JAMA Neurology | Original Investigation

Assessing Biological and Methodological Aspects of Brain Volume Loss in Multiple Sclerosis

Magí Andorra, MSc; Kunio Nakamura, PhD; Erika J. Lampert, BSc; Irene Pulido-Valdeolivas, MD, PhD; Irati Zubizarreta, MD; Sara Llufriu, MD, PhD; Eloy Martinez-Heras, PhD; Nuria Sola-Valls, MD; María Sepulveda, MD; Ana Tercero-Uribe, MD; Yolanda Blanco, MD, PhD; Albert Saiz, MD, PhD; Pablo Villoslada, MD, PhD; Elena H. Martinez-Lapiscina, MD, PhD

IMPORTANCE Before using brain volume loss (BVL) as a marker of therapeutic response in multiple sclerosis (MS), certain biological and methodological issues must be clarified.

OBJECTIVES To assess the dynamics of BVL as MS progresses and to evaluate the repeatability and exchangeability of BVL estimates with Jacobian Integration (JI) and Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL) (specifically, the Structural Image Evaluation, Using Normalisation, of Atrophy–Cross-Sectional [SIENA-X] tool or FMRIB's Integrated Registration and Segmentation Tool [FIRST]).

DESIGN, SETTING, AND PARTICIPANTS A cohort of patients who had either clinically isolated syndrome or MS was enrolled from February 2011 through October 2015. All underwent a series of annual magnetic resonance imaging (MRI) scans. Images from 2 cohorts of healthy volunteers were used to evaluate short-term repeatability of the MRI measurements (n = 34) and annual BVL (n = 20). Data analysis occurred from January to May 2017.

MAIN OUTCOMES AND MEASURES The goodness of fit of different models to the dynamics of BVL throughout the MS disease course was assessed. The short-term test-retest error was used as a measure of JI and FSL repeatability. The correlations (R^2) of the changes quantified in the brain using JI and FSL, together with the accuracy of the annual BVL cutoffs to discriminate patients with MS from healthy volunteers, were used to measure compatibility of imaging methods.

RESULTS A total of 140 patients with clinically isolated syndrome or MS were enrolled, including 95 women (67.9%); the group had a median (interquartile range) age of 40.7 (33.6-48.1) years. Patients underwent 4 MRI scans with a median (interquartile range) interscan period of 364 (351-379) days. The 34 healthy volunteers (of whom 18 [53%] were women; median [IQR] age, 33.5 [26.2-42.5] years) and 20 healthy volunteers (of whom 10 [50%] were women; median [IQR] age, 33.0 [28.7-39.2] years) underwent 2 MRI scans within a median (IQR) of 24.5 (0.0-74.5) days and 384.5 (366.3-407.8) days for the short-term and long-term MRI follow-up, respectively. The BVL rates were higher in the first 5 years after MS onset (R^2 = 0.65 for whole-brain volume change and R^2 = 0.52 for gray matter volume change) with a direct association with steroids (β = 0.280; P = .02) and an inverse association with age at MS onset, particularly in the first 5 years (β = 0.015; P = .047). The reproducibility of FSL (SIENA) and JI was similar for whole-brain volume loss, while JI gave more precise, less biased estimates for specific brain regions than FSL (SIENA-X and FIRST). The correlation between whole-brain volume loss using JI and FSL was high (R^2 = 0.92), but the same correlations were poor for specific brain regions. The area under curve of the whole-brain volume change to discriminate between patients with MS and healthy volunteers was similar, although the thresholds and accuracy index were distinct for JI and FSL.

CONCLUSIONS AND RELEVANCE The proposed BVL threshold of less than 0.4% per year as a marker of therapeutic efficiency should be reconsidered because of the different dynamics of BVL as MS progresses and because of the limited reproducibility and variability of estimates using different imaging methods.

JAMA Neurol. 2018;75(10):1246-1255. doi:10.1001/jamaneurol.2018.1596
Published online July 2, 2018.

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Elena H. Martinez-Lapiscina, MD, PhD, Center of Neuroimmunology, Department of Neurology, Hospital Clinic of University of Barcelona, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), Villarroel 170, Barcelona ES 08036, Spain (hernandez@clinic.ub.es).

Multiple sclerosis (MS) is an immune-mediated inflammatory and neurodegenerative disease of the central nervous system (CNS). To reduce the risk of long-term disability in MS, early optimization of treatment with disease-modifying therapies (DMTs) is essential.[1] Consequently, it has been proposed that a possible aim when treating MS is to achieve no evident disease activity (NEDA-3),[2] which is currently defined by 3 criteria: the absence of relapses, a worsening in disability, and new or enlarging T2/T2-fluid-attenuated inversion recovery (T2-FLAIR) or gadolinium-enhanced lesions. NEDA is a useful composite for capturing the inflammatory burden in MS,[3,4] although neuroaxonal injury and its prominent contribution to permanent disability in MS is likely to be underestimated in the NEDA-3 composite. Therefore, extensive research is ongoing to establish a validated marker of neuroaxonal injury in MS. Together with retinal thickness,[5] brain volume measurements[6] have raised much interest because of their reliable association with disability and the relative simplicity of image-processing algorithms. Recent randomized clinical trials have shown how new DMTs can reduce the rate of brain volume loss (BVL) in cohorts of patients.[7] Consequently, it has been suggested that NEDA-3 should be expanded to NEDA-4 by adding a BVL rate of less than 0.4% per year as a therapeutic target.[8]

Although an appealing proposal, certain issues must be clarified before BVL can be validated as a marker of therapeutic response. First, it is not clear if BVL is similar throughout the disease course[9,10] or if it progresses faster at any particular stage of MS.[11] Using a single cutoff would only be reasonable if the rate of BVL remained constant throughout the course of MS; otherwise, it might be necessary to establish different goals according to the stage of MS. Second, it is important to assess whether magnetic resonance imaging (MRI) methods are sufficiently reproducible to accurately track BVL at the individual-patient level, considering the expected magnitudes of BVL in MS.[12] Finally, we should appraise whether the changes in brain volume estimated with different methods are comparable.

As such, our primary aim was to evaluate the dynamic changes in the whole-brain and regional brain volume throughout the course of MS using Jacobian Integration (JI)[13] and Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL).[14-16] As secondary aims, we assessed the repeatability and the compatibility of the JI and FSL methods to estimate the changes in brain volume.

Key Points

Question Could brain volume loss of less than 0.4% per year serve as a marker of therapeutic response in multiple sclerosis?

Findings In this study, the rate of brain volume loss is faster in the first 5 years after multiple sclerosis onset, which affects the usability of a 0.4% threshold. Considering an expected loss of 0.5% to 1.35% per year, the poor repeatability of 2 separate imaging methods prevents their use at the individual level; also, volumetric estimates and the accuracy of thresholds to discriminate patients with multiple sclerosis from healthy volunteers differ in the 2 approaches.

Meaning Further clarification is needed before promoting brain volume loss as a marker of therapeutic response in multiple sclerosis.

Methods

Study Population

In this study, the first 147 consecutive patients with MS (with conditions defined according to the McDonald criteria[17,18]) who were enrolled into the prospective MS-VisualPath cohort[19] at the Hospital Clinic-University of Barcelona were evaluated for eligibility. We excluded 5 patients who had not completed any MRI follow-up visits, 1 patient who had MRI scans of insufficient quality, and 1 patient with a disease duration longer than 30 years. As a result, 140 patients with MS were included in this study. Each patient underwent annual MRI scans (median number per patient, 4; interquartile range [IQR], 3-4) with a median interscan period of 364 days (IQR, 351-379 days). In addition, we included 2 different cohorts of healthy volunteers: a cohort of 34 healthy volunteers who underwent 2 MRI scans within a median interscan period of 24.5 days, and a second cohort of 20 healthy volunteers who underwent 2 MRIs within a median interscan period of 384.5 days.

The institutional review board of the Hospital Clinic of Barcelona approved the study, and all participants provided their written informed consent. Moreover, the article follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. Further details can be found in eMethods in the Supplement.

Images Acquisition and Processing

All MRI studies were performed with a 3-T Magnetom Trio scanner (Siemens), using a 32-channel phased-array head coil.[19] In this study, 3-dimensional (3-D) structural T1-weighted isometric voxel magnetization-prepared rapid gradient echo (T1-MPRAGE) images with a voxel size of 0.9 × 0.9 × 0.9 mm^3 were used to estimate all the volumes, and 3-D T2-FLAIR with the same voxel size was used to make the lesion masks.

First, T2-FLAIR images were registered to T1-MPRAGE scans using a rigid transformation, with 6 *df* (FSL). Second, a trained neurologist manually created T1 lesion masks using ITK-SNAP, version 2.4 (Penn Image Computing and Science Laboratory, University of Pennsylvania, and the Scientific Computing and Imaging Institute, University of Utah).[20] Third, lesion in-painting was performed on T1-MPRAGE scans using lesion mask, a state-of-the-art method for volumetric analysis that avoids pixel misclassification.[21] Finally, we used the JI[13] and FSL methods[14-16] to quantify the change in volume in the whole brain, gray and white matter, thalamus, caudate nucleus, putamen, and hippocampus. Healthy volunteers were evaluated using the same protocol, scanner, and image processing pipeline, excluding the lesion in-painting.

Structural Image Evaluation, Using Normalisation, of Atrophy (SIENA) is a registration-based technology used to quantify changes in whole-brain volume between 2 time points. By contrast, SIENA-X and FMRIB's Integrated Registration and

Segmentation Tool (FIRST) are segmentation-based techniques that are useful for cross-sectional studies but less reliable for longitudinal studies. The SIENA-X and FIRST tools segment the voxels in the region of interest (ROI) (eg, gray matter) in a single 3-D T-1 scan, measuring the volume as the number of segmented voxels in the ROI multiplied by the voxel volume of the 3-D T-1 scan. The difference between 2 points can then be calculated by subtracting the volumes.

The JI method computes a nonrigid registration between the two 3-D T-1 scans, producing a Jacobian matrix for each voxel, the determinant of which reflects the expansion and contraction of that voxel. The integration of all determinants in the segmented region (eg, gray matter) gives an estimate of the change in volume.[22] We used a custom JI pipeline defined previously[13] that is essentially no different than commercial JI solutions (such as MSmetrix).[23]

Statistical Analyses

We used the median and IQRs to describe the quantitative variables, absolute frequencies, and the proportions of the qualitative variables. We used mixed-effects regression to model the rate of change in brain volume, accounting for intraparticipant correlation.[24] We compared third-order and second-order B-spline and linear models based on Akaike Information Criterion.[25] With disease duration (time from MS onset) as a main fixed-effect predictor, these models were used to fit the rates of brain volume changes between visits using the JI and FSL methods. We used the pseudoR^2 defined by Nakagawa and Schielzoth[26] to estimate the goodness of fit of the models. To evaluate the potential influence of sex, age, DMT use, and use of steroids, we included these variables as fixed effect predictors in mixed-effects models. We included age at MS onset instead of age at study inclusion to avoid collinearity with disease duration. For variables with a significant association, we evaluated whether the effect on the rate of brain volume change might differ during disease progression (effect modification) using an interaction term between MS duration and covariate.

Short-term test-retest measurement errors were estimated using the formula of 100 * (V2 − V1) / mean(V1,V2), where V1 and V2 were absolute values measured at the 2 time points for the segmentation technologies (SIENA-X and FIRST), while the absolute percentage change in volume between 2 points was used for registration-based technologies (SIENA and JI). We presented the 50th and 75th percentiles with their 95th CIs estimated using a bias-corrected accelerated bootstrap to summarize the test-retest measurement errors. We also calculated the same difference after coregistering both MPRAGE (which is not a standard step for SIENA-X).

To assess if the JI and FSL estimated brain volumes are comparable, R^2 was evaluated as a measure of the goodness of fit of the simple linear regression models for pairwise brain regions quantified with JI and FSL. Additionally, we evaluated the performance of the annual rate of whole and regional brain volume changes to classify 20 healthy volunteers and 100 age and sex-matched patients with MS. We used a Hosmer-Lemeshow goodness-of-fit test for calibration and receiver operating characteristics curve analyses for discrimination. We also obtained the area under the curve, sensitivity, and specificity of the best cutoffs according to the Youden *J* statistic.[27]

We performed 2 sensitivity analyses: complete case analyses to assess the influence of missing data and analyses excluding participants with more than a 10-fold interscan change (which would evaluate the influence of extreme values on results). Two-tailed *P* values <.05 were considered statistically significant, and all statistical analyses were performed using R language (R version 3.3.3; R Foundation for Statistical Computing). Data analysis was completed from January 2017 to May 2017.

Results

Study Population Features

The study population included 140 patients with MS (95 women [67.8%]; median [interquartile range] age, 40.7 [33.6-48.1] years) with a median (IQR) of 7.0 (3.3-12.6) years of disease duration at inclusion (**Table 1**). Most patients had relapsing-remitting MS, were mildly disabled (median [IQR] score on the Expanded Disability Status Scale, 1.5 [1-2] points), and were receiving DMTs at inclusion.

Patients were allowed to change therapy during the follow-up period, and 7 patients (5.0%) started a DMT (low to intermediate potency); 11 patients (7.9%) changed to another DMT within the same treatment group (low to intermediate potency or high potency); 8 patients (5.7%) changed from natalizumab to fingolimod because of a positive John Cunningham virus response; and 1 patient (0.7%) changed from low-to-intermediate-potency to high-potency treatment because of a lack of efficacy of the first regimen. Additionally, 22 of 138 patients (15.9%) received steroids during the first interscan period, 13 of 118 patients (11.0%) received them during the second, and 9 of 107 (8.4%) received them during the third interscan period.

Rate of Brain Volume Loss by MS Stages

Third-order B-spline mixed-effects models best fit the distribution of the annual changes in whole-brain and gray matter volume quantified with JI according to the Akaike Information Criterion[25] (eTable 1 in the Supplement). The annual whole-brain and gray matter volume changed faster during the first 5 years of disease (**Figure 1**) than in later years, with Nakagawa conditional R^2 values of 0.65 and 0.52, respectively. The JI method gave a similar dynamic pattern of annual changes in the thalamus, putamen, and hippocampus, but not the caudate nucleus (eFigure 1 in the Supplement). No significant model was found for the annual rate of changes in white matter volume. Similar annual rates of change in whole-brain volume were obtained with SIENA (eFigure 2 in the Supplement), yet no suitable model was found for the annual rate of gray matter volume change using SIENA-X (eFigure 2 in the Supplement) or for the deep gray matter volume changes using FIRST (eFigure 3 in the Supplement).

Table 1. Demographic and Clinical Characteristics of the Patients and Controls at Baseline

Characteristic	No. (%)		
	Patients With MS (n = 140)	Healthy Volunteer Reproducibility Cohort (n = 34)	Healthy Volunteer Annualized Brain Volume Loss Cohort (n = 20)
Female	95 (67.8)	18 (53)	10 (50)
White race/ethnicity	140 (100)	34 (100)	20 (100)
Age at baseline, median (IQR), y	40.7 (33.6-48.1)	33.5 (26.2-42.5)	33.0 (28.7-39.2)
Disease duration (baseline), median (IQR), y	7.0 (3.3-12.6)	NA	NA
Annualized relapse rate (2 y preinclusion), median (IQR)	0.5 (0-0.5)	NA	NA
Annualized relapse rate (during follow-up), median (IQR)	0 (0-0.33)	NA	NA
Disease type			
Clinically isolated syndrome	6 (4.3)	NA	NA
Relapsing-remitting multiple sclerosis	121 (86.4)	NA	NA
Secondary progressive multiple sclerosis	5 (3.6)	NA	NA
Primary progressive multiple sclerosis	8 (5.7)	NA	NA
Expanded Disability Status Scale score, median (IQR)	1.5 (1-2)	NA	NA
Disease modifying therapies			
None	35 (25.0)	NA	NA
Interferon beta 1b, subcutaneous	25 (17.9)	NA	NA
Interferon beta 1a, subcutaneous	29 (20.7)	NA	NA
Interferon beta 1a, intramuscular	13 (9.3)	NA	NA
Glatiramer acetate	24 (17.1)	NA	NA
Natalizumab	10 (7)	NA	NA
Other medications[a]	4 (2.9)	NA	NA
Normal brain volumes at baseline, median (IQR)			
Brain (parenchymal), cm^3	1532 (1460-1595)	1613 (1572-1692)	1609 (1563-1637)
Gray matter, cm^3	801 (762-836)	833 (798-885)	846 (793-862)
White matter, cm^3	733 (691-770)	782 (755-808)	762 (744-794)
Thalamus, mm^3	14 490 (13 476-15 605)	16 067 (15 158-16 672)	16 462 (15 351-17 450)
Caudate, mm^3	6521 (5965-7103)	7280 (6760-7799)	7111 (6787-8262)
Putamen, mm^3	9044 (8276-9560)	10 061 (9463-10 879)	10 094 (8942-11 136)
Hippocampus, mm^3	7197 (6681-7745)	7837 (7374-8379)	7776 (7095-8684)

Abbreviation: NA, not applicable.

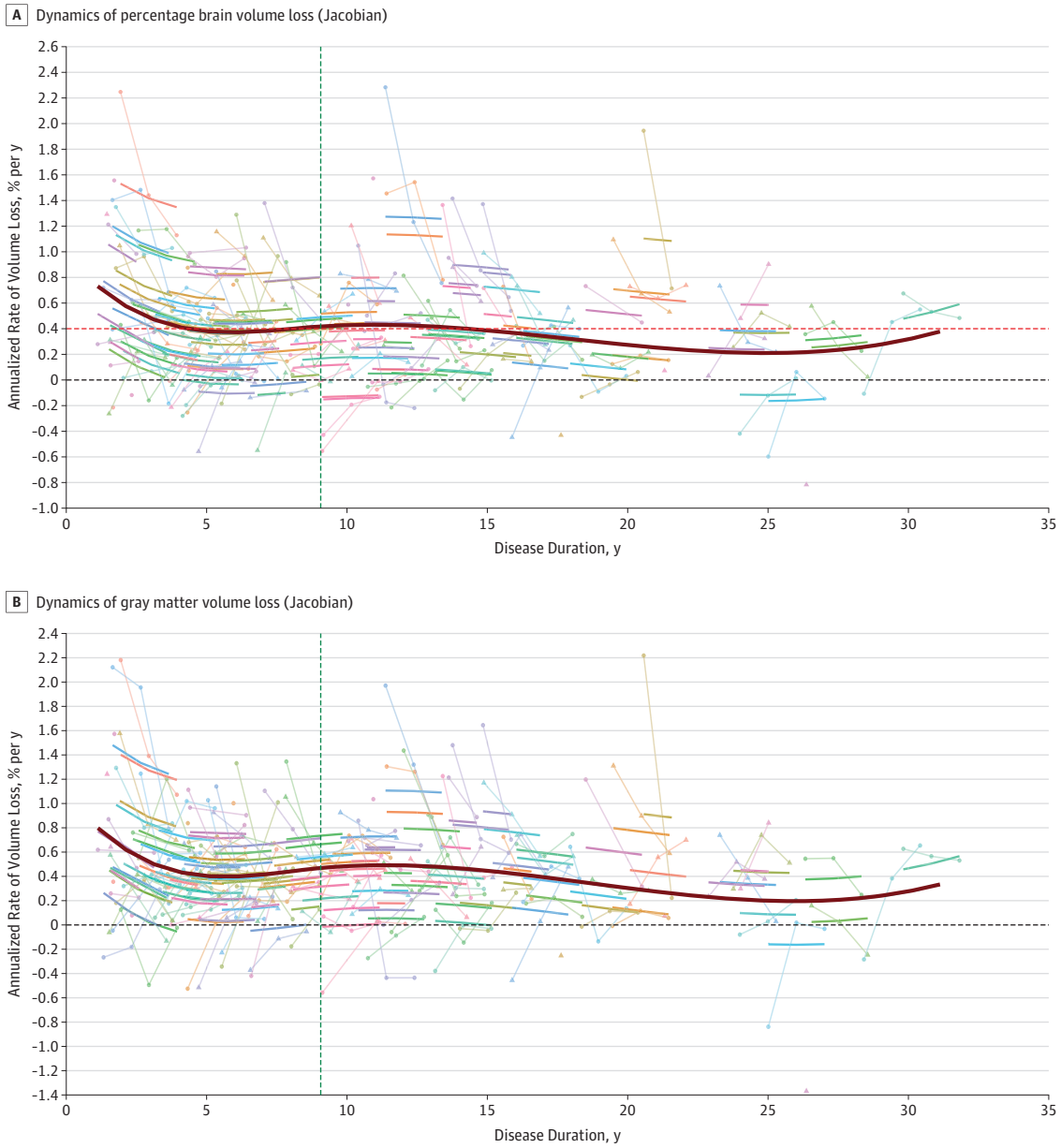
[a] Fingolimod (n = 1), diazoxide (n = 1), dimethyl-fumarate (n = 1), and teriflunomide (n = 1).

Demographic-Associated and MS-Associated Variables and Brain Volume Change Dynamics

We did not find a significant association with sex (β = −0.091; *P* = .21), age at inclusion (β = 0.001; *P* = .92), DMT (low-to-intermediate-potency drugs: β = 0.105; *P* = .12; high-potency drugs: β = 0.061; *P* = .63), or steroid use (β = 0.073; *P* = .19) in the model assessing whole-brain volume changes. Similarly, there was no significant coefficient for disease duration in the model of white-matter volume change conditioning on sex, age at MS onset, DMT use, and steroid use. In the model of changes in gray matter volume, no significant effect of sex (β = 0.051; *P* = .47) or age at MS onset (β = −0.008; *P* = .06) was found.

However, the near significance of the age at MS onset prompted exploration of the interaction of age at onset with other factors. The interaction between age at MS onset and MS duration was marginally significant (modeled as a binary variable with a cutoff in 5 years or >5 years: β = 0.015; *P* = .047). Controlling for sex, DMT, and steroid use, participants with MS duration of 5 years or less had 0.018% per year lower change in gray matter volume for each year of later MS onset (age: β = −0.003; interaction: β = −0.015), whereas the effect for those with an MS

Figure 1. Dynamics of the Annualized Rate of Brain Volume Loss in Multiple Sclerosis During Disease Progression



Whole-brain volume loss (A) and gray matter volume loss (B), both assessed with the Jacobian Integration method. The third-order B-spline mixed-effect model follows the equation y ≈ BS (disease duration) + (1/participant), where BS (x) is the B-spline and (1/participant) is a random intercept for each participant. Colored points joined by a line represent the individual trajectories of brain volume changes, the thicker curves represent the individual fit of the model, and the dark red line represents the population model. The dotted black horizontal line indicates no change in brain volume, and the green dotted vertical line represents the 50th percentile from the raw disease duration data, where the knot of the B-spline was placed. The red dotted horizontal line is the reference cutoff of 0.4% per year for the whole-brain volume change; there was no analogous cutoff for gray matter volume change.

duration longer than 5 years was a decrease of 0.003% per year (age: β = −0.003).

We did not find a significant association with DMT (low-to-intermediate-potency drugs: β = 0.015; *P* = .83; high-potency drugs: β = −0.088; *P* = .52) or steroid use (β = 0.110; *P* = .09) in an assessment with no interaction with MS duration (β = 0.164; *P* value = .23). Results from patients with MS who had a disease duration of less than 5 years, stratified by DMT and steroid use, are in eTable 2 in the Supplement.

For deep gray matter structures, we found a significant association with steroids in models assessing changes in thalamic volume (β = 0.280; *P* = .02), without interaction with disease duration (β = 0.145; *P* = .56). Moreover, we found a significant association with age at MS onset (β = 0.019; *P* = .01) in the dynamics of hippocampal volume change, yet no interaction with disease duration (β = −0.002; *P* = .86). We did not find any other significant associations for the other brain regions tested.

Table 2. Short-term Test-Retest Measurement Errors of Jacobian Integration and FSL in Changes in Whole-Brain and Regional Brain Volume in Healthy Volunteers[a]

Region	Percentile Value (95% CI)			
	FSL		Jacobian Integration	
	50th Percentile	75th Percentile	50th Percentile	75th Percentile
Main brain structures[b]				
Whole brain	0.156 (0.088-0.213)	0.266 (0.196-0.449)	0.230 (0.089-0.298)	0.361 (0.270-0.518)
Gray matter	0.812 (0.605-1.280)	1.776 (1.018-2.129)	0.180 (0.072-0.281)	0.377 (0.226-0.458)
White matter	1.253 (0.684-1.604)	1.955 (1.459-2.724)	0.192 (0.126-0.331)	0.557 (0.237-0.749)
Deep gray matter structures[c]				
Thalamus	0.918 (0.404-1.590)	2.151 (1.431-3.231)	0.471 (0.301-0.556)	0.711 (0.503-1.039)
Putamen	1.218 (0.746-1.534)	1.777 (1.373-2.567)	0.502 (0.395-0.766)	0.935 (0.531-1.231)
Caudate	1.109 (0.549-1.490)	1.763 (1.166-2.075)	0.289 (0.211-0.488)	0.586 (0.323-0.755)
Hippocampus	1.823 (1.169-2.548)	3.352 (2.24-4.47)	0.441 (0.358-0.643)	0.768 (0.515-0.993)

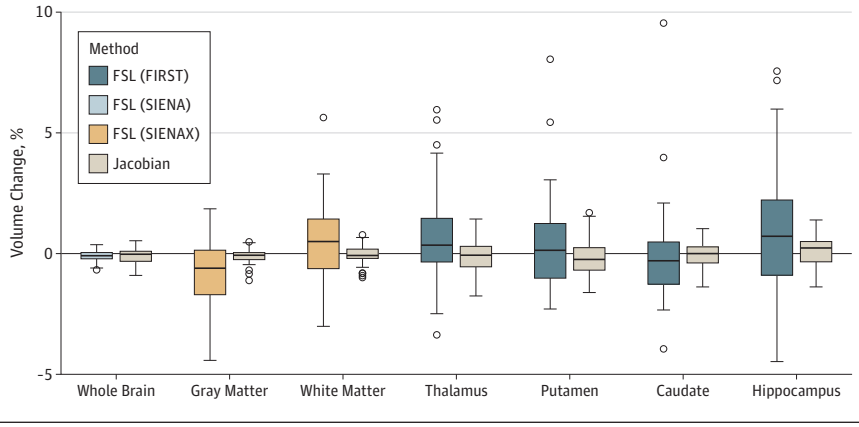
Abbreviations: FIRST, FMRIB's Integrated Registration and Segmentation Tool; FSL, Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library; SIENA, Structural Image Evaluation, Using Normalisation, of Atrophy; SIENA-X, Structural Image Evaluation, Using Normalisation, of Atrophy–Cross-Sectional.

[a] The magnitude of the deviation from 0 (true change) represents the error in the measurements. The data represents the 50th and 75th percentiles of the absolute difference between the first and second scans and 95% CIs using the bias-corrected and accelerated bootstrap. A cohort of 34 healthy volunteers were assessed.

[b] For main brain structures, FSL includes SIENA and SIENA-X approaches.

[c] For deep gray matter structures, FSL is by the FIRST approach.

Figure 2. Box Plots of the Repeatability of the Jacobian Integration and FSL Methods for Brain Volume Change In Healthy Volunteers



The figure represents the median percentage change, interquartile ranges, and the outliers of the short-term test-retest errors in the brain volume estimates for healthy volunteers (n = 34). FSL indicates Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library; FIRST, FMRIB's Integrated Registration and Segmentation Tool; SIENA, Structural Image Evaluation, Using Normalisation, of Atrophy; SIENA-X, Structural Image Evaluation, Using Normalisation, of Atrophy–Cross-Sectional.

Limited Reproducibility of Brain Volume Estimates and Use in Individuals

As registration-based methods, FSL and JI produced similar short-term test-retest errors for whole-brain volume changes (FSL: 75th percentile, 0.27; JI: 75th percentile, 0.36). The estimates with JI were more precise (with a smaller IQR) and less biased (with a median closer to 0) than those with FSL for all the brain regions except the putamen, for which JI was more precise but FIRST was less biased. The error in assessing the gray matter volume was nearly 5-fold higher for SIENA-X (75th percentile, 1.78) than for JI (75th percentile, 0.38) (**Table 2** and **Figure 2**). Coregistration of two 3D-T1 scans before running SIENA-X did not improve the results (eTable 3 in the Supplement).

Comparison of Algorithms for Brain Volume Change Assessments

An evaluation of the correlations between the changes in brain volume quantified with JI and FSL yielded a high correlation for the whole-brain volume changes (R^2 = 0.92; P < .001). By contrast, the correlations for different brain regions were low, with a maximum R^2 = 0.23 for the thalamus (eFigure 4 in the Supplement).

Brain Volume Cutoffs per Estimation Method

Finally, we evaluated the performance of the annual rates of brain volume changes estimated with FSL and JI to correctly classify 20 healthy volunteers (with a median [IQR] age of 33.0 [28.7-39.2] years; 10 female [50%]) and 100 patients with MS (median [IQR] age = 35.5 [32.6-41.3] years; 69 female [69%]) from the 140 patients with MS who were matched by age (P = .22) and sex (P = .17). Although the area under curve for whole-brain volume changes was similar for both methods, the optimal thresholds differed (**Table 3**).

Finally, we found different sensitivity and specificity for the cutoffs described previously for annual whole-brain volume changes[28] using the FSL (SIENA) and JI methods when

Table 3. Discrimination and Calibration Estimates for Annualized Rates of the Change in Whole-Brain and Regional Volume to Predict Multiple Sclerosis or Healthy Status

	FSL					Jacobian Integration				
Cutoff	AUC	Threshold	Specificity, %	Sensitivity, %	*P* Value[a]	AUC	Threshold	Specificity, %	Sensitivity, %	*P* Value
By region[b]										
Whole brain	0.632	0.335	75	53	.55	0.604	0.568	90	40	.10
Gray matter	0.518	0.314	85	30	.76	0.550	0.908	95	24	.04
Thalamus	0.558	−3.715	20	99	.57	0.609	0.831	70	53	.79
Putamen	0.615	2.077	95	29	.16	0.655	0.82	80	53	.55
Caudate	0.540	1.716	85	32	.41	0.670	0.508	85	55	.31
Hippocampus	0.535	−1.076	85	36	.30	0.670	−0.818	40	91	>.99
By value										
−0.52	NA	NA	95/85[c]	49/38[c]	NA	NA	NA	80[e]	41[e]	NA
−0.46	NA	NA	90/75[c]	56/43[c]	NA	NA	NA	80[e]	46[e]	NA
−0.40	NA	NA	80/75[c]	65/48[c]	NA	NA	NA	75[e]	50[e]	NA
−0.37[d]	NA	NA	80/75[c]	67/51[c]	NA	NA	NA	70[e]	53[e]	NA

Abbreviations: AUC, area under curve; BVL, brain volume loss; FIRST, FMRIB's Integrated Registration and Segmentation Tool; FSL, Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library; NA, not applicable; SIENA, Structural Image Evaluation, Using Normalisation, of Atrophy; SIENA-X, Structural Image Evaluation, Using Normalisation, of Atrophy–Cross-Sectional.

[a] Hosmer–Lemeshow goodness-of-fit test used 10 quantiles (standard) and 8 *df*. When $P < .05$, the null hypothesis of good calibration was rejected.

[b] Values presented by region were generated via SIENA, SIENA-X, and FIRST methods.

[c] Numbers in these cells are from De Stefano et al[28] (first number) and 20 healthy volunteers and 100 age-matched, sex-matched patients with MS in this analysis (second number); all values were generated via the SIENA approach.

[d] Best cutoff from De Stefano et al.[28]

[e] Numbers in these cells are from 20 healthy volunteers and 100 age-matched, sex-matched patients with MS in this analysis; all values were generated via the SIENA approach.

quantifying brain volume changes in the same study population. A Hosmer-Lemeshow goodness-of-fit test produced nonsignificant *P* values (favoring acceptable calibration) for all models except the gray matter volume change using the JI method (area under curve, 0.550; $P = .04$; Table 3).

Sensitivity Analyses

We included 107 participants (76.4%) with 4 MRI scans, 21 participants (15%) with 3 MRI scans, and 12 participants (8.6%) with 2 MRI scans. We found similar results from the complete case analyses (eFigure 5 in the Supplement) and the results were similar in the models after excluding participants with extreme changes (eFigure 6 in the Supplement).

Discussion

There were 3 main findings from this study: (1) the rates of BVL were highest during the first 5 years of the disease, with a significant association with age at MS onset and of steroids on the gray matter volume change; (2) repeatability was similar for FSL (SIENA) and JI when estimating whole-brain volume changes, but JI provided more precise and less biased estimates for gray matter and deep gray matter structures than FSL (by SIENA-X and FIRST, respectively); and (3) estimates from different image processing methods may not be directly comparable, as evident through the low R^2 for brain regions quantified by JI and FSL, together with the distinct performance of JI and FSL to assess the annualized brain volume change to discriminate patients with MS from healthy volunteer participants.

These findings may have implications for the therapeutic decision-making process, because when the dynamics of BVL are considered, a fixed cutoff (eg, < 0.4%[28]) may underestimate the therapeutic response in the first 5 years and overestimate the therapeutic response thereafter. Thus, different targets may be necessary at different stages of MS. Moreover, our results suggest that the younger the patient is at MS onset, the faster the rate of gray matter volume loss, consistent with previous reports.[29] Indeed, this effect was particularly prominent during the first 5 years of MS. We did not find a significant association with steroids on the rate of brain volume change during follow-up, although a significant effect on volume loss was evident in the thalamus (but not in other gray matter regions), and there was only a marginally significant effect on cortical gray matter volume loss. Both treatments have been associated with accelerated BVL, a phenomenon called *pseudoatrophy* that can be explained by resolution of the ongoing edema and inflammation at the time treatment is initiated. In this situation, the time elapsed between treatment initiation and the MRI scan is crucial.

In addition, the use of a DMT significantly decrease brain volume in the first 6 months of therapy.[30] However, steroids may produce a transient and reversible change in the estimated brain volume for approximately 1 month,[31] and particularly when high-dose tapering of oral prednisolone was used for a couple of weeks after high-dose intravenous steroids.[32] In this study, few patients started or shifted DMT use during the follow-up period, and more importantly, patients were recruited after 2 months of the use of steroids or the presence of relapses, and yearly routine examinations were performed under stable conditions. Therefore, although we cannot com-

pletely rule out a pseudoatrophy effect, it is likely to have had a limited effect in our study.

Since we consider inflammation to be the main driver for neuroaxonal injury in MS, we expect to find the highest rates of BVL in patients with the strongest MS activity. In fact, patients with either new T2 or gadolinium-enhanced lesions, particularly those with disease duration less than 5 years, had higher rates of retinal thinning than stable patients with MS.[33] Considering that steroids and DMT are usually prescribed to patients with active MS, the increased rate of BVL in patients using these drugs may reflect the association in our study between inflammation and neuroaxonal injury in MS.

We did not find any model for the rate of change in white matter volume. In MS, neuroaxonal loss may be associated with inflammatory lesions, astrogliosis, and myelin loss,[34] particularly in the white matter. Myelin loss contributes to BVL, while new inflammatory lesions and astrogliosis may compensate for it. The different relative contribution of these pathological substrates to changes in brain volume across participants may explain the failure to generate a suitable model for the changes in white matter volume.

Given the expected BVL of 0.5% to 1.35% per year[12] in MS, the accuracy of JI and FSL to estimate brain volume would argue against their use at the individual level. To compare BVL at the population level, the repeatability of whole-brain volume changes assessed with JI did not surpass that with FSL (SIENA). Although the gray matter displays stronger association with disability[11] and a weaker effect of pseudoatrophy[35] than the whole-brain volume, longitudinal studies often focus on the volume change in the whole brain rather than the gray matter because the reproducibility of SIENA-X is 4-fold worse than that of SIENA.[36] In this study, the JI method estimated gray matter volume changes more precisely and with less bias than FSL did, and the estimates for the cortical gray matter were similar to those for the estimates of whole-brain changes. Consequently, we recommend using registration-based technologies like JI to estimate gray matter volume changes in longitudinal studies, benefiting from the gray matter measurements without the large penalty error associated with registration-based technologies.

Finally, estimating BVL using segmentation and registration-based methods is not equivalent. Indeed, the best threshold of the annual rate of BVL to discriminate healthy volunteers from patients with MS differed in the same study population with the 2 registration-based methods (FSL and JI), suggesting that the variability in the technologies should be considered before establishing the rate of BVL to use as a therapeutic goal in MS.

Our study has several strengths. First, we evaluated the dynamic rate of changes in brain volume without any a priori assumption, testing linear and nonlinear models. Second, the estimates of BVL were made blind to any MS-associated characteristics. Third, we compared the repeatability and performance of 2 volumetric imaging methods in the same study population and with the same MRI protocol and scanner. Fourth, our sensitivity analyses suggest our results are robust to missing and extreme data. Finally, we provided estimates of measurement error, useful when estimating sample size for randomized clinical trials.

Limitations

Our study also has some limitations, not least of all that our sample size was insufficient to analyze relapsing vs progressive MS separately and model changes after more than 25 to 30 years of disease duration. Moreover, we did not find a model for the white matter volume change, which may indicate that the sample size is too small for such a complex model. In this context, the acquisition of additional time points or of 2 T1s (with repositioning) at each scan session may have been useful. We also did not assess repeatability in patients with MS, although similar test-retest error measurements have already been reported for whole-brain volume changes in patients with MS assessed with JI (MS metrics) and FSL (SIENA).[23] In addition, we did not evaluate other methods such as Freesurfer,[37] Statistical Parametric Mapping,[38] or the new unreleased version of SIENA-XL (a longitudinal version of SIENA-X),[39] sequences other than MPRAGE, or different scanners and strengths. Similar errors were found with JI (MSmetrics) and FSL (SIENA) using 1.5-T and 3-T scanners,[23] yet the impact of using scanners with different field strength on the error in BVL estimation was an order of magnitude lower for JI than for FSL (SIENA). Finally, we used a short-term rescan approach to assess the repeatability of the brain volume estimations, while a long-term rescan method would have been preferable to account for changes in brain shape or other long-term changes that were not addressed by our approach.

Conclusions

As a marker of therapeutic response, BVL is an appealing proposal. However, the threshold of 0.4% per year proposed as a therapeutic goal should be reconsidered based on the nonlinear dynamics of BVL, its limited reproducibility, and the variability with different methods. Future multicenter collaborations should address these issues with larger samples of patients with MS, including all phenotypes and disease stages, and should image them with different MRI scans or protocols, as well as using different MRI software and algorithms.

ARTICLE INFORMATION

Accepted for Publication: May 4, 2018.

Published Online: July 2, 2018. doi:10.1001/jamaneurol.2018.1596

Author Affiliations: Center of Neuroimmunology Department of Neurology, Hospital Clinic of Barcelona, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain (Andorra, Lampert, Pulido-Valdeolivas, Zubizarreta, Llufriu, Martinez-Heras, Sola-Valls, Sepulveda, Tercero-Uribe, Blanco, Saiz, Villoslada, Martinez-Lapiscina); Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio (Nakamura); Cleveland Clinic, Lerner College of Medicine, Cleveland, Ohio (Lampert); Now with Genentech, Inc, South San Francisco, California (Villoslada).

Author Contributions: Dr Martinez-Lapiscina had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Andorra, Tercero-Uribe, Saiz, Villoslada, Martinez-Lapiscina.

Acquisition, analysis, or interpretation of data: Andorra, Nakamura, Lampert, Pulido-Valdeolivas, Llufriu, Martinez-Heras, Sola-Valls, Sepulveda, Blanco, Saiz, Villoslada, Martinez-Lapiscina, Zubizarreta.
Drafting of the manuscript: Andorra, Villoslada, Martinez-Lapiscina.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Andorra, Pulido-Valdeolivas, Martinez-Lapiscina.
Obtained funding: Villoslada, Martinez-Lapiscina.
Administrative, technical, or material support: Pulido-Valdeolivas, Sola-Valls, Blanco.
Supervision: Martinez-Heras, Sepulveda, Tercero-Uribe, Villoslada, Martinez-Lapiscina.

Conflict of Interest Disclosures:
Dr Martinez-Lapiscina reports serving as a researcher in the Optical Coherence Tomography Trial in Multiple Sclerosis study, an observational study that involves no specific drugs sponsored by Novartis. Dr Martinez-Lapiscina reports having received honoraria from Biogen and Genzyme for speaking and a travel reimbursement from Genzyme and Roche for international and national meetings in the last 3 years; she is a member of the working committee of International Multiple Sclerosis Visual System Consortium. Dr Solà-Valls reports having received compensation for consulting services and speaker honoraria from Sanofi-Aventis, Bayer-Schering, Novartis, and Biogen-Idec. Dr Pulido-Valdeolivas reports having received travel reimbursement from Roche and Genzyme and holding stock in Aura Innovative Robotics. Dr Saiz reports having received compensation for consulting services and speaker honoraria from Bayer-Schering, Merck-Serono, Biogen-Idec, Sanofi-Aventis, Teva Pharmaceutical Industries Ltd, and Novartis. Dr Villoslada reports having received compensation for consulting services from Novartis, Biogen, Roche, and Genzyme; having received unrestricted research grants from Novartis, Biogen and Genzyme; holding stocks in Bionure Inc, Spire Bioventures, Mint-Labs, and Health Engineering; serving as an academic editor of *Multiple Sclerosis and Demyelinating Diseases*; and holding current employment at Genentech Inc. Dr Llufriu reports having received speaker honoraria from Biogen Idec, Novartis, Teva, Genzyme, and Merck. Dr Nakamura reports having received speaking honoraria from Sanofi Genzyme, consulting fees from NeuroRx Research, and royalty fees for license from Biogen. No other disclosures were reported.

Funding/Support: This study was supported by the Instituto de Salud Carlos III, Spain and Fondo Europeo de Desarrollo Regional (grants PI15/0061, Dr Villoslada; JR16/00006 and PI15/00587, Dr Martinez-Lapiscina; FI16/00168, Mr Andorra; and FI16/00251, Dr Sola-Valls), the CERCA Program of the Generalitat de Catalunya (Dr Villoslada), Grant for MS Innovation (Dr Martinez-Lapiscina), Fundació Privada Cellex (Dr Martinez-Lapiscina), Marató TV3 Charitable Foundation (Dr Martinez-Lapiscina), Predoctoral Grant for Health Research (PFIS; Mr Andorra and Dr Sola-Valls), Red Española de Esclerosis Múltiple (grant RD16/0015/0002, Dr Saiz), and a Proyecto de Investigación en Salud (grant FIS 2015-PI15/00587; Drs Saiz and Llufriu), which was integrated in the Plan Estatal de Investigación Científica y Técnica de Innovación 2013-2016 and cofunded by the Instituto de Salud Carlos III, subdirección General de Evaluación, and the Fondo Europeo de Desarrollo Regional ("Otra Manera de Hacer Europa").

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank the Fundació Cellex (Barcelona, Spain) for their support of Dr Martinez-Lapiscina's research activity and Elena Abad, PhD, Institut d'Investigacions Biomèdiques August Pi Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain, for her advice and discussion regarding the statistical models and approach to analysis. Dr Abad was not compensated by the funders for her contribution.

REFERENCES

1. Ziemssen T, Derfuss T, de Stefano N, et al. Optimizing treatment success in multiple sclerosis. *J Neurol*. 2016;263(6):1053-1065. doi:10.1007/s00415-015-7986-y

2. Giovannoni G, Turner B, Gnanapavan S, Offiah C, Schmierer K, Marta M. Is it time to target no evident disease activity (NEDA) in multiple sclerosis? *Mult Scler Relat Disord*. 2015;4(4):329-333. doi:10.1016/j.msard.2015.04.006

3. Hauser SL, Oksenberg JR. The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. *Neuron*. 2006;52(1):61-76. doi:10.1016/j.neuron.2006.09.011

4. Tallantyre EC, Bø L, Al-Rawashdeh O, et al. Clinico-pathological evidence that axonal loss underlies disability in progressive multiple sclerosis. *Mult Scler*. 2010;16(4):406-411. doi:10.1177/1352458510364992

5. Martinez-Lapiscina EH, Arnow S, Wilson JA, et al; IMSVISUAL consortium. Retinal thickness measured with optical coherence tomography and risk of disability worsening in multiple sclerosis: a cohort study. *Lancet Neurol*. 2016;15(6):574-584. doi:10.1016/S1474-4422(16)00068-5

6. Bovis F, De Stefano N, Steinerman JR, Knappertz V, Sormani MP. Validating the use of brain volume cutoffs to identify clinically relevant atrophy in RRMS[published online November 1, 2017]. *Mult Scler*. doi:10.1177/1352458517743801

7. Tsivgoulis G, Katsanos AH, Grigoriadis N, et al. The effect of disease modifying therapies on brain atrophy in patients with relapsing-remitting multiple sclerosis: a systematic review and meta-analysis. *PLoS One*. 2015;10(3). doi:10.1371/journal.pone.0116511

8. Kappos L, De Stefano N, Freedman MS, et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. *Mult Scler*. 2016;22(10):1297-1305. doi:10.1177/1352458515616701

9. De Stefano N, Giorgio A, Battaglini M, et al. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. *Neurology*. 2010;74(23):1868-1876. doi:10.1212/WNL.0b013e3181e24136

10. Kalkers NF, Ameziane N, Bot JCJ, Minneboo A, Polman CH, Barkhof F. Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype. *Arch Neurol*. 2002;59(10):1572-1576. doi:10.1001/archneur.59.10.1572

11. Fisher E, Lee J-C, Nakamura K, Rudick RA. Gray matter atrophy in multiple sclerosis: a longitudinal study. *Ann Neurol*. 2008;64(3):255-265. doi:10.1002/ana.21436

12. De Stefano N, Airas L, Grigoriadis N, et al. Clinical relevance of brain volume measures in multiple sclerosis. *CNS Drugs*. 2014;28(2):147-156. doi:10.1007/s40263-014-0140-z

13. Nakamura K, Guizard N, Fonov VS, Narayanan S, Collins DL, Arnold DL. Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis. *Neuroimage Clin*. 2013;4:10-17. doi:10.1016/j.nicl.2013.10.015

14. Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage*. 2002;17(1):479-489. doi:10.1006/nimg.2002.1040

15. Smith SM, Jenkinson M, Woolrich MW, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004;23(suppl 1):S208-S219. doi:10.1016/j.neuroimage.2004.07.051

16. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*. 2011;56(3):907-922. doi:10.1016/j.neuroimage.2011.02.046

17. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald criteria". *Ann Neurol*. 2005;58(6):840-846. doi:10.1002/ana.20703

18. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol*. 2011;69(2):292-302. doi:10.1002/ana.22366

19. Martínez-Lapiscina EH, Fraga-Pumar E, Gabilondo I, et al. The multiple sclerosis visual pathway cohort: understanding neurodegeneration in MS. *BMC Res Notes*. 2014;7(1):910. doi:10.1186/1756-0500-7-910

20. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-1128. doi:10.1016/j.neuroimage.2006.01.015

21. Sdika M, Pelletier D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Hum Brain Mapp*. 2009;30(4):1060-1067. doi:10.1002/hbm.20566

22. Boyes RG, Rueckert D, Aljabar P, et al. Cerebral atrophy measurements using Jacobian integration: comparison with the boundary shift integral. *Neuroimage*. 2006;32(1):159-169. doi:10.1016/j.neuroimage.2006.02.052

23. Lysandropoulos AP, Absil J, Metens T, et al. Quantifying brain volumes for multiple sclerosis patients follow-up in clinical practice - comparison of 1.5 and 3 Tesla magnetic resonance imaging. *Brain Behav*. 2016;6(2):e00422. doi:10.1002/brb3.422

24. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;(67):1-48. doi:10.18637/jss.v067.i01

25. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19 (6):716-723. doi:10.1109/TAC.1974.1100705

26. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods Ecol Evol*. 2013;4(2):133-142. doi:10.1111/j.2041-210x.2012.00261.x

27. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

28. De Stefano N, Stromillo ML, Giorgio A, et al. Establishing pathological cut-offs of brain atrophy rates in multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2015;87(1):93-99. doi:10.1136/jnnp-2014-309903

29. Tedeschi G, Lavorgna L, Russo P, et al. Brain atrophy and lesion load in a large population of patients with multiple sclerosis. *Neurology*. 2005; 65(2):280-285. doi:10.1212/01.wnl.0000168837.87351.1f

30. Zivadinov R, Reder AT, Filippi M, et al. Mechanisms of action of disease-modifying agents and brain volume changes in multiple sclerosis. *Neurology*. 2008;71(2):136-144. doi:10.1212/01.wnl.0000316810.01120.05

31. Rao AB, Richert N, Howard T, et al. Methylprednisolone effect on brain volume and enhancing lesions in MS before and during IFNbeta-1b. *Neurology*. 2002;59(5):688-694. doi:10.1212/WNL.59.5.688

32. Hoogervorst EL, Polman CH, Barkhof F. Cerebral volume changes in multiple sclerosis patients treated with high-dose intravenous methylprednisolone. *Mult Scler*. 2002;8(5):415-419. doi:10.1191/1352458502ms838oa

33. Ratchford JN, Saidha S, Sotirchos ES, et al. Active MS is associated with accelerated retinal ganglion cell/inner plexiform layer thinning. *Neurology*. 2013;80(1):47-54. doi:10.1212/WNL.0b013e31827b1a1c

34. Stadelmann C, Albert M, Wegner C, Brück W. Cortical pathology in multiple sclerosis. *Curr Opin Neurol*. 2008;21(3):229-234. doi:10.1097/01.wco.0000318863.65635.9a

35. Fisher E, Nakamura K, Lee J-C, You X, Sperling B, Rudick RA. Effect of intramuscular interferon beta-1a on gray matter atrophy in relapsing-remitting multiple sclerosis: A retrospective analysis. *Mult Scler*. 2016;22(5): 668-676. doi:10.1177/1352458515599072

36. Cover KS, van Schijndel RA, van Dijk BW, et al; neuGRID; Alzheimer's Disease Neuroimaging Initiative. Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans. *Psychiatry Res*. 2011;193(3):182-190. doi:10.1016/j.pscychresns.2011.02.012

37. Reuter M, Rosas HD, Fischl B. Highly accurate inverse consistent registration: a robust approach. *Neuroimage*. 2010;53(4):1181-1196. doi:10.1016/j.neuroimage.2010.07.020

38. Penny WD, Friston KJ, Ashburner J, Kiebel SJ, Nichols T. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Amsterdam, the Netherlands: Elsevier/Academic Press; 2006.

39. Battaglini M, Jenkinson M, De Stefano N; Alzheimer's Disease Neuroimaging Initiative. SIENA-XL for improving the assessment of gray and white matter volume changes on brain MRI. *Hum Brain Mapp*. 2017;39(3): 1063-1077. doi:10.1002/hbm.23828