, M.G., Frohman, E.M., Khan, O., 2008. Mechanisms of action of disease-modifying agents and brain volume changes in multiple sclerosis. Neurology 71, 136–144.