

# Evaluating and Reducing the Impact of White Matter Lesions on Brain Volume Measurements

Marco Battaglini,<sup>1</sup> Mark Jenkinson,<sup>2</sup> and Nicola De Stefano<sup>1\*</sup>

<sup>1</sup>Department of Neurological and Behavioral Sciences, University of Siena, Italy

<sup>2</sup>Department of Clinical Neurology, Oxford University Centre for Functional MRI of the Brain (FMRIB), University of Oxford, United Kingdom

---

**Abstract:** MR-based measurements of brain volumes may be affected by the presence of white matter (WM) lesions. Here, we assessed how and to what extent this may happen for WM lesions of various sizes and intensities. After inserting WM lesions of different sizes and intensities into T1-W brain images of healthy subjects, we assessed the effect on two widely used automatic methods for brain volume measurement such as SIENAX (segmentation-based) and SIENA (registration-based). To explore the relevance of partial volume (PV) estimation, we performed the experiments with two different PV models, implemented by the same segmentation algorithm (FAST) of SIENAX and SIENA. Finally, we tested potential solutions to this issue. The presence of WM lesions did not bias measurements for registration-based method such as SIENA. By contrast, the presence of WM lesions affected segmentation-based brain volume measurements such as SIENAX. The misclassification of both gray matter (GM) and WM volumes varied considerably with lesion size and intensity, especially when the lesion intensity was similar to that of the GM/WM interface. The extent to which the presence of WM lesions could affect tissue-class measures was clearly driven by the PV modeling used, with the mixel-type PV model giving a lower error in the presence of WM lesions. The tissue misclassification due to WM lesions was still present when they were masked out. By contrast, refilling the lesions with intensities matching the surrounding normal-appearing WM ensured accurate tissue-class measurements and thus represents a promising approach for accurate tissue classification and brain volume measurements. *Hum Brain Mapp* 00:000–000, 2011. © 2011 Wiley-Liss, Inc.

**Key words:** brain atrophy; white matter lesions; SIENA; SIENAX

---

## INTRODUCTION

The development of computational methods that, using conventional magnetic resonance (MR) images, are able to provide sensitive and reproducible measures of brain volumes, has allowed an indirect quantification of the brain.

These methods have been extensively used in the study of multiple sclerosis (MS) to estimate total and regional (i.e., white matter [WM] and gray matter [GM]) cerebral tissue loss, providing measures able to accurately assess and monitor the pathologic evolution of the disease [Battaglini et al., 2009; Bendfeldt et al., 2009; Chard et al., 2004; Chen et al., 2004; De Stefano et al., 2003].

Recently, much effort has been dedicated to improve the efficiency of automated segmentation algorithms for MR images [Nakamura and Fisher, 2009; Sdika and Pelletier, 2009], particularly in relation to the influence that the presence of WM lesions, such as those found in the brain of MS patients, can have in the measurement of tissue specific brain volumes [Chard et al., 2010; Nakamura and Fisher, 2009; Sdika and Pelletier, 2009]. In one recent study

---

\*Correspondence to: Nicola De Stefano, MD PhD, Department of Neurological and Behavioral Sciences, University of Siena, Viale Bracci 2, 53100 Siena, Italy. E-mail: destefano@unisi.it

Received for publication 14 June 2010; Revised 6 March 2011; Accepted 11 April 2011

DOI: 10.1002/hbm.21344

Published online in Wiley Online Library (wileyonlinelibrary.com).

[Sdika and Pelletier, 2009], WM lesions were shown to distort the output of non-linear registration, and the filling of the lesions with the intensity of the normal-appearing neighbor voxels appeared as an effective solution for this bias. In another recent study [Nakamura and Fisher, 2009], a new segmentation algorithm was developed and tested. This was able to calculate GM volumes avoiding the misclassification of WM lesions by using a combination of intensity, anatomical and morphological probability maps. The work pointed out that, in the presence of WM lesions, GM could be linearly underestimated (and consequently the WM overestimated), even when the misclassification of lesions was avoided, due to a misclassification of voxels with overlapping intensities [Nakamura and Fisher, 2009]. Finally, Chard et al. [2010] confirmed that lesions with an intermediate intensity between GM and WM produce an underestimation of GM volumes. They also investigated the effect of filling lesions before the segmentation, using intensities sampled from a single, global WM distribution, with the mean and standard deviation equal to that of the original WM.

Given the limited resolution of MR images and the irregular shape of brain tissue interfaces, the accuracy of any segmentation method in assigning partial volume (PV) voxels to a single tissue type is inherently limited [Niessen et al., 1999] and PV classification models need to be used [Santago and Gage, 1995; Van Leemput et al., 2003]. However, although it may be true that PV models fail to provide adequate tissue classification in the presence of WM lesions [Nakamura and Fisher, 2009], it is less certain to what extent this affects the measurements of global and regional brain volume. To address these issues, we performed the present study with the aims (i) to assess how and to what extent the presence of WM lesions of various sizes and intensities can affect brain volume measurements using a segmentation-based approach such as SIENAX [Smith et al., 2002], and with different PV models; (ii) to assess whether these issues hold for a registration-based approach such as SIENA [Smith et al., 2001]; (iii) to propose a robust and practical approach to solve the issues related to the presence of WM lesions in tissue classification analysis.

## METHODS

To estimate the accuracy of performing total and tissue type brain volume measurements in the presence of WM lesions that vary in size and intensity, we needed to work with brain volume measures with a known ground truth. Thus, we selected five normal T1-weighted three-dimensional gradient echo (T1-W) images (FFE, flip angle = 40, TR/TE = 35/10 ms, 256 × 256 matrix, 1 signal average, 250 × 250 mm<sup>2</sup> field of view, 50 contiguous 3-mm slices) of five healthy subjects obtained by using a Philips Gyroscan operating at 1.5 T (Philips Medical Systems, Best, The Netherlands), and six binarized lesion masks previously created from T2-weighted (T2-W)/proton-density (PD) MR

images of six different MS patients with different lesion loads (6, 12, 18, 24, 30 and 60 cm<sup>3</sup>). We then “created” lesions in the images using these six lesion masks, on each of the five “original” T1-W images, thus creating 30 “artificial” images with binarized regions of interest (ROIs) having a known and relatively wide range of volumes.

To fill the ROIs of each of the 30 “artificial” images with intensity values of cerebrospinal fluid (CSF), CSF/GM interface, GM, and GM/WM interface, this strategy was followed:

- i. Each “original” T1-W image of the healthy controls was hard segmented and binarized maps of WM, GM, and CSF were created.
- ii. Each of these binarized maps was multiplied by the original T1-W image, and the mean and standard deviation within the WM, GM, and CSF tissue masks were calculated.
- iii. Four different Gaussian intensity distributions were generated for each “original” image: CSF, CSF/GM, GM, GM/WM (Fig. 1- top). They had an intensity mean equal to the mean intensity of the CSF, CSF/GM interface, GM and GM/WM interface, respectively. The GM/WM and CSF/GM intensity means were defined as the average of the GM and WM mean values or the average of the CSF and GM mean values, respectively. The standard deviations of each of the intensity distributions were calculated by dividing the standard deviation of each tissue or tissue mixel types (i.e., tissue made by a mixture of different tissues) by 4, in order to obtain a narrow range around the mean. In this case, the GM/WM and CSF/GM standard deviations were set to be the difference between the GM and WM means, divided by 4, or the difference of the CSF and GM means, divided by 4, respectively.
- iv. Finally, ROIs of each of the 30 “artificial” images were filled by randomly extracting voxel values from each of the four different intensity (Gaussian) distributions (Fig. 1- bottom), creating a total of 120 “artificial” T1-W images.

## WM Lesions and Segmentation-Based Brain Volume Measurements

At this stage, brain volume measurements were obtained using a segmentation-based algorithm (SIENAX) to get values of normalized brain volume (NBV), normalized white matter volume (NWMV) and normalized gray matter volume (NGMV) from the “original” 5 T1-W images of healthy controls, as well as from the 120 “artificial” images with various lesion size, load, and intensities. Each “original” T1-W image was then compared to the corresponding 24 “artificial” images derived from it, to obtain changes in NBV, NWMV, and NGMV between the two images.

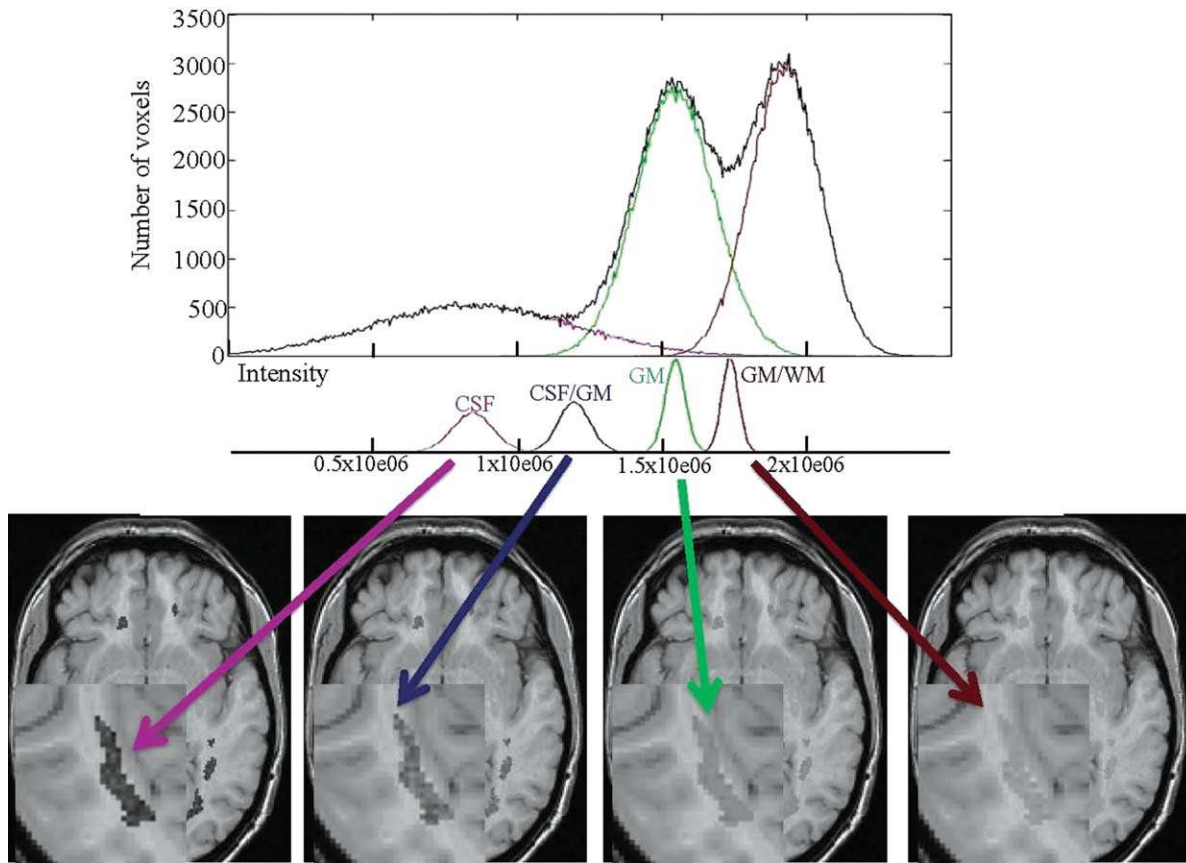


Figure 1.

Illustrative example of the creation of four different Gaussian intensity distributions (CSF, CSF/GM, GM, and GM/WM) from the histogram intensity of an “original” image (top). The binarized regions of interest of the “artificial” images (previously created

from the “original” image and a given lesion mask, see methods for details) were then filled by randomly extracting voxel values from each of the four different Gaussian intensity distributions (Fig. 1- bottom).

### WM Lesions and Registration-Based Brain Volume Measurements

Brain volume measurements were also performed using a linear registration-based algorithm (SIENA), able to obtain values of percentage brain volume change (PBVC) in two subsequent MR scans. In this two-time point approach, each “original” T1-W image of each healthy control was used as the first time point, and each of the corresponding 24 “artificial” images, containing the above-described variety of lesion load and intensity, was used as the follow-up scan.

### “Masked-Out” or “Refilled” WM Lesions and Segmentation-Based Brain Volume Measurements

Once the error due to lesion misclassification during brain volume measurements was assessed and quanti-

fied, possible strategies for minimizing this error were tested. Two “adjusted” T1-W images were created from each of the 30 “artificial” images with a binarized ROI of known and wide range (6, 12, 18, 24, 30, and 60 cm<sup>3</sup>) for a total of 60 additional “adjusted” T1-W images. The first set of 30 “adjusted” T1-W images was obtained by simply masking out the ROIs from the original T1-W images. The second set was created by refilling each two-dimensional lesion with intensities derived from a histogram that was matched closely to the histogram of the WM surrounding the lesion, obtained using a non-uniformly sampled histogram. The latter method was chosen since there may be only a small number of voxels immediately neighboring the lesions, and so a non-uniformly sampled histogram is better adapted to the available data. No “a-priori” choice of intensity distribution was imposed.

The following steps provide details of how the ROIs were refilled:

- i. An ROI ( $R_L$ ) was selected by using 2-D dilation and defined as the ROI comprising the voxels that are immediate neighbors of the binarized lesion mask ( $L$ ) and belong to the WM. From  $R_L$ , the number of voxels ( $N_R$ ) and their mean intensity ( $M_R$ ) were calculated. The intensity histogram of  $R_L$  was then constructed with the number of bins ( $n_{bins}$ ) equal to 10 if  $N_R$  was bigger than 40 and equal to round ( $N_R/4$ ) if lower. The bins were all of equal width. Finally, the fraction of voxels of  $R_L$  belonging to the  $i$ th bin was calculated:  $f_{Ri} = N_{Ri}/N_R$ , where  $N_{Ri}$  is the number of voxels of  $R_L$  falling in the  $i$ th bin.
- ii. Because we wanted the refilled intensities in  $L$  to vary smoothly at the boundary,  $L$  was divided into two additional binarized masks: the border voxels of  $L$  ( $\delta L$ , refilled using the method described below in Step iii) and the inner voxels ( $L_{in}$ , obtained from  $L$  by excluding the border voxels). The number of voxels in  $L_{in}$  is denoted as  $N_{Lin}$ . To create a histogram for  $L_{in}$  that is well matched to the histogram of the voxels in  $R_L$ , the same number of bins ( $n_{bins}$ ) were used and each bin had the same proportion of entries.

The number of voxels in  $L_{in}$  assigned to the  $i$ th bin is denoted as  $N_{L-Bi}$ . Initially we set  $N_{L-Bi}$  to be the integer giving the smallest difference between  $f_{Ri}$  and  $f_{Li}$ , where  $f_{Li}$  is the fraction of voxels of  $L_{in}$  in the  $i$ th bin. However, this definition did not guarantee that the sum of  $N_{L-Bi}$  was equal to  $N_{Lin}$ . The difference,  $N_{diff}$ , is defined as

$$N_{diff} = N_{Lin} - \sum_{i=1}^{i=n_{bins}} N_{L-Bi}$$

All the possible ways in which the  $N_{diff}$  voxels could be rearranged into the  $n_{bins}$  were then explored, and for each of them a new  $f_{Li}$  was calculated. Finally, a set of  $N_{L-Bi}$  was chosen to minimize the function

$$\sum_{i=1}^{i=n_{bins}} (f_{Ri} - f_{Li})^2.$$

The  $n_{bins}$  chosen previously were an arbitrary choice, made without taking into account the lesion size. To account for size, a non-uniform sampling of the histogram was obtained by dividing each bin into an appropriate number of sub-bins. This was achieved by applying the procedure described above (Step ii) to each of the previously defined equally-sized bins, substituting  $R_L$  with  $R_{Li}$  and  $N_{Lin}$  with  $N_{L-Bi}$ . Thus, each bin was potentially divided into sub-bins, each of which can have a different number of voxels falling in it.

We denote the range of intensities covered by the  $j$ th sub-bin of the  $i$ th bin as  $\Delta I_{ij}$ , which contains  $N_{Bij}$  voxels. The intensities chosen for refilling voxels are drawn from uniform distributions covering each  $\Delta I_{ij}$ .

- iii. Finally,  $\delta L$  voxels were refilled with the mean value of 8 in plane nearest neighboring voxels that belong to either  $R_L$  or  $L_{in}$ . In addition, the mean of all the voxel intensities used to refill  $L$  was constrained to be equal to  $M_R$  by simply adding an offset,  $(M_R - M_L)$ , to the intensity of each voxel, where  $M_L$  is the mean voxel intensity in  $L$ .

At this stage, SIENAX was used to obtain NGMV from (i) the 5 “original” T1-W images, (ii) the 30 “adjusted” T1-W images where the lesions were masked out, and (iii) the 30 “adjusted” T1-W images where the lesions were refilled with the procedure described above.

### PV Models’ Impact on Estimation of Volumes in the Presence of WM Lesions

All the analyses described above were repeated using two different PV estimation methods, as provided in two different versions of FAST (FMRIB’s Automated Segmentation Tool) [Zhang et al., 2001]: FAST version 3 (released in FSL-4.0); and FAST version 4 (released in FSL-4.1). This was done to assess two things: how the size of the differences in brain volume measurements, in the presence of WM lesions, depends on the different PV classification approaches; and in addition, to test whether or not the refilling method affects the PV estimations to different extents. The main difference between the partial volume modeling used in FAST-3 and FAST-4 is that the Markov Random Field (MRF) is applied to the partial volume fractions (i.e., WM, GM and CSF) in FAST-3, but it is applied to the mixel-type in FAST-4. The mixel-type represents the classification of the mixture present in each voxel (e.g., pure WM, or a mixture of GM and WM, etc.) and applying the MRF to the mixel-type makes the assumption that the same mixture of tissues will be spatially adjacent, rather than assuming that the partial volume fractions will be similar between spatially adjacent voxels. A consequence of this is that the borders appear sharper in the FAST-4 version, as a pure tissue type (partial volume fraction of 1.0) is more likely to occur in one voxel away from a boundary voxel. This is mainly due to the fact that in the FAST-3 version, the MRF on the partial volume fraction tends to blur out the boundaries, biasing voxels that actually contain pure tissue to have lower partial volume fractions in order to be similar to their neighboring (boundary) voxels. However, the effect of these differences in partial volume modeling in the presence of WM lesions is difficult to assess theoretically, which is why this empirical study was performed.

### Statistical Analysis

Statistical analysis was performed using the R software (www.r-project.org). A within-within analysis of variance (ANOVA) of the values of NBV, NWMV and NGMV and

PBVC was performed using both lesion load and the type of lesion intensity distribution as factors. These analyses were followed by a pair-wise post-hoc comparison using Tukey's honestly significant difference procedure.

A linear regression was performed to evaluate the dependence of NGMV on lesion load in "masked" and "refilled" images. Data were considered significant at a  $P$ -value  $< 0.05$ .

## RESULTS

### Preliminary Test for the Use of "Artificial" Images in SIENAx and SIENA

As preliminary step, we assessed whether the use of "artificial" images, which are identical to "original" images for the vast majority of voxels, did not cause unexpected or unwanted behavior of the software in both SIENAx and SIENA measurements.

We first tested if the skull-finding (used as scaling factor) could be altered when artificial lesions were inserted. This was done by calculating the coefficient of variation (CV) for all the scaling factors referring to the "artificial" images related to the image of each healthy subject. We found that mean of the CV was 0.48, indicating a very small dispersion of the data within the same subject. In addition, the mean scaling factor for all the "artificial" images was 1.34 (range 1.24–1.44), with a maximum variation of  $\pm 8\%$ .

We then tested the performance of the SIENA method by analyzing five pairs of identical images. The PBVC values obtained from this analysis was equal to 0%, ruling out possible errors resulting from this approach.

### Influence of WM Lesions on Segmentation-Based Brain Volume Measurements

The extent to which the WM lesions may bias brain volume measurements in a segmentation method was assessed by comparing the SIENAx results of each "original" T1-W image and the corresponding "artificial" images with different WM lesion load and lesion intensities, thus obtaining the pd-NBV, pd-NWMV, and pd-NGMV (see Fig. 2).

When FAST-3 was used for the analysis, the results showed:

- NBV measures were  $1664 \pm 19 \text{ cm}^3$  in the "original" T1-W images and were generally not influenced by the increase in lesion load. When the lesion load was  $6 \text{ cm}^3$ , the values of differences in NBV for the different intensity filling models were: GM/WM =  $6.15 \pm 23.3 \text{ cm}^3$ , GM =  $-0.33 \pm 6.6 \text{ cm}^3$ , CSF/GM =  $0.33 \pm 3.3 \text{ cm}^3$ , CSF =  $-7.15 \pm 5.0 \text{ cm}^3$ . The values of differences in NBV decreased significantly ( $P < 0.001$ ) with high lesion load only when the lesion intensity was similar to that of CSF (differences in NBV:  $-56.1 \pm 13.6 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ).

- NWMV measures were  $858 \pm 38 \text{ cm}^3$  in the "original" T1-W images. When the lesion load was  $6 \text{ cm}^3$ , the values of differences in NWMV for the different intensity filling models were: GM/WM =  $5.83 \pm 12.9 \text{ cm}^3$ , GM =  $-4.8 \pm 4.3 \text{ cm}^3$ , CSF/GM =  $-5.8 \pm 1.7 \text{ cm}^3$ , CSF =  $-11.6 \pm 2.6 \text{ cm}^3$ . The values of differences in NWMV appeared to increase significantly ( $P < 0.001$ ) with high lesion load when the lesion intensity was similar to that of the GM/WM interface (differences in NWMV:  $93.6 \pm 18.9 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ). By contrast, they decreased with high lesion load ( $P < 0.001$ ) when the lesion intensity was similar to that of the CSF (NWMV:  $-58.3 \pm 11.8 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ).
- NGMV measures were  $805 \pm 37 \text{ cm}^3$  in the "original" T1-W images. When the lesion load was  $6 \text{ cm}^3$ , the values of differences in NGMV for the different intensity filling models were: GM/WM =  $0.5 \pm 10.4 \text{ cm}^3$ , GM =  $4.4 \pm 4.8 \text{ cm}^3$ , CSF/GM =  $6.1 \pm 4.0 \text{ cm}^3$ , CSF =  $3.9 \pm 4.0 \text{ cm}^3$ . The values of differences in NGMV progressively decreased with increasing lesion load ( $P < 0.001$ ) only when the lesion intensity was similar to that of the GM/WM interface (differences in NGMV:  $-82.7 \pm 4.8 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ).

When FAST-4 was used for the analysis, the results showed:

- NBV measures were  $1,555 \pm 16 \text{ cm}^3$  in the "original" T1-W images and were generally not influenced by the increase in lesion load. When the lesion load was  $6 \text{ cm}^3$ , the values of differences in NBV for the different intensity filling models were: GM/WM =  $9.0 \pm 18.7 \text{ cm}^3$ , GM =  $2.6 \pm 6.2 \text{ cm}^3$ , CSF/GM =  $-1.5 \pm 3.11 \text{ cm}^3$ , CSF =  $-8.8 \pm 7.8 \text{ cm}^3$ . The values of differences in NBV decreased significantly ( $P < 0.001$ ) with high lesion load only when the lesion intensity was similar to that of CSF (differences in NBV:  $-51.3 \pm 14.5 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ).
- NWMV measures were  $802 \pm 17 \text{ cm}^3$  in the "original" T1-W images. When the lesion load was  $6 \text{ cm}^3$ , the values of differences in NWMV for the different intensity filling models were: GM/WM =  $3.3 \pm 9.6 \text{ cm}^3$ , GM =  $-3.92 \pm 3.2 \text{ cm}^3$ , CSF/GM =  $0.0 \pm 1.6 \text{ cm}^3$ , CSF =  $-6.3 \pm 4.0 \text{ cm}^3$ . The values of differences in NWMV decreased with high lesion load ( $P < 0.001$ ) when the lesion intensity was similar to that of GM (differences in NWMV:  $-42.4 \pm 7.7 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ) or CSF (differences in NWMV:  $-37.5 \pm 9.7 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ).
- NGMV measures were  $753 \pm 19 \text{ cm}^3$  in the "original" T1-W images. When the lesion load was  $6 \text{ cm}^3$ , the values of differences in NGMV for the different intensity filling models were: GM/WM =  $5.7 \pm 9.0 \text{ cm}^3$ , GM =  $6.6 \pm 3.0 \text{ cm}^3$ , CSF/GM =  $-1.7 \pm 1.6 \text{ cm}^3$ , CSF =  $-2.4 \pm 3.8 \text{ cm}^3$ . The values of differences in NGMV progressively increased with increasing lesion load ( $P < 0.001$ ) only when the lesion intensity was similar to that of the GM (differences in NGMV:  $50.8 \pm 6.8 \text{ cm}^3$  for lesion load of  $60 \text{ cm}^3$ ).

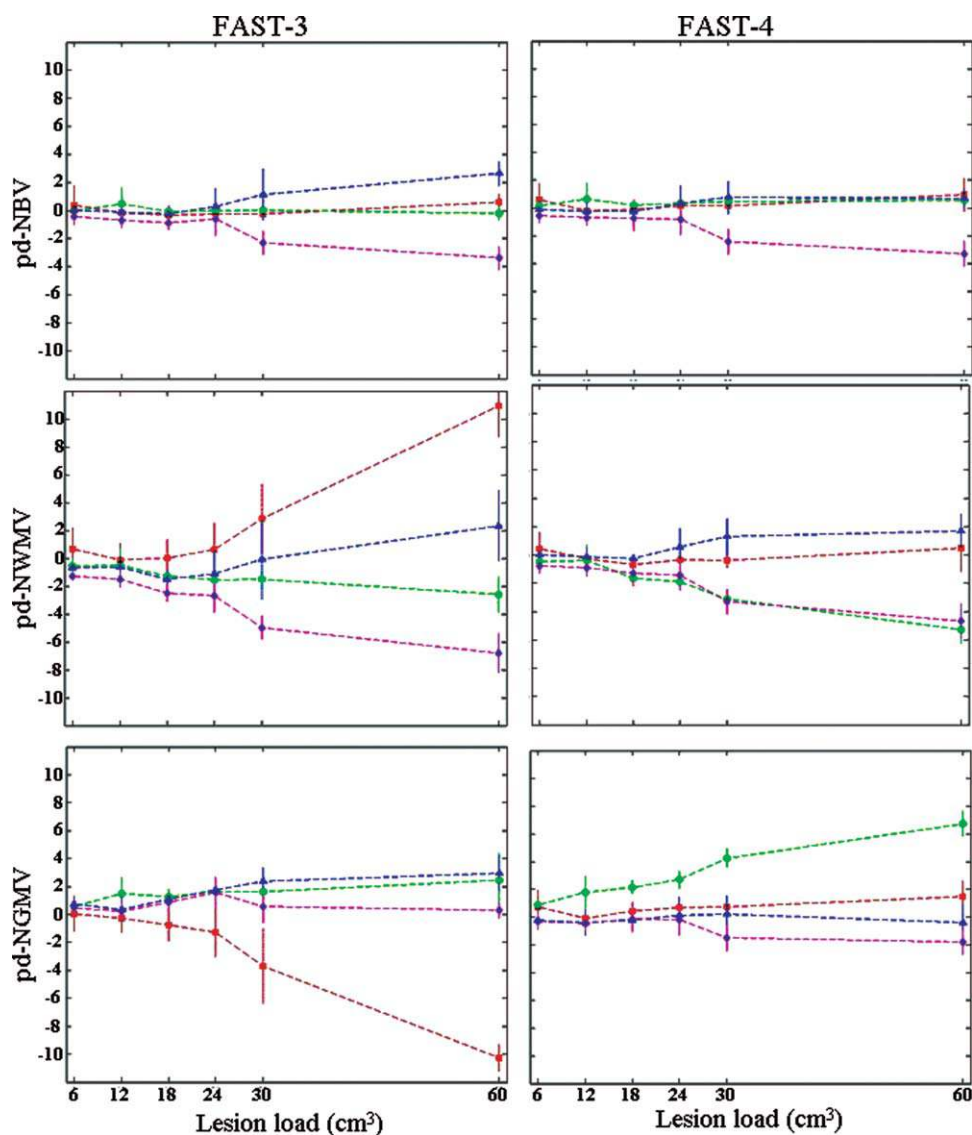


Figure 2.

The graphs illustrate the percentage differences (y-axis, defined as  $100 \times (V_2 - V_1)/V_1$ , where  $V_1$  and  $V_2$  represent the first and the second volume measurement, respectively) in the segmentation-based measurements (as assessed by SIENAX) of NBV (pd-NBV, top panels), NWMV (pd-NWMV, central panels), and NGMV (pd-NGMV, bottom panels) when lesions were inserted into the “original” T1-weighted images with an increasing lesion load (x-axis) and different intensities (■ for GM/WM interface,

● for GM, ▲ for CSF/GM and ◆ for CSF). The analysis was performed by using two different partial volume approaches as provided by FAST-3 (left column) and FAST-4 (right column) (see Methods for details). Each dot and vertical line in the graphs represents the mean and standard deviations of the five percentage differences obtained by comparing each “original” T1-weighted image with the related “artificial” T1-weighted image of a given lesion load and intensity.

### Influence of WM Lesions on Registration-Based Brain Volume Measurements

An analysis similar to that performed with the segmentation-based algorithm (SIENAX) was performed to test the extent to which the WM lesions could bias brain vol-

ume measurements in a registration-based method such as SIENA. In this case, PBVC was generally not influenced by the increase in lesion load and no differences were found by using FAST-3 or FAST-4 when performing the analysis (see Fig. 3). In both FAST-3 and FAST-4 analyses, the PBVC was always  $<0.1$  for a lesion load of  $6 \text{ cm}^3$  with

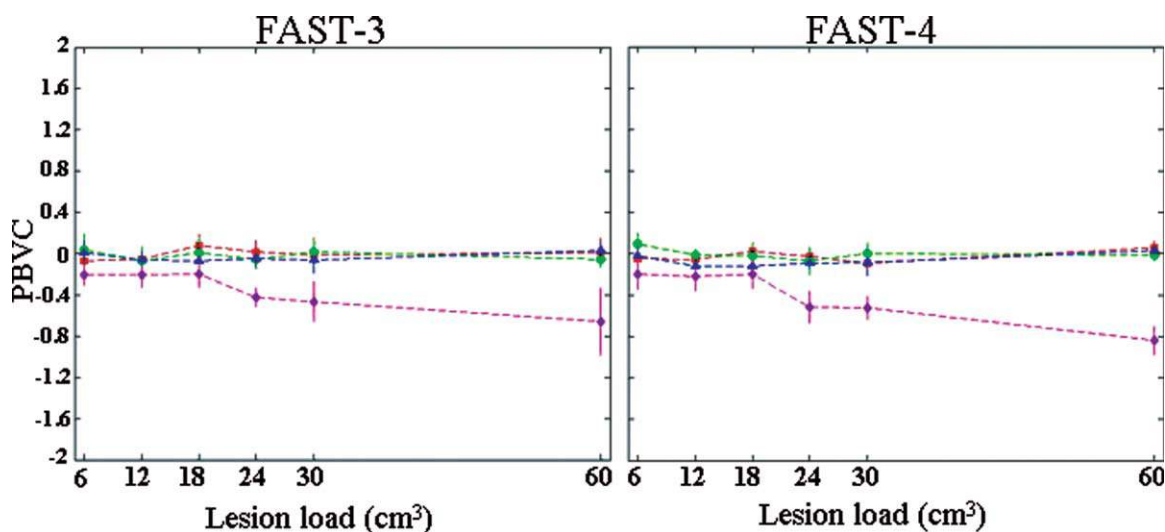


Figure 3.

The graphs illustrate the percentage brain volume changes (PBVC,  $y$ -axis) in the registration-based measurements (as assessed by SIENA) when lesions were inserted into the “original” T1-weighted images with an increasing lesion load ( $x$ -axis) and different intensities (■ for GM/WM interface, ● for GM, ▲ for CSF/WM and ◆ for CSF). The analysis was performed by using two different partial volume approaches as provided by

FAST-3 (left column) and FAST-4 (right column) (see Methods for details). Each dot and vertical line in the graphs represents the mean and standard deviations of the five percentage differences obtained by comparing each “original” T1-weighted image with the related “artificial” T1-weighted image of a given lesion load and intensity.

intensities similar to that of the GM/WM interface, GM and the CSF/GM interface, and it did not appear to increase with increasing lesion loads (see Fig. 3). The PBVC showed a significant difference ( $P < 0.001$ ) with high lesion load only when the lesion intensity was similar to that of CSF (for lesion load of  $60 \text{ cm}^3$ , PBVC with FAST-3:  $-0.66 \pm 0.33$ ; PBVC with FAST-4:  $-0.84 \pm 0.14$ ).

refilled with intensity similar to that of the surrounding normal appearing WM (FAST-3: slope  $0.051 \pm 0.093$ ; FAST-4:  $0.011 \pm 0.037$ ). By contrast, in both cases an inverse dependence between lesion load and NGMV was found when the lesions were masked out (FAST-3:  $-0.35 \pm 0.15$ ; FAST-4:  $-0.20 \pm 0.14$ ).

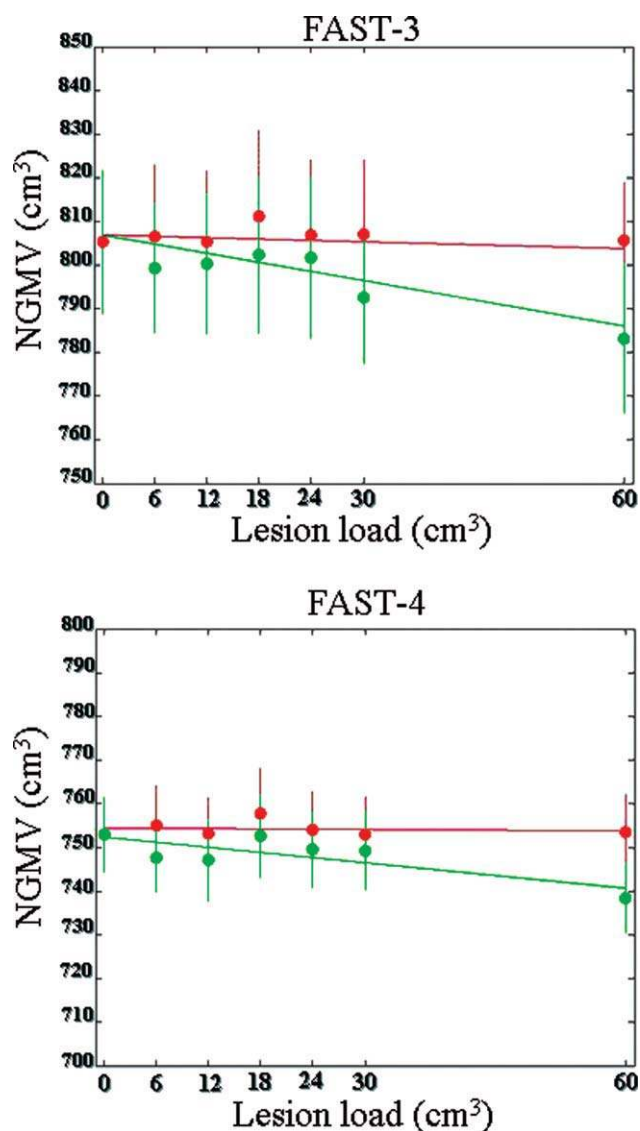
## DISCUSSION

### “Masked-Out” and “Refilled” WM Lesions in Segmentation-Based Brain Volume Measurements

In the SIENAX analysis, possible strategies for minimizing the errors due to WM lesion misclassification were tested (see Fig. 4). This was done by comparing two mean slopes derived by the NGMV measurements in (i) the “original” 5 T1-W images and each of the “adjusted” images where the six lesion masks with increasing lesion loads were masked out and (ii) the “original” 5 T1-W images and each of the “adjusted” images where the six lesion masks were filled with intensities similar to that of the surrounding normal appearing WM (following the method described previously).

Interestingly, with both FAST-3 and FAST-4 approaches (see Fig. 4), the analysis showed that NGMV measures were not dependent on WM lesions when these were

Recent work has shown that the presence of focal WM abnormalities such as those found in MS can affect MR-based quantitative measurements of the brain, especially if these rely on tissue type segmentation [Nakamura and Fisher, 2009] or registration algorithms [Sdika and Pelletier, 2009]. In the present study, using two widely utilized methods for brain volume measurements in MS research, a segmentation-based method for the measurement of atrophy state (SIENAX) and a registration-based approach for the measurement of atrophy rate (SIENA), we assessed how and to what extent WM lesions of various size and intensity may affect brain volume measurements when they are artificially inserted into the “original” T1-W brains. The main results of the study were: (i) WM lesions do affect segmentation-based measurements of brain volume, especially when tissue-class segmentation is performed; (ii) in these conditions the measurement error may vary considerably with lesion size and intensity as well as with the type of PV



**Figure 4.**

The graphs illustrate the regression lines of NGMV measurements (y-axis) in relation to lesion load (x-axis) when lesions were either “masked out” (●) or “refilled” with the surrounding normal-appearing WM (●). Values of NGMV at 0 lesion load are given by those of the “original” T1-weighted images of the healthy controls. The analysis was performed by using two different partial volume approaches as provided by FAST-3 (left column) and FAST-4 (right column) (see Methods for details).

estimation used; (iii) the linear registration-based approach (SIENA) is relatively insensitive to the presence of WM lesions, with the exception of cases with large, CSF-like intensity lesion loads; (iv) the error found in the segmentation-based measurements might be substantially solved by refilling the WM lesions with an intensity similar to that of the surrounding normal appearing WM.

By comparing the SIENAX results of “original” T1-W images from healthy controls and the corresponding “artificial” images where WM lesions with different sizes and intensities were inserted, we were able to accurately assess the influence of WM lesions on total and tissue-class (i.e., WM and GM) brain volume measures. These results showed that a certain degree of influence of WM lesions on segmentation-based brain volume measurements was generally present. This was, in the case of total brain measures (i.e., NBV), relevant only in the presence of very high lesion load with CSF intensity, which is very rare on clinical grounds. By contrast, substantial changes were seen in both WM and GM measures in most of the tested “artificial” images, showing that tissue misclassification is often found in the presence of WM lesions of different intensities and even with moderate lesion load. Differently from previous work [Chard et al., 2010; Nakamura and Fisher, 2009], we investigated the effect on partial volume estimation and how this can affect measures of atrophy calculated by the segmentation-based methods (SIENAX) and registration-based methods (SIENA). Interestingly, tissue misclassification was greater in the FAST-3 measures, with a pronounced underestimation of the GM volume (and a consequent overestimation of the WM volume) with increasing lesion load and lesion intensity similar to that of the GM/WM interface (see Fig. 2). This is particularly important in a real-world setting since intensities between that of WM and GM are very likely to be present in hypointense WM lesions on the conventional T1-W images that are generally used in clinical studies for MR-based measurements of brain volumes.

The misclassification of MS lesions as GM in T1-W images is probably not the only source of error in brain volume measurements. In support of this, the error in the assessment of GM volumes was also present when the NGMV were calculated after the exclusions of lesional voxels misclassified as GM (data not shown). Furthermore, for example, when “artificial” T1-W images with 60 cm<sup>3</sup> of lesions were filled with GM/WM intensity, the absolute volume of the FAST3 output underestimated the GM of about 69 cm<sup>3</sup> and when the same images were filled with the GM intensity, the absolute volume of the FAST4 output overestimated the GM of about 43 cm<sup>3</sup>. This demonstrates that the observed brain volume changes due to the lesions are not simply the volume of those (misclassified) lesion voxels themselves, but that the presence of lesions may affect the tissue classification of the segmentation algorithm. Given the irregular shape of brain tissue interfaces, the MRI voxels in specific brain regions (i.e., the GM/WM and CSF/GM interfaces) contain a mixture of tissue types. In the presence of WM lesions this may lead to classification errors during segmentation due to the failure of PV models to provide accurate tissue classification [Nakamura and Fisher, 2009]. Interestingly, in our study, this issue has shown to produce very different results when two different PV approaches were implemented in the same segmentation algorithm (i.e., FAST-3 and



FAST-4, see Fig. 2). Indeed, the influence of WM lesions and the consequent error in brain volume measurements appeared much lower in the FAST-4 analyses, probably due to the use of a mixel-type MRF in its PV model (see Methods for details). The mixel-type represents the classification of the mixture present in each voxel, which in FAST-4 included a six-tissue-class PV modeling (i.e., pure WM, WM/GM, pure GM, GM/CSF, pure CSF and WM/CSF) rather than the three-tissue-class approach of FAST-3 and other widely used segmentation algorithms [Ashburner and Friston, 2005; Nakamura and Fisher, 2009]. Thus, although it is not possible to generalize from data in this study, a mixel-type PV modeling such as that used in FAST-4 seems to be the best approach in presence of WM lesions such as those found in MS brains.

In the present study we also tested the extent to which the WM lesions could bias a registration-based method for global measurement of brain volume changes, such as SIENA. Results showed that PBVC measures were insensitive to increases in lesion load and to different intensity filling models, independently of PV modeling used, with the exception of the CSF intensity-filling model. This is particularly interesting as it is very similar to what was found in the segmentation-based analysis of global brain volume measurements. Taken together, these findings suggest that when the WM and GM segmentation is not attempted, the error due to the presence of WM lesions is limited to CSF misclassification that, on clinical grounds, might be present only in T1-W black holes exhibiting extreme tissue loss [Barkhof and van Walderveen, 1999]. Finally, it is worth noting that the present data suggest that the linear registration methods do not seem to suffer from the same problems encountered by the non-rigid registration approaches in the presence of WM lesions of MS brain images [Sdika and Pelletier, 2009].

Once the error in brain volume measurements due to WM lesions was quantified, possible strategies for minimizing this error were tested. Our results showed that, in the SIENAX analysis, the errors due to GM misclassification in the presence of WM lesions could not be corrected by simply masking the WM lesions out, an approach widely used in clinical studies [Chard et al., 2002]. By contrast, the refilling of the lesions with intensities that match their surrounding normal-appearing WM appeared to solve most of the issues related to GM misclassification. This refilling used a methodological approach similar to that previously reported by Sdika and Pelletier [2009]. The presence of similar results with different PV modeling methods (similar slopes were found with both FAST-3 and FAST-4 analyses) adds further support to this statement. It must be stressed here that the use of “artificial” T1-W images obtained from T1-W images of healthy controls may have created an easier scenario than the one that needs to be faced in routine MR images of MS brains. Certainly, for example, the normal-appearing WM of MS patients might not be normal, especially in perilesional regions [Vrenken et al., 2006]. In this case, however, the

enlargement of the surrounding WM area beyond the perilesional “dirty” WM, which could be easily done with our non-uniformly sampled method, may help to solve or minimize the problem.

In conclusion, the results of this study show that the presence of WM lesions does not bias longitudinal, linear-registration-based measurements of global brain atrophy, where tissue-class classification is not required. By contrast, WM lesions may significantly affect GM measurements, especially when their intensity is between that of WM and GM, a condition that is very likely to occur in the hypointense WM lesions found on the conventional T1-W images that are used in clinical settings. However, the extent to which the presence of WM lesions may affect tissue-class measures is clearly driven by the PV modeling used by the segmentation algorithm. The use of both a mixel-type PV model and the refilling of the lesions with the surrounding normal-appearing WM seem to solve the problems created by the presence of WM lesions and provide accurate tissue-class measurements.

## ACKNOWLEDGMENTS

The authors thank Dr. Antonio Giorgio (Dept. of Neurological Sciences, University of Siena) for thoughtful discussion. The authors are grateful to Arlene Cohen for revising the manuscript language.

## REFERENCES

- Ashburner J, Friston KJ (2005): Unified segmentation. *Neuroimage* 26:839–851.
- Barkhof F, van Walderveen M (1999): Characterization of tissue damage in multiple sclerosis by nuclear magnetic resonance. *Philos Trans R Soc Lond B Biol Sci* 354:1675–1686.
- Battaglini M, Giorgio A, Stromillo ML, Bartolozzi ML, Guidi L, Federico A, De Stefano N (2009): Voxel-wise assessment of progression of regional brain atrophy in relapsing-remitting multiple sclerosis. *J Neurol Sci* 282:55–60
- Bendfeldt K, Kuster P, Traud S, Egger H, Winkhofer S, Mueller-Lenke N, Naegelin Y, Gass A, Kappos L, Matthews PM, Nichols TE, Radue EW, Borgwardt SJ (2009): Association of regional gray matter volume loss and progression of white matter lesions in multiple sclerosis—A longitudinal voxel-based morphometry study. *Neuroimage* 45:60–67.
- Chard DT, Griffin CM, McLean MA, Kapeller P, Kapoor R, Thompson AJ, Miller DH (2002): Brain metabolite changes in cortical grey and normal-appearing white matter in clinically early relapsing-remitting multiple sclerosis. *Brain* 125 (Part 10):2342–2352.
- Chard DT, Griffin CM, Rashid W, Davies GR, Altmann DR, Kapoor R, Barker GJ, Thompson AJ, Miller DH (2004): Progressive grey matter atrophy in clinically early relapsing-remitting multiple sclerosis. *Mult Scler* 10:387–391.
- 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.

- 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.
- De Stefano N, Matthews PM, Filippi M, Agosta F, De Luca M, Bartolozzi ML, Guidi L, Ghezzi A, Montanari E, Cifelli A, Federico A, Smith SM (2003): Evidence of early cortical atrophy in MS: Relevance to white matter changes and disability. *Neurology* 60:1157–1162.
- 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.
- Niessen WJ, Vincken KL, Weickert J, Haar Romeny BM, Viergever MA (1999): Multiscale segmentation of three-dimensional MR brain images. *Int Comput Vis* 31:185–202.
- Santago P, Gage HD (1995): Statistical models of partial volume effect. *IEEE Trans Image Process* 4:1531–1540.
- 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.
- Smith SM, De Stefano N, Jenkinson M, Matthews PM (2001): Normalized accurate measurement of longitudinal brain change. *J Comput Assist Tomogr* 25:466–475.
- 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.
- Van Leemput K, Maes F, Vandermeulen D, Suetens P (2003): A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans Med Imaging* 22:105–119.
- Vrenken H, Geurts JJ, Knol DL, Polman CH, Castelijns JA, Pouwels PJ, Barkhof F (2006): Normal-appearing white matter changes vary with distance to lesions in multiple sclerosis. *AJNR Am J Neuroradiol* 27:2005–2011.
- 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.