

The Effect of Hypointense White Matter Lesions on Automated Gray Matter Segmentation in Multiple Sclerosis

Rose Gelineau-Morel,^{1,2} Valentina Tomassini,^{1,3*} Mark Jenkinson,¹
Heidi Johansen-Berg,¹ Paul M. Matthews,^{1,4,5} and Jacqueline Palace¹

¹Oxford Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom

²Baylor College of Medicine, Houston, Texas, USA

³Department of Neurology and Psychiatry, "Sapienza" University, Rome, Italy

⁴GSK Clinical Imaging Centre, GlaxoSmithKline, London, United Kingdom

⁵Department of Clinical Neuroscience, Imperial College, London, United Kingdom

Abstract: Previous imaging studies assessing the relationship between white matter (WM) damage and matter (GM) atrophy have raised the concern that Multiple Sclerosis (MS) WM lesions may affect measures of GM volume by inducing voxel misclassification during intensity-based tissue segmentation. Here, we quantified this misclassification error in simulated and real MS brains using a lesion-filling method. Using this method, we also corrected GM measures in patients before comparing them with controls in order to assess the impact of this lesion-induced misclassification error in clinical studies. We found that higher WM lesion volumes artificially reduced total GM volumes. In patients, this effect was about 72% of that predicted by simulation. Misclassified voxels were located at the GM/WM border and could be distant from lesions. Volume of individual deep gray matter (DGM) structures generally decreased with higher lesion volumes, consistent with results from total GM. While preserving differences in GM volumes between patients and controls, lesion-filling correction revealed more lateralised DGM shape changes in patients, which were not evident with the original images. Our results confirm that WM lesions can influence MRI measures of GM volume and shape in MS patients through their effect on intensity-based GM segmentation. The greater effect of lesions at increasing levels of damage supports the use of lesion-filling to correct for this problem and improve the interpretability of the results. Volumetric or morphometric imaging studies, where lesion amount and characteristics may vary between groups of patients or change over time, may especially benefit from this correction. *Hum Brain Mapp* 33:2802–2814, 2012. © 2011 Wiley Periodicals, Inc.

Key words: MRI; multiple sclerosis; white matter lesions; gray matter segmentation; lesion-filling; brain atrophy

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Multiple Sclerosis Society UK; Contract grant number: 829/05

Contract grant sponsors: MS Society Italy (V.T.'s Multiple Sclerosis Society Fellowship 2005–2006); UK NIHR Oxford Biomedical Research Centre;

*Correspondence to: Valentina Tomassini, MD, PhD, FMRIB Centre, University of Oxford, Nuffield Department of Clinical

Neurosciences, John Radcliffe Hospital, OX3 9DU Oxford. E-mail: valentt@fmrrib.ox.ac.uk

Received for publication 14 December 2010; Revised 28 April 2011; Accepted 9 June 2011

DOI: 10.1002/hbm.21402

Published online 5 October 2011 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Multiple Sclerosis (MS) is a chronic inflammatory disease of the central nervous system causing axonal demyelination and neuronal degeneration [Lassmann, 2010]. Previous studies have shown a strong correlation between white matter (WM) lesion volume and gray matter (GM) atrophy [Ceccarelli et al., 2008; De Stefano et al., 2001; Henry et al., 2008; Ramasamy et al., 2009; Sepulcre et al., 2006; Tao et al., 2009], prompting investigations on the relationship between GM and WM damage [Bendfeldt et al., 2010; Ceccarelli et al., 2009; Henry et al., 2009; Herz et al., 2010]. While most of these studies have assumed a pathological cause for GM atrophy in MS, very few have investigated possible methodological reasons to explain changes in GM volume with increasing WM damage [Chard et al., 2002, 2010; Chen et al., 2004; Nakamura and Fisher, 2009]. These studies have suggested that WM lesions may affect mean WM intensity values and shift segmentation boundaries, thus leading to inaccuracies in the GM segmentation considered for atrophy quantification. Therefore, they have recommended solutions to correct for WM lesion effects on GM segmentation. One study addressed this by using a scaling factor to adjust segmented GM volumes [Nakamura and Fisher, 2009], while another developed a technique to fill lesions with normal appearing white matter (NAWM) intensities prior to automated brain segmentation [Chard et al., 2010]. The method of lesion-filling was able to correct the misclassification of lesion voxels in simulated images and provided a more accurate quantification of overall GM volumes [Chard et al., 2010].

The aim of this study is to extend the lesion-filling approach from simulated images to images of MS brains. As MS lesions can change over time [Bagnato et al., 2003; Ciccarelli et al., 1999], with gender [Pozzilli et al., 2003], disease types [Nijeholt et al., 1998] and severity of damage [Brex et al., 2000], accounting for these changes and their effects on GM segmentation in MS patients seems crucial to improve the interpretability of quantitative imaging measures of damage, as well as our understanding of disease processes and intervention effects [Barkhof et al., 2009].

Here, using a lesion-filling method, we first quantified and localized the effect of WM lesions on segmented GM volumes both in the simulation images, where lesions were morphologically similar in size, shape, intensity and location to MS lesions, and in a cohort of relapsing-remitting MS (RRMS) patients. Second, we assessed the effect of WM lesions on the segmentation of deep gray matter (DGM) structures of MS patients. The proximity of these structures to WM lesions may directly impact their segmentation and may allow us to disentangle the neighboring effect of lesions on segmentation from the influence of lesions on mean WM intensity values. Third, we tested the impact of lesion-filling correction on differences between patients and healthy volunteers in volumetric and mor-

phometric GM measures, which is relevant to determine the importance of this correction in clinical studies.

METHODS

Image Acquisition

Structural images were acquired in 23 RRMS patients (18 women, mean±SD age: 43.08 ± 2.7 years; median EDSS 4.0, range, 0–7.0) and 12 healthy volunteers (9 women, mean±SD age: 30.1 ± 5.3 years) using a 1.5 Tesla Siemens Sonata scanner, with a maximum gradient strength of 40 mT m^{-1} . A T1-weighted 3D FLASH image (T1-WI) (TR = 12 ms, TE = 5.65 ms, flip angle = 19° , with elliptical sampling of k space, giving a voxel size of $1 \times 1 \times 1 \text{ mm}^3$) and a turbo spin echo sequence [TR = 3,000 ms, TE = 22/87 ms, FOV = 256 mm, slice thickness 3 mm, giving proton density (PD-WI) and T2-weighted images (T2-WI)] were acquired in a single session in each subject.

MS Lesion Masks in Patients

An experienced researcher (V.T.) identified brain hypointense and hyperintense WM lesions on T1-WI and PD-WI, respectively. To increase confidence in lesion detection and masking on T1-WI, the correspondence of lesions between the T1-WI and PD-WI scans was verified. All further work with lesion volumes and lesion masks was based entirely on the lesions outlined on the T1-WI.

MS lesions were outlined in each patient using a semi-automated lesion segmentation tool (Jim, Version 5.0, Xinapse Systems Ltd., Northants, UK) (Fig. 1). Inaccuracies in lesion segmentation were corrected by manually refining the lesion contour on the T1-WI. Total lesion volume was determined in patients by calculating the volume of their binarized lesion mask.

Patient Lesion-Filled Images

MS lesions were filled by replacing the lesion voxel intensity values with values that were randomly sampled from an intensity distribution that was measured from the surrounding WM voxels (Fig. 1), like the method used by Battaglini et al. (personal correspondence). Similar methods of lesion filling have been used previously to correct for effects of lesions on GM volumes [Chard et al., 2010] and to improve nonlinear registration of MS brains with healthy appearing brains [Sdika and Pelletier, 2009]. Briefly, the histogram of intensities from GM and WM voxels (as segmented by FSL's FAST, <http://www.fmrib.ox.ac.uk/fsl/fast4/>, [Zhang et al., 2001]) neighboring a WM lesion (i.e., one voxel deep) was generated (neighboring histogram). Then samples were taken from this histogram using a uniform random value passed through an interpolated version of the empirical cumulative distribution function of the neighboring histogram (linearly

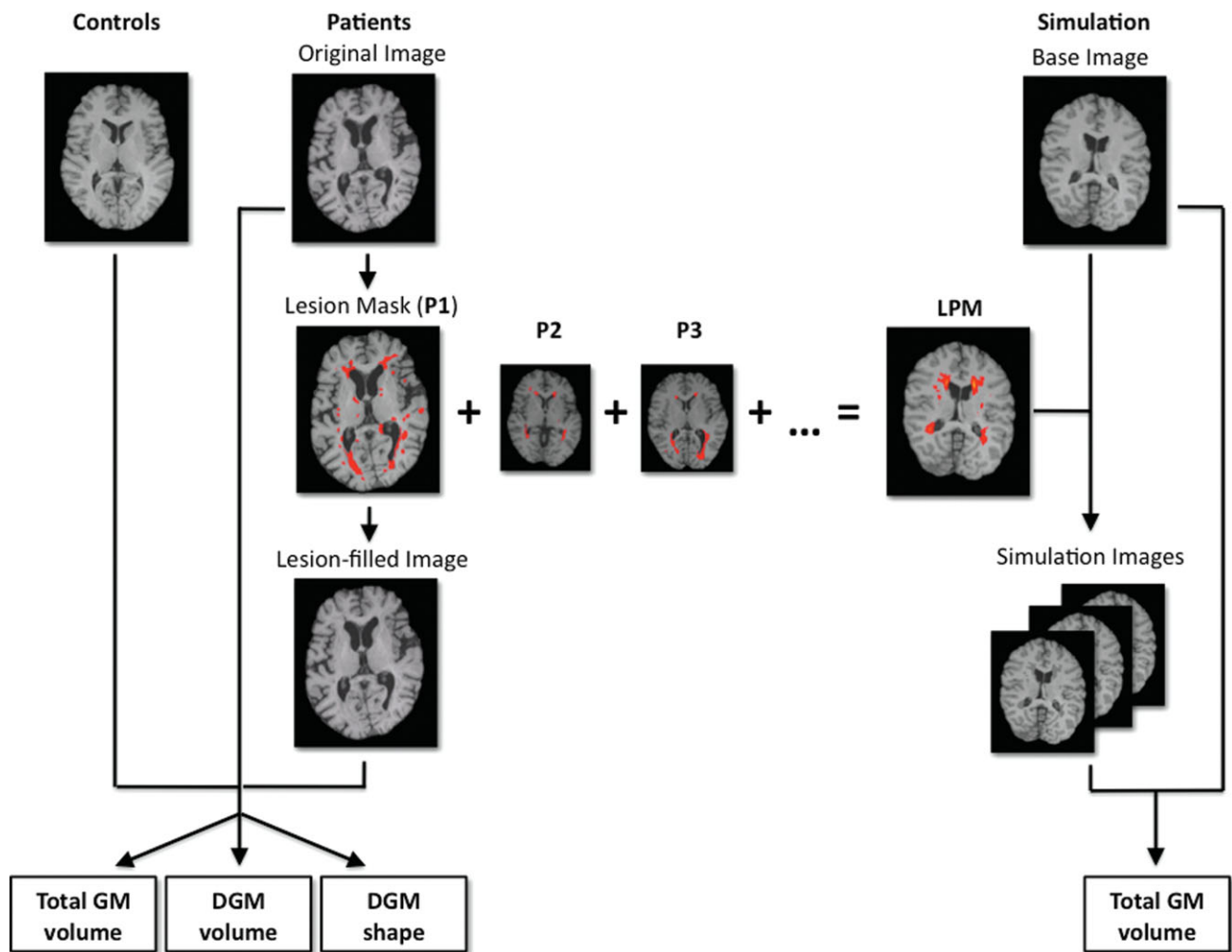


Figure 1.

Left: Flow diagram showing the steps adopted for data analysis. In each patient, we segmented total GM and DGM structures before and after lesion-filling. We quantified the volume of these structures and parameterised their shapes to test the effects of lesion-filling on volumetric and morphometric GM measures. Healthy volunteers' scans were used as control group for comparisons with patients before and after lesion-filling (i.e., patient original images and patient lesion-filled images) to assess the

interpolating between every fifth value). These samples were used as the intensities to fill the interior lesion voxels. Only neighboring GM and WM voxels, as opposed to distant WM voxels, were used in order to avoid the effects of inhomogeneities in the magnetic field, which are very slight over the range of sizes of the individual lesions but can be considerable across the whole brain. Both GM and WM voxels were included in the neighboring histogram in order to represent the surrounding tissue and allow the filled lesions to best visually blend into their environment.

The one-voxel layer at the inside edge of each lesion was also filled, but with a different method. In this case, the original

impact of lesion-filling in clinical studies. Right: Simulation images with WM areas morphologically similar to MS lesions were created from the scan of a healthy volunteer (base image) using patient LPM. We compared total GM volumes of base and simulation images to quantify the effects of various WM lesion volumes and intensities on segmented GM volume. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

intensities were replaced by the mean value of the nearest neighboring voxels, either from the voxels outside the lesion or from the previously filled interior lesion voxels, but not from other inside edge voxels. This method ensured that a smooth intensity gradient was created from the intensities outside the lesion to those of the interior lesion voxels and avoided an abrupt change in intensity from one voxel to the next.

Examples of lesions from various locations of WM before and after lesion-filling are provided in Figure 1. The lesion-filled image was then used for total GM segmentation and subcortical GM segmentation (Fig. 1), as described in detail below.

Lesion Probability Map

Individual lesion masks were used to produce a lesion probability map (LPM), in which each voxel represented the proportion of patients with a WM lesion at that voxel (see Fig. 1). This was achieved by registering individual T1-WI to MNI standard space using four levels of FSL's nonlinear image registration tool, FNIRT (<http://www.fmrib.ox.ac.uk/fsl/fnirt>, [Klein et al., 2009]). The registration of individual T1-WI to standard space was optimized by masking out lesions and by using the intensity mapping to account for intensity inhomogeneities in the image. The final warp obtained from individual T1-WI registration was applied to register the individual lesion mask to MNI standard space. Each registered lesion mask was thresholded at 50% to include in the final lesion mask only voxels surrounded by at least 50% lesion voxels. This provided a more conservative estimate of lesion volume and prevented neighboring GM voxels from being included in the WM lesion mask. All of the final registered and thresholded individual lesion masks were added together and divided by the total number of patients to create a LPM, in which each voxel intensity represented the proportion of patients with a WM lesion at that voxel. This LPM was used in subsequent steps of the analysis.

Simulation Images

To simulate patients with hypointense WM lesions of different volumes and intensities, we created simulation images by varying lesion volumes and lesion intensities within the LPM generated on MS patients (Fig. 1).

The LPM was nonlinearly registered using FNIRT to a healthy brain T1-WI, which was randomly selected to be the base structural image for the simulation image. The LPM was then thresholded at 13 levels in order to create 13 simulated lesion masks, which contained different lesion volumes (0–34,373 mm³), approximating those found in our patients (1,425–33,300 mm³). These simulated lesion masks were binarized and multiplied by six intensity values, which represented increasing percentages of the average WM intensity of simulation image. The percentages of WM were determined by calculating the mean intensity of the WM, as segmented by FSL's FAST (<http://www.fmrib.ox.ac.uk/fsl/fast4/>, [Zhang et al., 2001]), in the T1-WI used as the base image. In order to mirror the average lesion intensities determined from our patient group, and therefore reflect typical lesion pathology in MS patients, we took a thresholded LPM and homogeneously filled it with intensities set to 68%, 74%, 79%, 84%, 89% and 95% of the WM intensity of the simulation image.

Each thresholded and multiplied LPM was used to create a simulation image, in which the intensity of the base image voxels within the LPM was replaced with the intensity of the thresholded and multiplied LPM voxels. This created 78 simulation images with various lesion volumes and intensities.

Total GM Volume

Skull-stripped simulation, patient, and healthy volunteer images were segmented with FAST [Zhang et al., 2001], which uses a hidden Markov Random Field model and an Expectation-Maximization algorithm to segment brain tissue. We created three separate partial volume maps for GM, WM, and cerebrospinal fluid (CSF), in which each voxel value represented the proportion of that tissue class in that voxel. The total volume for each tissue class was obtained by summing the partial volume estimates over all voxels for each class separately. One patient was excluded due to insurmountable inaccuracies in image segmentation owing to extensive brain atrophy and ventricular enlargement. This left 22 patients and 12 controls which were used for the remainder of the study. For both the simulation and patient images, any lesion voxels misclassified as GM in the tissue-segmentation were removed prior to analyzing GM volumes by masking out voxels contained within the lesion mask from the segmented GM tissue.

The percent (%) difference in GM volume between images with and without WM lesions was calculated for both the simulation images (with versus without lesions) and the patient images (original versus lesion-filled) to determine the effects of WM lesions on GM volumes. This was done using the following equation:

$$\text{GM volume difference} = [(GMV_L - GMV_{\emptyset L}) / GMV_{\emptyset L}] \times 100,$$

where GMV_L represents the total GM volume from the image with lesions (either simulation image or original patient image) and $GMV_{\emptyset L}$ represents the total GM volume from the image with no lesions (either simulation image without lesions or lesion-filled patient image). Using this formula, a negative % GM volume difference would indicate a smaller GM volume when lesions are present in the image than with no lesions in the image, while a positive percent GM volume change would indicate a larger GM volume with lesions in the image than with no lesions in the image.

As automated segmentation tools are known to have difficulty in segmenting cerebellar GM [Datta et al., 2009; Lewis and Fox, 2004; Suckling et al., 1999], we tested whether the addition of cerebellar GM to cerebral GM volume could introduce some bias due to tissue misclassification. We first drew a mask of the cerebellum in the simulation image and calculated total GM volumes both with and without the inclusion of the cerebellar GM volume. Then, we calculated the percent GM volume difference and determined that the impact of lesion-filling on percent GM volume difference was independent from the inclusion of cerebellum. After ensuring similar pattern of results in patients, we included cerebellar volumes in our patient total GM volume calculation to provide more clinically informative results of the impact of WM lesions on GM segmentation. Simulation image results are reported without cerebellar GM volumes.

To compare the effects of lesions in patients and simulated images, we drew polynomial trendlines demonstrating the change in GM volume with increasing lesion volume at each of the six simulated lesion intensities ($R^2 = 0.999$ for each lesion intensity). For each patient, we chose the simulated lesion intensity which most closely approximated that of the patient's lesions and used the polynomial trendline equation to determine the expected GM volume change given the patient's lesion volume. We divided the patient's percent GM volume change by the expected percent GM volume change to calculate how much of the expected GM volume change was produced by lesion-filling in patients.

DGM Volume and Shape

To examine the effects of lesion-filling on differences between patients and controls in volume and shape of DGM, patient images (original and lesion-filled), as well as healthy volunteer images, were analysed using FIRST. The FIRST tool is an integrated registration and segmentation tool in FSL (<http://www.fmrib.ox.ac.uk/fsl/first>, [Patenaude et al., 2011]) that uses a model of shape and intensity, built from a training-set of manually segmented images, to guide the segmentation of DGM structures on T1-weighted images. Briefly, in FIRST individual images are registered to MNI standard space using an affine transformation with 12° of freedom. A second stage of registration applies a MNI subcortical mask to exclude voxels that are located outside the subcortical regions. Once in MNI space, FIRST segments each DGM structure using a multivariate Gaussian model that contains shape and intensity information for each structure based on the vertex locations and intensity profiles associated with a deformable mesh. The mesh has a fixed number of vertices for each structure and point-correspondence is maintained across subjects. To obtain an image representation, FIRST fills all voxels within the mesh and then performs boundary correction for voxels that the mesh passes through, using an optimized, structure-specific, voxel-wise classification method [Patenaude et al., 2011].

Volume Analysis

To represent the differences in GM volume between the original and the lesion-filled images of patients, the segmented lesion-filled image was subtracted from the segmented original image to produce a GM volume difference image for each DGM structure, as shown below:

$$\text{DGM volume difference} = \text{DGMV}_L - \text{DGMV}_{\emptyset L}$$

where DGMV_L represents the DGM volume from the segmented original image and $\text{GMV}_{\emptyset L}$ represents the DGM volume from the segmented lesion-filled image. In this difference image, voxels with a positive intensity represented

voxels included in the segmented original image, but not in the segmented lesion-filled image. Voxels with a negative intensity represented voxels included in the segmented lesion-filled image, but not in the segmented original image. For each structure, we explored the relationship between DGM volume difference and WM lesion volume using a Pearson correlation, with P -values < 0.05 considered significant.

A post-hoc analysis of the correlation between DGM volume difference and WM lesion volume was performed after thresholding the volume of the difference image at an intensity of 0. This thresholding selected voxels contained either in the segmented original image (voxel intensity of difference image >0) or in segmented lesion-filled image (voxel intensity of difference image <0). This analysis was done to disentangle the effect of increasing WM lesion volume on DGM volumes under the alternative hypotheses that increasing WM lesion volume may result in (Hypothesis 1) a decrease in DGM volume, due to the effect of lesions on the intensity threshold used for GM-WM segmentation, or in (Hypothesis 2) an increase in DGM volume, due to the effect of neighboring hypointense lesions surrounding DGM structures being included in the DGM segmentation.

Within-group comparisons of patient DGM volumes from original and lesion-filled images were performed using a two-tailed paired t -test. Between-group comparisons of DGM volumes were performed using a two-tailed unpaired t -test. Significance was attributed to P -values < 0.05 .

Shape Analysis

To compare shapes of DGM structure between patients and controls before and after lesion-filling in patients, we performed a multivariate test at each vertex of each DGM structure mesh. This vertex analysis tested whether the locations of the vertices were significantly different between groups. After performing a correction for multiple comparisons using False Discovery Rate (FDR) [Benjamini and Yekutieli, 2005], the multivariate F-statistic for vertex differences was displayed on a mean surface mesh using surface coloration for location of shape differences in each structure. In these areas, vectors indicated the direction of shape differences between groups.

RESULTS

Effect of WM Lesions and Their Filling on Total GM Volume

In the simulation image, the creation and filling of WM areas morphologically similar to MS lesions allowed us to test the effect of different lesion volumes and intensities on segmentation of total GM. Increasing WM lesion volume at all the physiological intensities tested was

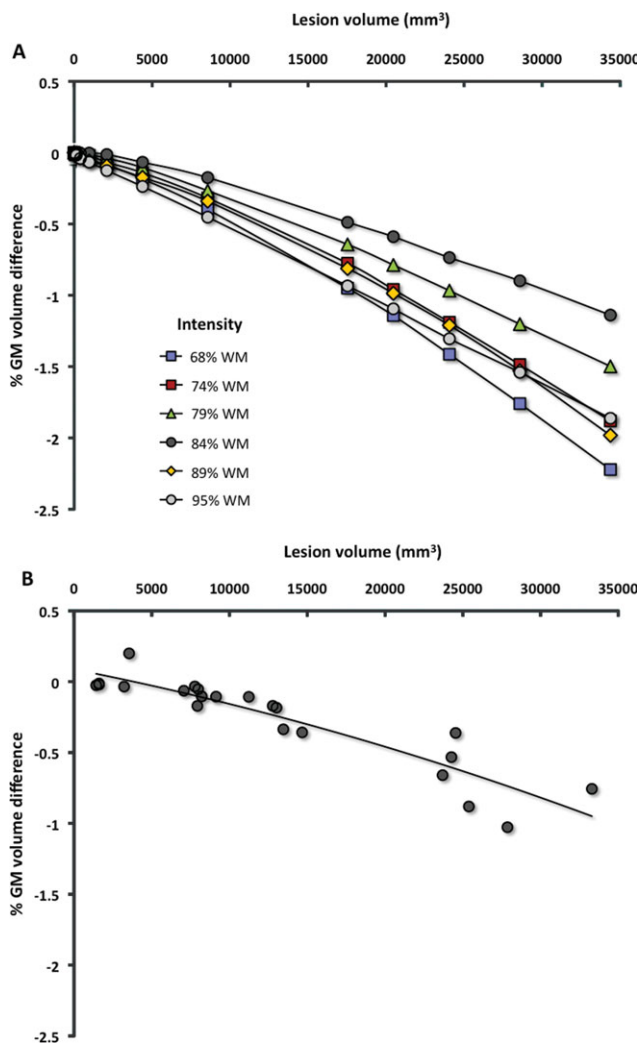


Figure 2.

A: Relationship between WM lesion volume and % GM volume difference in simulation images. % GM volume difference was calculated as the difference between the images without and the images with WM lesions at varying volumes and intensities. Lesions misclassified as GM were not included in the GM volume quantification. Increasing lesion volumes, at all intensities, were associated with decreasing % GM volume difference in a nonlinear relationship. **B:** Relationship between % GM volume difference between original and lesion-filled image in patients with increasing lesion volumes. Polynomial trendline in black ($R^2 = 0.84$) demonstrates the relationship between lesion volume and % GM volume difference. In all patients but one, the GM volume was smaller in the original image than in the lesion-filled image. Patients with greater lesion volumes generally had a larger change in GM volume with lesion-filling. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

associated with decreasing GM volume. Specific levels of hypointensity exerted a greater effect on GM. WM lesions with intensity values equal to 68% of the mean intensity of

WM showed the greatest effect on GM volume, while hypointense WM lesions with 84% of the WM mean intensity had the least effect (Fig. 2A). While all lesion intensities showed this pattern, there was no linear relationship between the intensities of WM lesions and their effect on GM volume. % GM volume difference varied from about 0% (for intensity of 84% of WM and lesion volume 58 mm³) to -2.22% (for intensity of 68% WM and lesion volume 34373 mm³) depending on the lesion volume and intensity, with an average % GM volume difference of -0.55%.

All patients except one showed negative % GM volume difference, i.e., smaller GM volume in the original than in the lesion-filled image (Fig. 2B). As with the simulation image, patients with higher lesion volume had larger changes in GM volume between the original and the lesion-filled image (Fig. 2B). However, the average % GM volume difference in patients was only -0.26% (range, -0.02 to -1.02%), less than that found in the simulation image. When individual patient images are compared to the simulation image with the most similar lesion intensity and the same lesion volume, the percent change in GM volume produced in patients after lesion-filling is, on average, 71.5% of that produced by the most similar simulation image. In other words, if adding lesions of a certain volume and intensity to a simulation image led to 5% reduction in total GM volume, patients with lesions of the same volume and intensity would only have a 3.58% ($5\% \times 0.715 = 3.58\%$) reduction in total GM volume compared to their lesion-filled image.

When the location of GM volume differences was categorised, GM voxels at the border of the cortical GM with WM, as well as voxels in the DGM were less likely to be segmented as GM in the presence of lesions. These effects were true both in the simulation image (Fig. 3A) and in

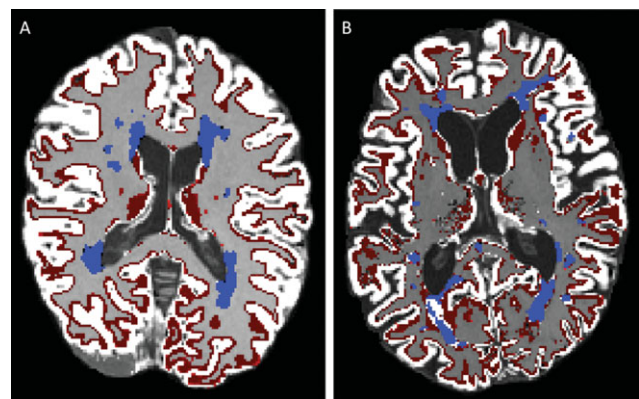


Figure 3.

Brain-extracted T1-WI scans showing the location of lower estimated GM volume (in red) in the presence of WM lesions (in blue), either with simulation image (**A**) or in original patient image (**B**). The location of GM volume difference was largely in areas along the cortical GM/WM boundary and in the DGM structures both in simulation and in original patient images. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

patients (Fig. 3B), with more extensive effects when the lesion volume was greater.

Effect of Lesions and Their Filling on DGM Volume

In patients, all DGM structures except the hippocampus had smaller average volumes in the original images than in the lesion-filled images, regardless of lesion volume. The hippocampus was the only structure that had a higher average volume in the original image than in the lesion-filled image (Fig. 4A).

Since the DGM structures did not have the same consistent direction of correlation with lesion volume (Fig. 4A) that we found with the total GM volumes (Fig. 2B) likely due to the dual effect of lesions on mean intensity threshold for segmentation (Hypothesis 1), as well as their proximity to DGM structures (Hypothesis 2), we explored the relationships between lesion volume and the volume of DGM difference image when either negative or positive voxels only were considered. The volume of difference image containing only negative voxels were significantly negatively correlated with the lesion volume in the caudate ($r = -0.43$, $P < 0.05$), pallidum ($r = -0.45$, $P < 0.05$), hippocampus ($r = -0.67$, $P = 0.001$), and amygdala ($r = -0.525$, $P = 0.01$), with a trend towards a significant correlation in the putamen ($r = -0.42$, $P = 0.05$). Consistent with total GM volume results in simulation images and in patients, this relationship indicated that with increasing lesion volumes, less GM volume was included in the segmented DGM of the original image (Fig. 4B).

The volume of difference image containing only positive voxels showed a significant positive correlation with lesion volume only in the caudate ($r = 0.59$, $P < 0.005$) and hippocampus ($r = 0.57$, $P = 0.005$) (Fig. 4C), suggesting that significantly more GM was included in the original image of the caudate and hippocampus in patients with greater lesion volumes.

Impact of Lesion-Filling on Within-Group and Between-Group Comparisons of GM Volume and Shape

Lesion-filling significantly affected both total GM and DGM volume. Total GM volume in the original patient images was significantly smaller than total GM volume in the lesion-filled images ($P = 0.001$) (Fig. 5). DGM volumes in the original patient images were significantly smaller than volumes in the lesion-filled images ($P < 0.05$) in all structures except for the hippocampus ($P = 0.2$), with a trend towards a significant difference in the amygdala ($P = 0.06$) (Fig. 5).

When the total GM volumes in patients before and after lesion-filling were compared with the control group, it was found that patients had a significantly smaller GM volume than controls when either the original images ($P < 0.01$) or the lesion-filled images ($P < 0.01$) were used. Simi-

larly, the comparison of DGM volumes between controls and patients both in the original images and after lesion-filling showed significantly smaller DGM structure volumes in patients for all the structures ($P < 0.05$) except for the thalamus ($P > 0.2$) (Fig. 5).

Vertex analysis demonstrated that DGM shape differences between groups were dependent on whether lesions were corrected. For example, while the original images showed an almost equal extent of shape differences in the right and left caudate when compared to controls, the comparison of the patients' lesion-filled images to controls revealed more lateralized shape differences, with the left caudate showing much greater differences in shape than the right caudate when compared to controls (Fig. 6).

DISCUSSION

This study confirms that hypointense WM lesions can cause an apparent reduction of GM volume where there is none. In simulation images, total GM volumes decrease with greater lesion volumes, with certain lesion intensities having a larger effect. A similar effect of WM lesions on total GM segmentation is confirmed in MS patients. However, for similar lesion volume and intensity, the impact of lesion-filling correction on GM volume is lower in patients than would be predicted from simulation. In both simulation images and patients, the location of the tissue segmentation error occurs at the GM/WM boundary and can be distant from lesions. Two opposing mechanisms can explain this misclassification, as demonstrated by DGM structures. Changes in intensity threshold of WM due to lesions can cause a decrease in volumes of DGM structures with increasing lesion volume. However, WM lesions can also increase the volumes of DGM structures in the original image through a neighboring effect. Therefore, lesion-filling correction can significantly affect volumetric and morphometric measures of DGM structures, influencing results of comparisons within patients, as well as between patients and controls.

The first part of this study investigated the effect of WM lesions at different volumes and intensities within the range found in a cohort of RRMS patients. The finding of decreasing GM volumes with increasing lesion volumes, both in simulation and in patients, is likely due to changes in the tissue classification of intermediate intensity voxels between the GM and WM [Chard et al., 2010; Nakamura and Fisher, 2009]. The presence of hypointense areas in WM can lead to a decrease in the average WM intensity and thus reduce the intensity threshold used to segment GM and WM. GM voxels then have a reduced probability of being classified as such, leading to an overall reduction of GM volume throughout the brain. This explanation is suggested by previous findings demonstrating that GM and WM tissue intensity differences at the cortical GM/WM interface can affect measures of cortical GM thickness in MS patients [Chen et al., 2004]. This explanation is also supported by our results showing that the location of

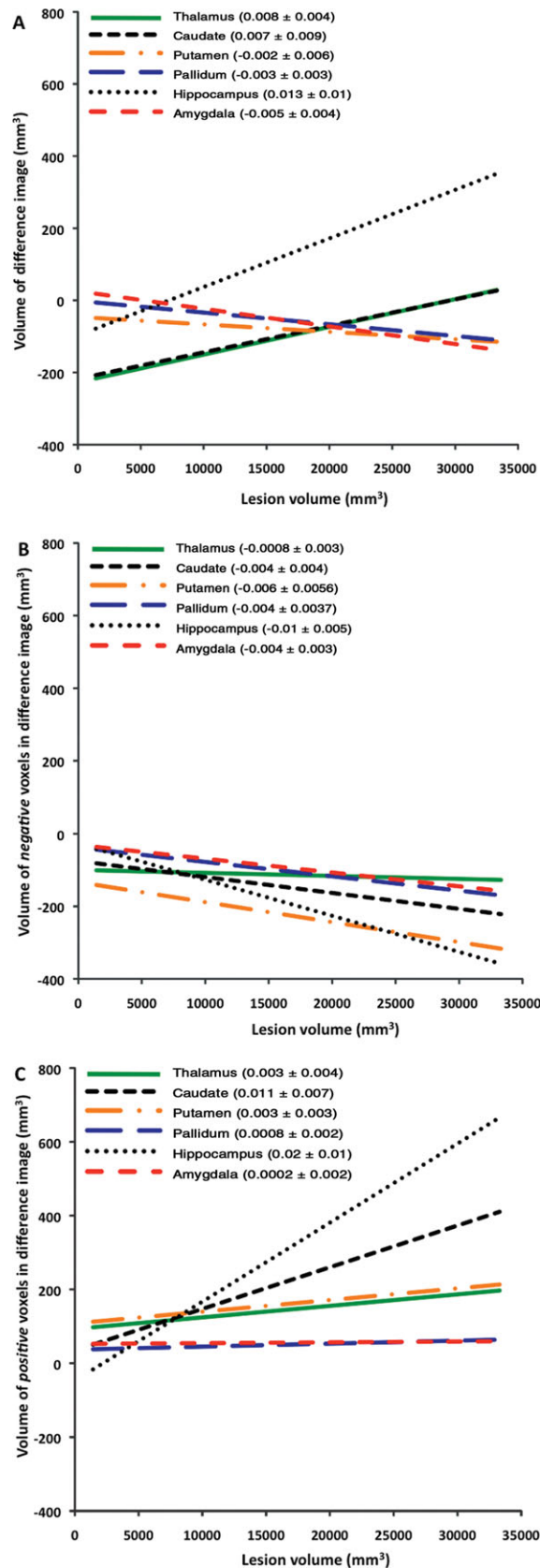


Figure 4.

A: Relationship between lesion volume and difference in DGM volumes from original to lesion-filled image in patients with heterogeneous lesion intensities. Negative values indicate a smaller DGM volume in the original image than in the lesion-filled image. While all structures except for the hippocampus have smaller volumes in the original image, there was no consistent relationship between DGM volume and lesion volume across GM structures. In the legend to the plot, slopes and 95% confidence intervals are shown in brackets beside structure name. **B:** Post-hoc analysis performed to test WM lesion impact on the intensity threshold used to segment brain tissue classes (Hypothesis 1). Greater volume of negative voxels indicates a smaller DGM volume in original image than in the lesion-filled image. All DGM structures except for the thalamus showed a significant negative correlation between lesion volume and volume of negative voxels in difference image ($P < 0.05$). Slopes and 95% confidence intervals are shown in brackets beside structure name. **C:** Post-hoc analysis performed to isolate the effect of neighboring WM lesions on DGM segmentation (Hypothesis 2). Greater volume of positive voxels indicates a larger GM volume in segmented original image than in lesion-filled image. The caudate and the hippocampus show a significant positive correlation between lesion volume and volume of positive voxels in the difference image ($P < 0.01$). Slopes and 95% confidence intervals are shown in brackets beside structure name. Supporting Information Figures 1S, 2S, and 3S show plots of individual structures. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

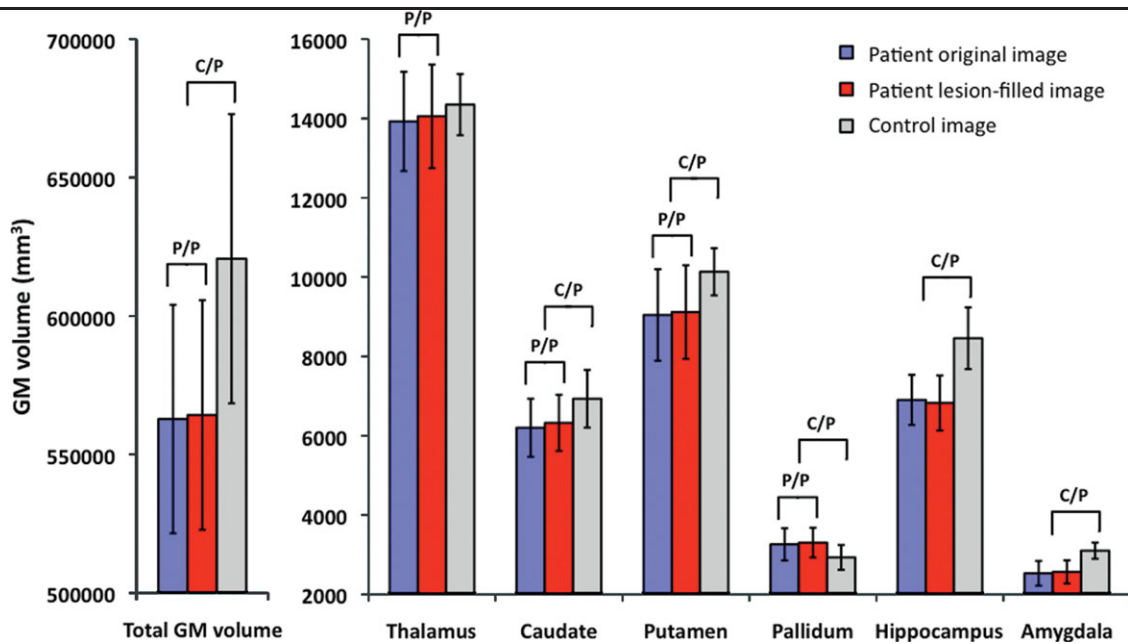


Figure 5.

GM volumes in controls and in patients before and after lesion-filling. Significant difference ($P < 0.05$) between groups is shown in brackets above columns. P/P indicates significant difference between original and lesion-filled images of patients using paired t -

test. C/P indicates significant difference between control group and both patient groups using unpaired t -test. Error bars represent standard deviation for each group. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

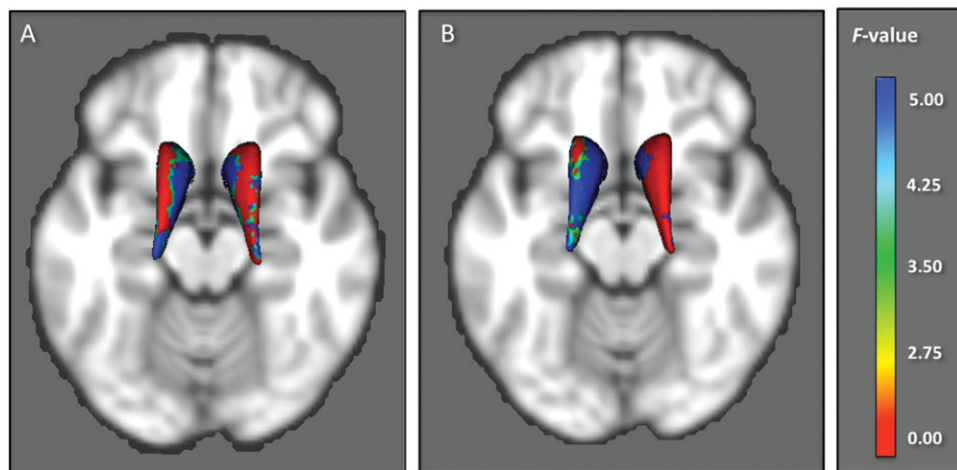


Figure 6.

Results of vertex analysis to assess between-group localized differences in the shape of the caudate nuclei. In this structure, the control group has been compared with the patient group before (original image) (A) and after (B) lesion-filling. Surface coloration shows the multivariate F -statistic for the vertex difference between groups, with higher values (more blue) indicating greater significance. These values are corrected for multiple

comparisons using FDR. The between-group comparison using lesion-filled images shows more lateralized shape differences in the left caudate, while the comparison using the original images shows a more bilateral and symmetric distribution of shape changes when compared to controls. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

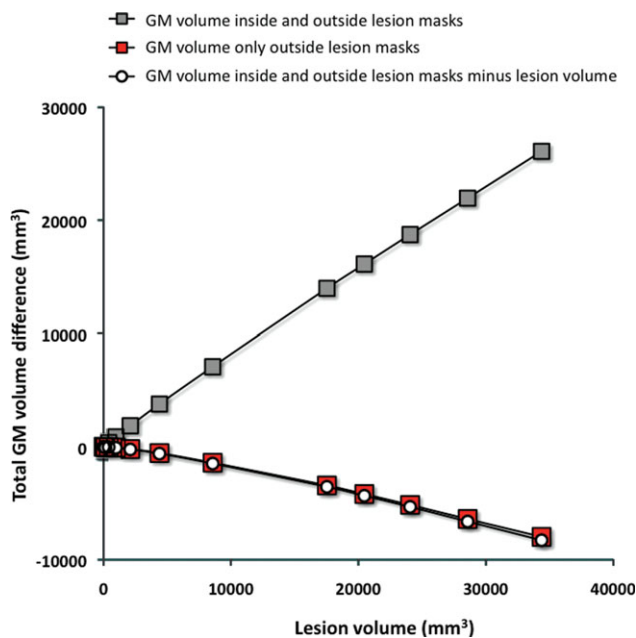


Figure 7.

Effect of misclassified lesional voxels on the relationship between total GM volume difference and increasing lesion volume in simulation. WM lesions were simulated using 74% of WM intensity (approximately equivalent to GM intensity). Red squares show results when misclassified lesions were excluded from total GM volume. Gray squares show results when lesions misclassified as GM were not excluded from the total GM volume calculation and agree with the results reported by Chard et al. [2010] for a similar lesion intensity. When lesion volume is removed from the total GM volume including misclassified lesions (gray squares), the resulting curve (white circles) agrees with the results produced when misclassified lesions are excluded from total GM volume (red squares). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Similar effects of lesions on GM volume were found in patients, where higher lesion volume was associated with a greater change in segmented GM volume when their lesions were filled. However, the extent of change was not as great as expected from the simulation, with images from patients showing about 72% of the percent GM volume change seen in the simulation images with similar lesion intensities and volumes. Both methodological and biological reasons could contribute to explain this smaller effect of lesion-filling in patients when compared to the simulation.

Areas of dirty appearing white matter not included in the lesion mask can affect the mean intensity threshold and thus the segmentation results [Chen et al., 2003; Vrenken et al., 2010]. The presence of these regions, found only in patients, could lead to underestimation of GM volumes even after lesion-filling and thus contribute to the smaller effect of lesion-filling correction on GM volumes in patients. Any hypointensities outside the simulated lesions that may have been in the control brain used for simulation would have

affected every simulation image equally, and therefore would not have influenced the GM volume difference resulting from the presence of lesions.

Hypointense areas surrounding a lesion could also affect the lesion-filling process, as this technique uses neighboring WM intensities to determine the intensity histogram used in filling the lesions. If the lesions were filled with hypointense WM intensities, instead of those of NAWM, this would not adequately remove the effects of lesions on GM segmentation, and may lead to a residual underestimation of GM volumes. While using a more global WM intensity histogram could help to address this problem, this method could be influenced by intensity inhomogeneities in the scan, producing an intensity histogram that is not a good representation of the WM intensities at each lesion. A better and simpler solution would be to ensure that more areas of hypointensity surrounding lesions are included within the lesion mask. Although this would still not correct for GM segmentation differences induced by diffuse hypointensities such as dirty appearing white matter, it might remove the effects of hypointensities surrounding WM lesions.

The occurrence of GM pathology and its relationship with WM damage, which are present in the patients as a result of MS pathology [Tomassini and Palace, 2009], but not in simulation images, should be taken into account when interpreting the patients' results. In MS, GM tissue loss can develop independently from [Geurts and Barkhof, 2008] or gradually following the appearance of WM lesions [Miller et al., 2002]. Pathological processes affecting GM in patients can alter the intensity of GM tissue and thus shift the boundary of GM/WM segmentation. If GM intensity were reduced as a result of pathology, this would partially negate the effect of lesion-filling and may explain the smaller impact of lesion-filling correction on GM volumes in patients compared to simulation.

Our study extends previous findings on the relationship between WM lesions and apparent reduction in GM volume to the investigation of DGM structures. We found that WM lesions could exert opposing effects on DGM. The caudate, thalamus and hippocampus showed increasing volumes, while putamen, pallidum, and amygdala showed decreasing volumes in original image with increasing WM lesion volume.

Our post-hoc analyses suggested that two opposing, but not mutually exclusive, mechanisms exist through which lesions affect GM segmentation, and both contribute to explain these results. A reduction in the intensity threshold used for GM/WM segmentation could explain the reduction in the DGM volume of the original image with increasing lesion volumes. As a change in the intensity threshold should lead to a global reduction in GM volumes, we would expect that it would affect every DGM structure equally. Our results support this, with all structures except the thalamus showing a significant correlation between lesion volume and the amount of GM volume loss in the DGM segmentation of the original image. As there is a low

contrast gradient between the GM and WM at the border of the thalamus, especially in the face of GM pathology in MS, the thalamic shape model in FIRST likely drives the structure segmentation in this region, thus making the thalamus less susceptible to changes in the intensity threshold that affect the GM/WM segmentation in other structures.

We also found a significant correlation between lesion volume and additional DGM volume in the original image segmentation of the caudate and hippocampus. We hypothesized that this effect could be due to the presence of neighboring lesions bordering these DGM structures and thus artificially increasing their volumes. This effect is not evident when examining total GM volumes, which included mostly cortical GM, but can become pronounced in those DGM structures bordering the lateral ventricle, which is most surrounded by WM lesions, especially at higher lesion volumes. In fact, this effect appears to be so strong in the hippocampus that it outweighs the former effect of the reduced intensity threshold used in GM segmentation, leading to an overall increase in hippocampal volume with increasing lesion volume.

The direct comparison of GM volumes in MS patients before and after lesion-filling confirmed the significant effect that lesions can have on GM segmentation by showing significantly smaller total GM volumes in the original image compared to the lesion-filled image. Except for the hippocampus and amygdala, DGM volumes showed a similar significant difference, suggesting that lesion-filling prior to segmentation is needed to more accurately represent GM volumes.

To assess the plausibility and relevance of our results, we compared patients before and after lesion-filling correction with healthy volunteers. This comparison confirmed that lesion-filling maintains the expected reduction in total and DGM volumes in patients compared to controls. While the use of lesion-filling does lead to a significantly higher GM volume in patients as described above, it would be concerning if there were no longer any difference between the lesion-filled patient group and the control group as many previous studies have reported histological signs of GM damage in MS patients [Amadio et al., 2010; Bo et al., 2003; Cifelli et al., 2002; Geurts et al., 2007; Wegner et al., 2006].

As shape analysis was developed to be more sensitive to local changes in DGM structures, it is encouraging that the effects of lesion-filling are evident on shape analysis. In this specific case, the use of lesion-filling revealed a more lateralized difference in caudate shape between controls and patients, which had not been evident prior to lesion-filling. The improvements in shape analysis provided by lesion-filling further encourage the use of this technique in between-group or between-session comparisons.

CONCLUSION

Our results demonstrate that hypointense WM lesions in MS patients can lead to significant apparent reductions in

total GM and DGM volumes, unless adequately corrected for by lesion-filling methods. While these results cannot be directly applied to other segmentation methods, any segmentation method that does not entirely exclude lesions and uses intensity values to determine GM/WM boundaries may be affected to some extent, as the intensities of the lesions will affect the determination of the boundaries. This effect of lesions becomes more evident at increasing lesion volumes, making the application of correction methods important for the interpretability of the results in patients with higher burden of pathology. While the use of lesion-filling can improve the accuracy of GM measures in all cases of automated GM segmentation, lesion-filling would be especially relevant when comparing groups of patients with similar levels of GM loss but different amount of WM lesions (e.g., RRMS vs. primary progressive groups). Correction through lesion-filling before GM segmentation (as opposed to simple removal of misclassified lesions after GM segmentation) would be also important in longitudinal studies, in which patients are monitored over periods of time with potential changes in their lesion volume and/or lesion intensities [Bagnato et al., 2003]. Pharmacological studies affecting lesion volume or altering lesion intensity [Di Rezze et al., 2007] would benefit from this correction to ensure the interpretability of their results.

Future investigations into the effects of dirty appearing WM on automated GM segmentation in MS patients are required to determine the extent to which more diffusely hypointense areas could affect GM segmentation and lead to overestimation of GM atrophy. In addition, other neurological conditions such as cerebrovascular accidents can be explored to determine if WM hypointensities could contribute to apparent GM volume loss [Appelman et al., 2010].

ACKNOWLEDGMENTS

The authors wish to thank people with MS and healthy volunteers who kindly participated in the study. They thank the MS Society UK (V.T., P.M.M.) and Italy (V.T.), the MS International Federation (www.msif.org) (V.T.) and the UK NIHR Oxford Biomedical Research Centre (J.P. and V.T.) for their generous support. P.M.M. is a full-time employee of GlaxoSmithKline, which is engaged in the development of new drugs for MS.

REFERENCES

- Amadio S, Montilli C, Magliozzi R, Bernardi G, Reynolds R, Volonte C (2010): P2Y12 receptor protein in cortical gray matter lesions in multiple sclerosis. *Cereb Cortex* 20:1263–1273.
- Appelman AP, Vincken KL, van der Graaf Y, Vlek AL, Witkamp TD, Mali WP, Geerlings MI (2010): White matter lesions and lacunar infarcts are independently and differently associated with brain atrophy: The SMART-MR study. *Cerebrovasc Dis* 29:28–35.

- Bagnato F, Jeffries N, Richert ND, Stone RD, Ohayon JM, McFarland HF, Frank JA. (2003): Evolution of T1 black holes in patients with multiple sclerosis imaged monthly for 4 years. *Brain* 126(Part 8):1782–1789.
- Barkhof F, Calabresi PA, Miller DH, Reingold SC (2009): Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. *Nat Rev Neurol* 5:256–266.
- Bendfeldt K, Blumhagen JO, Egger H, Loetscher P, Denier N, Kuster P, Traud S, Mueller-Lenke N, Naegelin Y, Gass A, Hirsch J, Kappos L, Nichols TE, Radue EW, Borgwardt SJ (2010): Spatio-temporal distribution pattern of white matter lesion volumes and their association with regional grey matter volume reductions in relapsing-remitting multiple sclerosis. *Hum Brain Mapp* 31:1542–1555.
- Benjamini Y, Yekutieli D (2005): Quantitative trait Loci analysis using the false discovery rate. *Genetics* 171:783–790.
- Bo L, Vedeler CA, Nyland HI, Trapp BD, Mork SJ (2003): Subpial demyelination in the cerebral cortex of multiple sclerosis patients. *J Neuropathol Exp Neurol* 62:723–732.
- Brex PA, Parker GJ, Leary SM, Molyneux PD, Barker GJ, Davies CA, Thompson AJ, Miller DH (2000): Lesion heterogeneity in multiple sclerosis: A study of the relations between appearances on T1 weighted images, T1 relaxation times, and metabolite concentrations. *J Neurol Neurosurg Psychiatry* 68:627–632.
- Ceccarelli A, Rocca MA, Pagani E, Colombo B, Martinelli V, Comi G, Filippi M (2008): A voxel-based morphometry study of grey matter loss in MS patients with different clinical phenotypes. *Neuroimage* 42:315–322.
- Ceccarelli A, Rocca MA, Valsasina P, Rodegher M, Pagani E, Falini A, Comi G, Filippi M (2009): A multiparametric evaluation of regional brain damage in patients with primary progressive multiple sclerosis. *Hum Brain Mapp* 30:3009–3019.
- 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 Imaging* 32:223–228.
- Chard DT, Parker GJ, Griffin CM, Thompson AJ, Miller DH (2002): The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology. *J Magn Reson Imaging* 15:259–267.
- Chen JT, Narayanan S, Collins DL, Smith SM, Matthews PM, Arnold DL (2004): Relating neocortical pathology to disability progression in multiple sclerosis using MRI. *Neuroimage* 23:1168–1175.
- Chen SC, Chung HW, Liou M (2003): Measurement of volumetric lesion load in multiple sclerosis: Moving from normal- to dirty-appearing white matter. *AJNR Am J Neuroradiol* 24:1929–1930.
- Ciccarelli O, Giugni E, Paolillo A, Mainero C, Gasperini C, Bastianello S, Pozzilli C (1999): Magnetic resonance outcome of new enhancing lesions in patients with relapsing-remitting multiple sclerosis. *Eur J Neurol* 6:455–459.
- Cifelli A, Arridge M, Jezard P, Esiri MM, Palace J, Matthews PM (2002): Thalamic neurodegeneration in multiple sclerosis. *Ann Neurol* 52:650–653.
- Datta S, Tao G, He R, Wolinsky JS, Narayana PA (2009): Improved cerebellar tissue classification on magnetic resonance images of brain. *J Magn Reson Imaging* 29:1035–1042.
- De Stefano N, Narayanan S, Francis GS, Arnautelis R, Tartaglia MC, Antel JP, Matthews PM, Arnold DL (2001): Evidence of axonal damage in the early stages of multiple sclerosis and its relevance to disability. *Arch Neurol* 58:65–70.
- Di Rezze S, Gupta S, Durastanti V, Millefiorini E, Pozzilli C, Bagnato F (2007): An exploratory study on interferon beta dose effect in reducing size of enhancing lesions in multiple sclerosis. *Mult Scler* 13:343–347.
- Geurts JJ, Barkhof F (2008): Grey matter pathology in multiple sclerosis. *Lancet Neurol* 7:841–851.
- Geurts JJ, Bo L, Roosendaal SD, Hazes T, Daniels R, Barkhof F, Witter MP, Huitinga I, van der Valk P (2007): Extensive hippocampal demyelination in multiple sclerosis. *J Neuropathol Exp Neurol* 66:819–827.
- Henry RG, Shieh M, Amirbekian B, Chung S, Okuda DT, Pelletier D (2009): Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. *J Neurol Sci* 282(1-2):61–66.
- Henry RG, Shieh M, Okuda DT, Evangelista A, Gorno-Tempini ML, Pelletier D (2008): Regional grey matter atrophy in clinically isolated syndromes at presentation. *J Neurol Neurosurg Psychiatry* 79:1236–1244.
- Herz J, Zipp F, Siffrin V (2010): Neurodegeneration in autoimmune CNS inflammation. *Exp Neurol* 225:9–17.
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV (2009): Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46:786–802.
- Lassmann H (2010): Axonal and neuronal pathology in multiple sclerosis: what have we learnt from animal models. *Exp Neurol* 225:2–8.
- Lewis EB, Fox NC (2004): Correction of differential intensity inhomogeneity in longitudinal MR images. *Neuroimage* 23:75–83.
- Miller DH, Barkhof F, Frank JA, Parker GJ, Thompson AJ (2002): Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. *Brain* (Part 8):1676–1695.
- Nakamura K, Fisher E (2009): Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. *Neuroimage* 44:769–776.
- Nijeholt GJ, van Walderveen MA, Castelijns JA, van Waesberghe JH, Polman C, Scheltens P, Rosier PF, Jongen PJ, Barkhof F (1998): Brain and spinal cord abnormalities in multiple sclerosis. Correlation between MRI parameters, clinical subtypes and symptoms. *Brain* 121(Part 4):687–697.
- Patenaude B, Smith S, Kennedy D, Jenkinson M (2011): A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56:907–922.
- Pozzilli C, Tomassini V, Marinelli F, Paolillo A, Gasperini C, Bastianello S (2003): “Gender gap” in multiple sclerosis: magnetic resonance imaging evidence. *Eur J Neurol* 10:95–97.
- Ramasamy DP, Benedict RH, Cox JL, Fritz D, Abdelrahman N, Hussein S, Minagar A, Dwyer MG, Zivadinov R (2009): Extent of cerebellum, subcortical and cortical atrophy in patients with MS: A case-control study. *J Neurol Sci* 282(1-2):47–54.
- 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.
- Sepulcre J, Sastre-Garriga J, Cercignani M, Ingle GT, Miller DH, Thompson AJ (2006): Regional gray matter atrophy in early primary progressive multiple sclerosis: A voxel-based morphometry study. *Arch Neurol* 63:1175–1180.
- Suckling J, Sigmundsson T, Greenwood K, Bullmore ET (1999): A modified fuzzy clustering algorithm for operator independent

- brain tissue classification of dual echo MR images. *Magn Reson Imaging* 17:1065–1076.
- Tao G, Datta S, He R, Nelson F, Wolinsky JS, Narayana PA (2009): Deep gray matter atrophy in multiple sclerosis: A tensor based morphometry. *J Neurol Sci* 282(1-2):39–46.
- Tomassini V, Palace J (2009): Multiple sclerosis lesions: Insights from imaging techniques. *Expert Rev Neurother* 9:1341–1359.
- Vrenken H, Seewann A, Knol DL, Polman CH, Barkhof F, Geurts JJ (2010): Diffusely abnormal white matter in progressive multiple sclerosis: In vivo quantitative MR imaging characterization and comparison between disease types. *AJNR Am J Neuroradiol* 31:541–548.
- Wegner C, Esiri MM, Chance SA, Palace J, Matthews PM (2006): Neocortical neuronal, synaptic, and glial loss in multiple sclerosis. *Neurology* 67:960–967.
- Zhang 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.