Postmortem Validation of MRI Cortical Volume Measurements in MS

Veronica Popescu,^{1†} Roel Klaver,^{2†} Adriaan Versteeg,¹ Pieter Voorn,²
 Jos W.R. Twisk,³ Frederik Barkhof,¹ Jeroen J.G. Geurts,² and
 Hugo Vrenken^{1,4*}

¹Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, The Netherlands

²Department of Anatomy and Neurosciences, VU University Medical Center, Amsterdam, The Netherlands

³Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

⁴Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, The Netherlands

Abstract: Grey matter (GM) atrophy is a prominent aspect of multiple sclerosis pathology and an important outcome in studies. GM atrophy measurement requires accurate GM segmentation. Several methods are used in vivo for measuring GM volumes in MS, but assessing their validity in vivo remains challenging. In this postmortem study, we evaluated the correlation between postmortem MRI cortical volume or thickness and the cortical thickness measured on histological sections. Sixteen MS brains were scanned in situ using 3DT1-weighted MRI and these images were used to measure regional cortical volume using FSL-SIENAX, FreeSurfer, and SPM, and regional cortical thickness using FreeSurfer. Subsequently, cortical thickness was measured histologically in 5 systematically sampled cortical areas. Linear regression analyses were used to evaluate the relation between MRI regional cortical volume or thickness and histological cortical thickness to determine which postprocessing technique was most valid. After correction for multiple comparisons, we observed a significant correlation with the histological cortical thickness for FSL-SIENAX cortical volume with manual editing (std. $\beta = 0.345$, adjusted $R^2 = 0.105$, $P = 0.005$), and FreeSurfer cortical volume with manual editing (std.

Contract grant sponsor: Dutch MS Research Foundation, The Netherlands; Contract grant number: 10-718 MS, 09-358d, and 14-358e

V.P. and R.K. contributed equally to this article.

Disclosures: V. Popescu is supported by the Dutch MS Research Foundation grant 10-718. R. Klaver is supported by the Dutch MS Research Foundation grant 09-358d. A. Versteeg reports no disclosures. Dr Voorn reports no disclosures. Dr Twisk reports no disclosures. Dr Barkhof serves on the editorial boards of Brain, European Radiology, Neuroradiology, Multiple Sclerosis journal and Radiology and serves as a consultant for Bayer-Schering Pharma, Sanofi-Aventis, Biogen-Idec, Teva, Novartis, Roche, Synthon BV, Jansen Research. Dr Geurts is an editor for Multiple Sclerosis Journal, associate editor for BMC Neurology, and serves on the editorial boards of Neurology and MS International; he is a member of the scientific advisory board of the Dutch MS Research Foundation, of the MS Society of Canada and a scientific

steering committee member of the International Progressive MS Alliance; he has served as a consultant for MerckSerono, Novartis, Biogen Idec, Genzyme and Teva Pharmaceuticals. Dr Vrenken has received funding for collaborative research projects from Pfizer, Novartis, and Merck-Serono, and a speaker honorarium from Novartis; all funds paid directly to his institution.

*Correspondence to: H. Vrenken, VU University Medical Center, Department of Radiology and Nuclear Medicine and Department of Physics and Medical Technology, De Boelelaan 1118, 1081 HZ Amsterdam, The Netherlands. E-mail: h.vrenken@vumc.nl

Received for publication 9 June 2015; Revised 19 February 2016; Accepted 21 February 2016.

DOI: 10.1002/hbm.23168

Published online 4 March 2016 in Wiley Online Library (wileyonlinelibrary.com).

$\beta = 0.379$, adjusted $R^2 = 0.129$, $P = 0.003$). In addition, there was a significant correlation between FreeSurfer cortical thickness with manual editing and histological cortical thickness (std. $\beta = 0.381$, adjusted $R^2 = 0.130$, $P = 0.003$). The results support the use of FSL-SIENAX and FreeSurfer in cases of severe MS pathology. Interestingly none of the methods were significant in automated mode, which supports the use of manual editing to improve the automated segmentation. *Hum Brain Mapp* 37:2223–2233, 2016. © 2016 Wiley Periodicals, Inc.

Key words: brain atrophy; FSL-SIENAX; SIENAX; FreeSurfer; SPM; multiple sclerosis; grey matter

INTRODUCTION

Multiple sclerosis (MS) is a demyelinating disease of the central nervous system, characterized by inflammation and neurodegeneration. Magnetic resonance imaging (MRI) reveals lesions in both the white (WM) and grey matter (GM). WM lesions are insufficient to explain clinical disability [Giorgio et al., 2008]. Possibly, brain atrophy is a better correlate of clinical disability. Both WM and GM atrophy play a role from the earliest phases of the disease [Geurts et al., 2009]. GM atrophy seems to accelerate especially with disease progression [Fisher et al., 2008] and is directly related to cognitive impairment [Benedict et al., 2009]. In contrast to WM and GM lesions, GM atrophy seems to represent the neurodegenerative component of MS pathology [Popescu et al., 2015].

Several methods have been used in the literature for measuring GM volumes in vivo in MS, with the most popular being FreeSurfer, FSL-SIENAX, and SPM. While some validation of these methods in the in vivo setting in MS is available [Derakhshan et al., 2010; Durand-Dubief et al., 2012], this does not include direct histological measurement of the tissue. In general, outside the MS field, only a few studies did a comparison between MRI-based cortical thickness and cortical thickness measured directly in tissue [Cardinale et al., 2014; Rosas et al., 2002]. These studies report a relatively good estimation of cortical thickness in Huntington's disease [Rosas et al., 2002] and epilepsy [Cardinale et al., 2014]. However, these results are lacking in MS.

In this postmortem study, we aimed to assess the relation of FreeSurfer, FSL-SIENAX, and SPM regional cortical GM volume measurements in five systematically sampled cortical areas, with a histological reference: cortical thickness measured on Nissl stained sections. In addition, we also assessed cortical thickness measured by FreeSurfer.

MATERIALS AND METHODS

Tissue was collected through our collaboration with the Netherlands Brain Bank (<http://www.brainbank.nl/>), using rapid autopsies with postmortem MRI scanning.

The MRI protocol included both whole-brain in situ MRI on a 1.5 T Siemens MRI scanner, and MRI of 10-mm-thick coronal brain slices, cut at autopsy, as described in

more detail before [Popescu et al., 2015; Seewann et al., 2009]. In short, after arrival in the hospital, the patient undergoes an MRI scan with the brain in situ, similar to in vivo MRI scanning. Subsequently, the craniotomy takes place and the brain is cut in 10-mm-thick coronal slices. From these slices, tissue is sampled from five standardized anatomical regions: the superior and inferior frontal gyrus, the anterior cingulate gyrus, the inferior parietal lobule, and the superior temporal gyrus. Specifically, these regions were selected because they are spread throughout the cerebral cortex. Figure 1 contains a schematic representation of our workflow.

Standard Protocol Approvals, Registrations, and Patient Consents

The study was approved by the local institutional ethics review board. Before death, the donors or their next of kin provided written informed consent for the use of their tissue and clinical information for research purposes to The Netherlands Brain Bank.

MR Imaging

The in situ imaging protocol included a sagittal 3D T1-weighted imaging sequence (TR = 2700 ms, TE = 5.1 ms, TI = 950 ms, FA = 8, voxel size = $1.2 \times 1.2 \times 1.3$ mm), which was used to measure cortical volume globally, and in specific cortical regions corresponding with the tissue samples extracted. 3D-FLAIR images were acquired (TR = 6500 ms, TE = 355 ms, TI = 2200 ms, voxel size = $1.2 \times 1.2 \times 1.3$ mm) and used for pial surface refinement in FreeSurfer. A dual-echo T2-weighted sequence (TR = 5640 ms, TE = 22/98 ms, FA = 150, 3 mm axial slices, in-plane resolution = 1×1 mm) was acquired for WM lesion assessment.

MR Image Processing

The analyses were performed on the "NCA" computer cluster of the Neuroscience Campus Amsterdam running under Linux CentOS (<http://www.neurosciencecampus-amsterdam.nl/en/news-agenda/news-archive/2013/NCA-GRID-cluster-computer.asp>). All volume measures described below are normalized for head size using the

Figure 1.

A schematic representation of our workflow from autopsy to statistical analyses.

volumetric scaling factor obtained from the FSL-SIENAX software described below.

WM lesion volumes

WM lesions were manually outlined on the short echo of the T2-weighted sequence using MIPAV software and WM lesion volumes were quantified.

Lesion-filling

MS lesions may appear hypointense on 3DT1 images and may hamper the automated GM segmentation [Battaglini et al., 2012]. This was avoided by lesion-filling [Sanfilippo et al., 2005] using LEAP [Chard et al., 2010]. The lesion masks used in the lesion-filling were manually outlined on the 3DT1 images by a single rater (VP) using MIPAV (<http://mipav.cit.nih.gov/>). All subsequent analyses described in this article were performed using the lesion-filled images. Images were corrected for geometrical distortion due to nonlinear gradients using vendor standard software. Apart from this, no preprocessing or correction were applied besides the correction standardly done by FSL-SIENAX, FreeSurfer, or SPM.

FSL-SIENAX

Normalized brain volume was estimated with SIENAX [Smith et al., 2002] part of FSL [Jenkinson et al., 2012] (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>). SIENAX starts by extracting brain and skull images from the single whole-head input data. The brain image was affine-registered to MNI152 space (using the skull image to determine the registration scaling), to obtain the volumetric scaling factor used to normalize for head size. The total brain volume (including separate estimates of GM and WM volumes) was calculated using tissue-type segmentation with partial volume estimation. The Harvard-Oxford atlas, delivered with FSL, was nonlinearly registered (using FNIRT) to the GM partial volume estimates and regional cortical volumes were calculated. The brain extraction step in SIENAX was applied in two different ways: (1) using automated BET with optimized settings [Popescu et al., 2012] and (2) using manually edited brain masks created using the protocol described in a previous study [Popescu et al., 2012, 2015]. Results for each method are reported.

FreeSurfer

Cortical reconstruction was performed with the FreeSurfer image analysis suite version 5.3, which is documented

and freely available for download online (<http://surfer.nmr.mgh.harvard.edu/>). Briefly, this processing includes removal of nonbrain tissue, automated Talairach transformation, segmentation of the subcortical structures, intensity normalization, tessellation of the GM/WM boundary, automated topology correction, and surface deformation following intensity gradients to optimally place the GM/WM and GM/CSF borders at the location where the greatest shift in intensity defines the transition to the other tissue class. Once the cortical models are complete, the surface is inflated and registered to a spherical atlas and the cerebral cortex was parcellated into units based on gyral and sulcal structure [Dale et al., 1999; Fischl et al., 1999]. The FreeSurfer GM volume estimations were also done in two different ways: (1) using FreeSurfer's default automated settings and (2) using manually edited brain and WM masks. The accuracy of these brain and WM masks was improved by refining the pial surface using 3D-FLAIR images (TR = 6500 ms, TE = 355 ms, TI = 2200 ms, voxel size = $1.2 \times 1.2 \times 1.3$ mm). The final GM segmentation was visually checked for correctness. Results for each method are reported.

SPM

SPM8 (Statistical Parametric Mapping Functional Imaging Laboratory, University College London, London, UK) accessed through the VBM8 toolbox, automatically identifies and quantifies per voxel GM, WM, and CSF in all scans. The AAL atlas delivered as part of SPM toolboxes was registered with FSL-FLIRT and FSL-FNIRT to the native scans and then used for calculating the regional GM volumes from the GM segmentation.

MRI whole-brain cortical volume

For FSL-SIENAX, FreeSurfer, and SPM, whole-brain cortical volume was calculated by taking the volume of all cortical areas minus the hippocampus and cerebellar cortex (both hemispheres).

Histopathology

Tissue blocks were formalin fixed, paraffin-embedded, cut and mounted on Superfrost Plus glass slides (VWR international; Leuven, Belgium). Sections were deparaffinized and rehydrated in graded series of xylene and ethanol. The sections were stained for cell bodies using a Nissl staining, which mainly stains nucleotides and therefore cell nucleus and cytoplasm. The staining was done as follows: the sections were dipped in thionin solution (5 mg/ml; Thermo Fisher Scientific Inc, Waltham, MA, USA) for 5 s, rinsed in distilled water and differentiated in ethanol with 3 ml acetic acid for 10 s, dehydrated with ethanol, placed in xylene, and coverslipped using Entellan (Merck, Darmstadt, Germany) mounting solution.

Figure 2.

Histological cortical thickness measurements. The histological cortical thickness was measured using a Nissl staining. The average of five thickness measurements (red) equally spread in the ROI was used as the final value of cortical thickness for that ROI. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Digitization and Morphometric Analysis

On the Nissl staining, regions of interest (ROIs) were systematic random selected based on a grid with frames of 4×4 mm, starting from a random first frame. Every subsequent third frame was checked for the presence of GM and selected as an ROI only if it contained all cortical layers. Due to the random sampling, the ROIs included regions of the top of gyri, the bottom of sulci, and the middle of sulcal banks.

Depending on the size of the tissue block, 2–16 ROIs were selected. Images of the ROIs were digitized using a Leica DM/RBE photomicroscope with a Xillix MicroImager digital camera (1280×1024 pixels) attached to an MCID Elite image analysis system (Imaging Research Inc, Ontario, Canada). During digitization, the exposure conditions were kept constant. Images were taken as vertical stacks of 6 μ m at $50\times$ magnification. The thickness was measured using the MCID Elite image analysis system, by drawing a line perpendicular to the WM/GM border from the pia to the WM/GM border. The average of five thickness measurements equally spread in the ROI was used as value of cortical thickness in the respective ROI (Fig. 2).

Histological whole-brain cortical thickness

The histological whole-brain cortical thickness was calculated by taking the average cortical thickness of the analyzed cortical areas per brain.

Statistical Analysis

All analyses were performed using SPSS 20.0. Intraclass correlation coefficient (ICC) for absolute agreement was used for the intrarater reliability for the histological cortical thickness measurements, which was measured twice in

TABLE I. Clinical and radiological characteristics of the patients

Patient	Age (y)	Gender	MRI PMD (h)	Autopsy PMD (h)	MS	DD (y)	Cause of death	T2LV (ml)	NBV (L)	NGMV (L)
1	84	F	3.3	7.6	SP	50	Euthanasia	9.3	1.25	0.65
2	51	M	4.0	11	SP	20	Pneumonia	28.8	1.33	0.66
3	86	M	3.5	10.2	SP	61	Pneumonia	28.3	1.28	0.64
4	56	M	5.0	9.8	SP	14	MODS	17.4	1.32	0.65
5	71	F	2.5	7.1	SP	34	MODS	16.2	1.13	0.57
6	56	F	4.0	8.4	SP	32	Pneumonia	25.4	1.26	0.65
7	57	F	3.5	8.7	SP	25	Euthanasia	10.8	1.38	0.66
8	75	M	7.0	10.2	SP	50	Pneumonia	32.3	1.26	0.63
9	53	M	4.0	10	SP	25	Euthanasia	32.1	1.27	0.61
10	80	M	4.5	9.75	SP	45	Pneumonia	6.0	1.24	0.62
11	70	M	4.0	8.75	PP	26	MODS	6.0	1.29	0.64
12	78	M	3.0	8.75	SP	33	Euthanasia	25.0	1.22	0.62
13	59	M	4.0	9.25	SP	27	Pneumonia	19.3	1.22	0.65
14	59	M	5.0	10.75	RR	21	Euthanasia	18.7	1.22	0.62
15	60	M	2.0	7.5	SP	31	Pneumonia	6.5	1.28	0.68
16	61	F	3.5	7.5	PP	11	Cachexia	2.7	1.48	0.78

MRI PMD: postmortem delay between death and MRI; Autopsy PMD: postmortem delay between death and end of autopsy; SP: secondary progressive MS; PP: primary progressive MS; DD: disease duration; MODS: multiple organ dysfunction syndrome; T2LV: T2 lesion volume; NBV: normalized brain volume; NGMV: normalized grey matter volume. Volume measurements were done using SIENAX.

56 ROIs from 11 subjects. Differences in mean regional cortical volume between FSL-SIENAX, FreeSurfer, and SPM were assessed using paired Student's *t*-tests. Last, we analyzed the relation between MRI measures and the histological cortical thickness. To rule out the effect of the size of different anatomical regions on this analysis, the different variables were first homogenized. Specifically, MRI cortical volume (ml), MRI thickness (mm), and histological cortical thickness (mm) were converted to Z scores based on the full set of measurements for each region separately. Linear regression analyses were used to investigate the relation between MRI regional cortical volume or thickness and histological cortical thickness. Of the MRI volume and thickness measurements, we only used those obtained from the same hemisphere as the tissue blocks for histological analysis.

In addition, we also evaluated the relation between MRI whole-brain cortical volume and the histological whole-brain thickness using a linear regression analysis.

All analyses were controlled for type 1 error across test by applying a false discovery rate (FDR) correction [Benjamini and Hochberg, 1995]. In addition, residual plots were used to check for biases in the linear regression analyses.

RESULTS

Between 2010 and 2013, brain tissue was collected from 50 MS cases (mean postmortem delay between death and craniotomy: 6.5 ± 2.7 h, MS type: 38 SPMS, 10 PPMS, 1 RRMS, male/female ratio: 24/26). Sixteen of these cases met our inclusion criteria and underwent postmortem in situ MRI scanning. The inclusion criteria were clinical

diagnosis MS and a postmortem delay of <12 h. Exclusion criteria were clinical indications of comorbidities such as Alzheimer's disease. Most patients were excluded due to MRI unavailability.

From these 16 MS patients, we collected 83 tissue blocks. Fourteen of the patients had secondary progressive (SPMS), one had primary progressive (PPMS), and one relapsing-remitting MS (RRMS). Eleven patients were male. Mean age $\pm$ SD of patients with MS was (66 ± 11.6 years) and mean disease duration was 31.6 ± 13.8 years. The mean postmortem delay between death and MRI was 3.9 ± 1.1 h, while the mean postmortem delay between death and craniotomy was 9.3 ± 0.7 h. Radiological and clinical characteristics are described in Table I.

Within the 83 tissue blocks, 139 ROIs were investigated. The mean histological cortical thickness measured in the 139 ROIs was 2.76 ± 0.84 mm. The intrarater reliability for the histological thickness was 0.855 (ICC for absolute agreement).

FSL-SIENAX, FreeSurfer, and SPM Cortical Volumes and FreeSurfer Cortical Thickness

The mean cortical volume and thickness per method are described in Table II.

FSL-SIENAX with automated BET extraction had an averaged regional normalized cortical volume of 5.11 ± 3.01 ml ranging from 9.80 ± 0.95 in the superior frontal gyrus to 1.93 ± 0.25 ml in the superior temporal gyrus. Manual editing of FSL-SIENAX BET extraction changed the cortical volume significantly in the superior frontal gyrus, from 2.58 ± 0.56 ml to 2.63 ± 0.60 ml

TABLE II. Mean cortical volume and thickness per method

Outcome measure	Anatomical region				
	icg (mean $\pm$ SD)	ifg (mean $\pm$ SD)	sfg (mean $\pm$ SD)	stg (mean $\pm$ SD)	ipl (mean $\pm$ SD)
FSL-SIENAX normalized cortical volume with automated BET extraction (ml)	5.39 $\pm$ 1.14	2.58 $\pm$ 0.56	9.80 $\pm$ 0.95	1.93 $\pm$ 0.25	4.23 $\pm$ 0.49
FSL-SIENAX normalized cortical volume with manual BET extraction (ml)	5.40 $\pm$ 1.14	2.63 $\pm$ 0.60 ^a	10.47 $\pm$ 0.90 ^a	1.90 $\pm$ 0.25 ^a	4.23 $\pm$ 0.46
FreeSurfer normalized cortical volume without manual editing (ml)	4.10 $\pm$ 0.86	8.74 $\pm$ 0.98	18.38 $\pm$ 3.00	10.21 $\pm$ 1.94	10.97 $\pm$ 1.85
FreeSurfer normalized cortical volume with manual editing (ml)	4.22 $\pm$ 0.80	9.46 $\pm$ 1.51	19.59 $\pm$ 2.79 ^a	10.77 $\pm$ 1.66 ^a	11.48 $\pm$ 2.15 ^a
SPM normalized cortical volume with automated settings (ml)	6.88 $\pm$ 1.50	4.99 $\pm$ 1.07	9.05 $\pm$ 1.28	8.20 $\pm$ 1.12	8.20 $\pm$ 2.17
Cortical thickness FreeSurfer with automated settings (mm)	2.25 $\pm$ 0.37	1.76 $\pm$ 0.22	1.88 $\pm$ 0.20	1.91 $\pm$ 0.28	1.80 $\pm$ 0.17
Cortical thickness FreeSurfer with manual editing (mm)	2.30 $\pm$ 0.32	1.85 $\pm$ 0.10	1.96 $\pm$ 0.17	1.93 $\pm$ 0.19	1.84 $\pm$ 0.19
Histological cortical thickness (mm)	2.49 $\pm$ 0.84	2.53 $\pm$ 0.88	2.97 $\pm$ 0.78	2.84 $\pm$ 0.62	2.97 $\pm$ 1.02

Icg: inferior cingulate gyrus; ifg: inferior frontal gyrus; sfg: superior frontal gyrus; stg: superior temporal gyrus; ipl: inferior parietal lobule.

^aA statistically significant difference between GM volume with automated settings and GM volume with manual editing. $p < 0.05$, as determined by paired sample Student's t -test.

($p < 0.001$, FDR-corrected). There was no significant difference found in normalized regional cortical volume between FSL-SIENAX with and without manual editing in other cortical areas than the superior frontal gyrus. For FreeSurfer without manual editing of GM segmentation, the averaged regional cortical volume was 10.67 ± 5.34 ml and ranged from 18.38 ± 2.99 ml in the superior frontal gyrus to 4.10 ± 0.86 ml in the inferior cingulate gyrus. For FreeSurfer, cortical volume significantly changed ($p = 0.008$, FDR-corrected) only in the superior frontal gyrus after manual editing.

SPM was run with default parameters and resulted in an average regional normalized cortical volume of 7.53 ± 2.01 ml, ranging from 9.05 ± 1.28 ml in the superior frontal gyrus to 4.99 ± 1.07 ml in the inferior frontal gyrus. Finally, irrespective of manual editing, regional normalized cortical volume differed significantly between FreeSurfer, FSL-SIENAX, and SPM ($p < 0.05$, FDR-corrected) in all anatomical regions.

FreeSurfer was the only method with the possibility to measure cortical thickness, which did not significantly change when manual editing was applied. The average regional cortical thickness was 1.95 ± 0.30 mm, and ranged from 1.78 ± 0.22 mm in the inferior frontal gyrus to 2.25 ± 0.37 mm in the inferior cingulate gyrus without manual editing. With manual editing, the averaged regional cortical thickness was 2.00 ± 0.28 mm, and ranged from 1.85 ± 0.10 mm in the inferior frontal gyrus to 2.30 ± 0.32 mm in the inferior cingulate gyrus.

Postmortem Validation of FSL-SIENAX, FreeSurfer, and SPM

The histological cortical thickness was used to validate cortical volume, measured using FSL-SIENAX, FreeSurfer, or SPM. Table II summarizes the results. After correction for multiple comparisons, only FSL-SIENAX with manual editing of the BET extraction (std. $\beta = 0.345$, adjusted $R^2 = 0.105$, $P = 0.005$), and FreeSurfer, with manual editing of WM/GM segmentation (std. $\beta = 0.379$, adjusted $R^2 = 0.129$, $P = 0.003$) had a significant correlation with the histological cortical thickness. SPM was only run using default parameters and did not correlate with the histological cortical thickness (std. $\beta = 0.151$, adjusted $R^2 = 0.008$, $P = 0.226$). The results are illustrated in Figure 3. Comparable to the results observed for volumes, the cortical thickness measured using FreeSurfer only exhibited a significant association with the histological cortical thickness when manual editing was applied (std. $\beta = 0.381$, adjusted $R^2 = 0.130$, $P = 0.003$, FDR-corrected) (Fig. 4 and Table III).

Finally, we repeated the analyses to evaluate the possible correlations between MRI cortical volumes *not normalized* for head size, and histological cortical thickness. Regardless of manual editing, after correction for multiple comparisons, none of the methods correlated with the histological cortical thickness.

In general, SPM and FSL-SIENAX are used to measure whole brain cortical volume and not regional cortical volumes. Therefore, we evaluated the relation between total cortical volume, measured using SPM, FSL-SIENAX, and

Figure 3.

Scatterplots illustrating the correlation between MRI normalized cortical volume and histological cortical thickness for FSL-SIENAX, FreeSurfer, and SPM run with automated settings or with manual editing. 95% confidence interval is indicated with a curved line. In addition, the black $x = y$ line illustrates the perfect correlation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

FreeSurfer, and the averaged histological cortical thickness. No significant correlations between MRI whole-brain cortical volume and histological whole-brain cortical thickness were found.

DISCUSSION

This study demonstrated that regional cortical volume measured using FSL-SIENAX with manual BET extraction, and FreeSurfer with manual WM/GM segmentation, correlated significantly with histological cortical thickness in postmortem MRI of patients with multiple sclerosis. Interestingly, none of the evaluated methods significantly correlated with histological thickness in automated mode, stressing the use of manual editing to improve the automated segmentation, especially in a disease like MS in which the average GM atrophy rate is only 0.2–0.6% per year [Fisher et al., 2008]. This is quite slow because the GM atrophy rate is already up to 0.1% in healthy controls [Fisher et al., 2008]. In addition, it calls for caution when interpreting in vivo results from fully automated FreeSurfer and FSL-SIENAX analyses.

An epilepsy study by Cardinale et al. [2014] also investigated the relation between histological cortical thickness and FreeSurfer cortical thickness. Both the results from

Cardinale et al. and our results show a significant correlation between FreeSurfer cortical thickness and the histological cortical thickness, suggesting FreeSurfer is a suitable method to estimate cortical thickness [Cardinale et al., 2014]. The mean cortical thickness found by Cardinale and colleagues in 27 epileptic patients with FreeSurfer and with histological measurements were respectively, 3.65 ± 0.44 mm and 3.72 ± 0.36 mm and were not significantly different ($P = 0.32$). However, Cardinale et al. used FreeSurfer in an *automated* mode and report significant results, while only FreeSurfer with *manual* editing produced significant result in our study. This difference in results could be due to the fact that Cardinale et al. sampled a mixture of brain areas (dependent on the location of the epileptic seizure focus), while we systematically sample five standard brain areas. In addition, they were using material from epileptic patients and we were studying material from chronic MS patients. WM of MS patients is characterized by WM lesions and diffuse WM pathology. Although we corrected for the effect of WM lesions by applying lesion filling, still diffuse WM pathology can have a pronounced effect on automated GM segmentation [Westlye et al., 2009]. In addition, MS pathology is characterized by cortical thinning and ventricle widening, which can influence automated GM segmentation as well

Figure 4.

Scatterplots illustrating the correlation between MRI normalized cortical thickness and histological cortical thickness for FreeSurfer run with automated settings or with manual editing. 95% confidence interval is indicated with a curved line. In addition, the black $x = y$ line illustrates the perfect correlation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

[Djamanakova et al., 2013]. These differences therefore may be additionally responsible for the discrepancies between our findings and those of Cardinale et al.

Rosas et al. [2002] also investigated the relation between cortical atrophy determined on MRI and in tissue blocks, and found a good agreement within 0.20 mm. They did not use histology to measure the cortical thickness, and only used cortical area with low curvature. In addition, they used an in-house software tool to estimate the cortical thickness in their MRI data. As they used different methods and investigated a population with Huntington's

disease, it is difficult to compare their results to those of our present study. However, these two studies and ours illustrate that the correlation between volume and cortical thickness estimation methods and the cortical thickness measured in tissue blocks vary and the agreement is likely influenced by multiple factors such as pathology and sampling techniques.

Our results are in apparent contradiction to the results of an earlier in vivo study, from Derakshan et al. [2010], which investigated the accuracy of five volume measurement methods, including FSL-SIENAX, FreeSurfer, and

TABLE III. The relation between MRI cortical volume or thickness and histological cortical thickness

Outcome measure	Std. β	adjusted R^2	p -value
FSL-SIENAX normalized cortical volume with automated BET extraction	0.318	0.087	0.009
FSL-SIENAX normalized cortical volume with manual BET extraction	0.345	0.105	0.005*
FreeSurfer normalized cortical volume with automated settings	0.173	0.015	0.164
FreeSurfer normalized cortical volume with manual editing	0.379	0.129	0.003*
SPM normalized cortical volume with automated settings	0.151	0.008	0.226
Cortical thickness FreeSurfer with automated settings	0.161	0.011	0.197
Cortical thickness FreeSurfer with manual editing	0.381	0.130	0.003*

Std. β Standardized β for histological cortical thickness. In each case, the regression model included MRI GM volume or MRI GM thickness (FreeSurfer cortical thickness) as dependent variable and the histological thickness as independent variable. *Remained significant after correction for multiple comparisons.

Figure 5.

Three examples of regional cortical segmentation using FSL-SIENAX, FreeSurfer, and SPM. For illustrative purposes, different patients were used for FSL-SIENAX, SPM, and FreeSurfer, because the performance per method varied considerably per patient. Images A and D, respectively, FSL-SIENAX without and with manual BET. When using FSL-SIENAX with automated BET extraction significant parts of the skull are segmented as cortex, particularly in the posterior brain regions such as the parietal

and occipital cortex. Images B and E illustrate FreeSurfer without and with manual editing of GM segmentation. FreeSurfer used with automated settings, often underestimates parts of the frontal cortex. Finally image C illustrates cortical segmentation using SPM. SPM often underestimates cortical volume in the frontal and parietal cortex in our sample. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

SPM. As a reference, a manually outlined GM mask of three brain slices of three MS patients were used in that in vivo study. Interestingly, SPM cortical masks corresponded better with the manually outlined masks than FSL-SIENAX and FreeSurfer, indicating a more accurate GM segmentation. The discrepancy in results might be due to the difference in sample size and due to the fact that we measured GM volume in brain regions more spread throughout the brain, compared with the regions included in the three brain slices as used by Derakhshan et al. Visual examination of our segmentation results showed for SPM a combination of local underestimation and overestimation of cortical volume in different regions. For example, often substantial parts of the parietal and frontal cortex were not segmented (Fig. 5). Another reason for the discrepancy with the work by Derakhshan et al. is

likely the fact that we used postmortem histologically measured cortical thickness as a standard reference instead of manually edited GM volume masks.

Taking into account postmortem effects, such as an absence of blood pressure, as well as differences in age, the global grey matter volume of 0.65 L is within the range of previous work (0.52–0.83), indicating the validity of postmortem MRI volume measurements [Steenwijk et al., 2015; Lee and Prohovnik, 2008; Good et al., 2001; Taki et al., 2004].

Although the histologically measured cortical thickness is likely close to the true cortical thickness, it is still influenced by factors such as dehydration, rehydration, and the cutting directions both at autopsy and during sectioning. We tried to minimize this error by taking the average of multiple measurements per ROI and per tissue block,

including a range of topologically different tissue, like the tops of gyri but also the bottoms of sulci. In addition, this systematic error applies to every comparison made between MRI measurement methods and the histological cortical thickness and does not change whether FSL-SIENAX, FreeSurfer, or SPM is used as dependent variable in the linear regression.

Both the MRI data and the histological data had a relatively high variation, which could in part explain the relatively low explained variance in the regression models. Although we tried to keep the variation to a minimum, still multiple factors contribute to this high variation, like biological variability and cause of death, but also tissue processing like fixation, nonorthogonal cutting direction, dehydration, and rehydration. Nevertheless, there was still a significant association of FSL-SIENAX and FreeSurfer measures with histological cortical thickness.

Because gyri were not removed as intact anatomical structures but instead parts of larger gyri were often selected, we could not measure the histological cortical volume (dependent on the presence of the total anatomical volume) and instead used histological cortical thickness as the primary outcome measure. By contrast, the MRI methods generated GM volumes and for FreeSurfer also cortical thickness, and to quantify these in specific anatomical regions, an atlas had to be applied. As a result of differences in the size of these atlas regions, the MRI GM volume measured by a given method varies first and foremost by anatomical region. Therefore, to be able to compare histological cortical thickness and MRI cortical volume, all values were converted to Z-scores based on the data for the corresponding anatomical region. By converting the values to Z-scores, all variables (both MRI and histology variables) were represented on the same standardized scale, enabling us to analyze the relation of MRI cortical volume with histological cortical thickness.

The statistical analysis method was chosen to accommodate the small number of cases available in this study (even after the long inclusion period, described in the results). Due to the low number of patients with postmortem *in situ* MRI data, statistical power to take dependency of the samples adequately into account was lacking, and we have treated the multiple Z score converted measures from a single patient as if they were independent in the linear regression analyses. Future, larger studies will probably have to consider these measures as independent.

CONCLUSION

In postmortem MS brain imaged *in situ*, FSL-SIENAX cortical volumes with manual editing, and manually edited FreeSurfer cortical thicknesses and volumes correlated with the histological cortical thickness. There were no significant correlations with histological cortical thickness for FSL-SIENAX, and FreeSurfer without manual editing, nor for SPM. In addition, although it is a

time-consuming procedure, manual editing of the BET nonbrain tissue removal incorporated in FSL-SIENAX and manual editing of FreeSurfer's GM segmentation improves MRI cortical GM volume and MRI cortical thickness measurements considerably.

ACKNOWLEDGMENT

The authors thank the MS-MRI Autopsy team and the Netherlands Brain Bank (<http://www.brainbank.nl/>) for their help in acquiring the data.

REFERENCES

- 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.
- Benedict RH, Ramasamy D, Munschauer F, Weinstock-Guttman B, Zivadinov R (2009): Memory impairment in multiple sclerosis: Correlation with deep grey matter and mesial temporal atrophy. *J Neurol Neurosurg Psychiatry* 80:201–206.
- Benjamini Y, Hochberg Y (1995): Controlling the false discovery rate - A practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57:289–300.
- Cardinale F, Chinnici G, Bramero M, Mai R, Sartori I, Cossu M, Lo Russo G, Castana L, Colombo N, Caborni C, De Momi E, Ferrigno G (2014): Validation of FreeSurfer-estimated brain cortical thickness: Comparison with histologic measurements. *Neuroinformatics* 12:535–542.
- Chard DT, Jackson JS, Miller DH, Wheeler-Kingshott CA (2010): Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *J Magn Reson Imag* 32:223–228.
- Dale AM, Fischl B, Sereno MI (1999): Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 9:179–194.
- Derakhshan M, Caramanos Z, Giacomini PS, Narayanan S, Maranzano J, Francis SJ, Arnold DL, Collins DL (2010): Evaluation of automated techniques for the quantification of grey matter atrophy in patients with multiple sclerosis. *NeuroImage* 52:1261–1267.
- Djamanakova A, Faria AV, Hsu J, Ceritoglu C, Oishi K, Miller MI, Hillis AE, Mori S (2013): Diffeomorphic brain mapping based on T1-weighted images: Improvement of registration accuracy by multichannel mapping. *J Magn Reson Imag* 37:76–84.
- Durand-Dubief F, Belaroussi B, Armspach JP, Dufour M, Roggerone S, Vukusic S, Hannoun S, Sappey-Marini 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. *AJNR* 33:1918–1924.
- Fischl B, Sereno MI, Dale AM (2008): Gray matter atrophy in multiple sclerosis: A longitudinal study. *Ann Neurol* 64:255–265.
- Fisher E, Lee JC, Nakamura K, Rudick RA (1999): Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9:195–207.
- Geurts JJ, Stys PK, Minagar A, Amor S, Zivadinov R (2009): Gray matter pathology in (chronic) MS: Modern views on an early observation. *J Neurol Sci* 282:12–20.

- Giorgio A, Battaglini M, Smith SM, De Stefano N (2008): Brain atrophy assessment in multiple sclerosis: Importance and limitations. *Neuroimaging Clin N Am* 18:675–686, xi.
- Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS (2001): A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 14:21–36.
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012): *Fsl*. *NeuroImage* 62:782–790.
- Lee H, Prohovnik I (2008): Cross-validation of brain segmentation by SPM5 and SIENAX. *Psychiatry Res* 164:172–177.
- Popescu V, Battaglini M, Hoogstrate WS, Verfaillie SC, Sluimer IC, van Schijndel RA, van Dijk BW, Cover KS, Knol DL, Jenkinson M, Barkhof F, de Stefano N, Vrenken H (2012): Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. *Neuroimage* 61:1484–1494.
- Popescu V, Klaver R, Voorn P, Galis-de Graaf Y, Knol D, Twisk J, Versteeg A, Schenk G, Van der Valk P, Barkhof F, De Vries H, Vrenken H, Geurts J (2015): What drives MRI-measured cortical atrophy in multiple sclerosis? *Multiple sclerosis (Houndmills, Basingstoke, England)*.
- Rosas HD, Liu AK, Hersch S, Glessner M, Ferrante RJ, Salat DH, van der Kouwe A, Jenkins BG, Dale AM, Fischl B (2002): Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58:695–701.
- Sanfilipo MP, Benedict RH, Sharma J, Weinstock-Guttman B, Bakshi R (2005): The relationship between whole brain volume and disability in multiple sclerosis: A comparison of normalized gray vs. white matter with misclassification correction. *Neuroimage* 26:1068–1077.
- Seewann A, Kooi EJ, Roosendaal SD, Barkhof F, van der Valk P, Geurts JJ (2009): Translating pathology in multiple sclerosis: The combination of postmortem imaging, histopathology and clinical findings. *Acta Neurol Scand* 119:349–355.
- Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, De Stefano N (2002): Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 17:479–489.
- Steenwijk MD, Geurts JJ, Daams M, Tijms B, Wink AM, L JB, Tewarie PK, Uitdehaag BM, Barkhof F, Vrenken H, Pouwels PJ (2015): Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant.
- Taki Y, Goto R, Evans A, Zijdenbos A, Neelin P, Lerch J, Sato K, Ono S, Kinomura S, Nakagawa M, Sugiura M, Watanabe J, Kawashima R, Fukuda H (2004): Voxel-based morphometry of human brain with age and cerebrovascular risk factors. *Neurobiol Aging* 25:455–463.
- Westlye LT, Walhovd KB, Dale AM, Espeseth T, Reinvang I, Raz N, Agartz I, Greve DN, Fischl B, Fjell AM (2009): Increased sensitivity to effects of normal aging and Alzheimer's disease on cortical thickness by adjustment for local variability in gray/white contrast: A multi-sample MRI study. *NeuroImage* 47:1545–1557.