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; Maria 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 (R²) 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 (R² = 0.65 for whole-brain volume change and R² = 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 (R² = 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.
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. Consequently, it has been proposed that a possible aim when treating MS is to achieve no evident disease activity (NEDA-3), 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, 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, brain volume measurements 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. 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.

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 or if it progresses faster at any particular stage of MS. 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. 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) and Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL). As secondary aims, we assessed the repeatability and the compatibility of the JI and FSL methods to estimate the changes in brain volume.

**Methods**

**Study Population**

In this study, the first 147 consecutive patients with MS (with conditions defined according to the McDonald criteria) who were enrolled into the prospective MS-VisualPath cohort 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. 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 mm³ were used to estimate all the volumes, and 3-D T2-FLAIR images 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 neurorlogist 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). 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. Finally, we used the JI and FSL methods 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 previously23 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 pseudod2 defined by Nakagawa and Schielzoth26 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 characteristic curves 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 Criterion25 (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).
Demographic-Associated and MS-Associated Variables and Brain Volume Change Dynamics

We did not find a significant association with sex ($\beta = -0.091; P = .21$), age at inclusion ($\beta = 0.001; P = .92$), DMT (low-to-intermediate-potency drugs: $\beta = 0.105; P = .12$; high-potency drugs: $\beta = 0.061; P = .63$), or steroid use ($\beta = 0.073; P = .19$) 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 conditioned 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 ($\beta = 0.051; P = .47$) or age at MS onset ($\beta = -0.008; P = .06$) 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 and MS onset and MS duration was marginally significant (modeled as a binary variable with a cutoff in 5 years or >5 years: $\beta = 0.015; P = .047$). 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: $\beta = -0.003$; interaction: $\beta = -0.015$), whereas the effect for those with an MS...
duration longer than 5 years was a decrease of 0.003% per year (age: \( \beta = -0.003 \)).

We did not find a significant association with DMT (low-to-intermediate-potency drugs: \( \beta = 0.015; P = .83 \); high-potency drugs: \( \beta = -0.088; P = .52 \)) or steroid use (\( \beta = 0.110; P = .09 \)) in an assessment with no interaction with MS duration (\( \beta = 0.164; P \text{ value} = .23 \)). 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 (\( \beta = 0.280; P = .02 \)), without interaction with disease duration (\( \beta = 0.145; P = .56 \)). Moreover, we found a significant association with age at MS onset (\( \beta = 0.019; P = .01 \)) in the dynamics of hippocampal volume change, yet no interaction with disease duration (\( \beta = -0.002; P = .86 \)). We did not find any other significant associations for the other brain regions tested.
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 using the FSL (SIENA) and JI methods when...
quantifying brain volume changes in the same study population. A Hosmer-Lemeshow goodness-of-fit test produced non-significant P values (favoring acceptable calibration) for all models except the gray matter volume change using the JI method (area under curve, 0.550; 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 R² 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%) 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. 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. However, steroids may produce a transient and reversible change in the estimated brain volume for approximately 1 month, particularly when high-dose tapering of oral prednisolone was used for a couple of weeks after high-dose intravenous steroids. 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-

### 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

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>AUC</th>
<th>Threshold</th>
<th>Specificity, %</th>
<th>Sensitivity, %</th>
<th>P Value*</th>
<th>AUC</th>
<th>Threshold</th>
<th>Specificity, %</th>
<th>Sensitivity, %</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>By region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole brain</td>
<td>0.632</td>
<td>0.335</td>
<td>75</td>
<td>53</td>
<td>.55</td>
<td>0.504</td>
<td>0.506</td>
<td>90</td>
<td>40</td>
<td>.10</td>
</tr>
<tr>
<td>Gray matter</td>
<td>0.518</td>
<td>0.314</td>
<td>85</td>
<td>30</td>
<td>.76</td>
<td>0.550</td>
<td>0.908</td>
<td>95</td>
<td>24</td>
<td>.04</td>
</tr>
<tr>
<td>Thalamus</td>
<td>0.558</td>
<td>−3.715</td>
<td>20</td>
<td>99</td>
<td>.57</td>
<td>0.609</td>
<td>0.831</td>
<td>70</td>
<td>53</td>
<td>.79</td>
</tr>
<tr>
<td>Putamen</td>
<td>0.615</td>
<td>2.077</td>
<td>95</td>
<td>29</td>
<td>.16</td>
<td>0.655</td>
<td>0.82</td>
<td>80</td>
<td>53</td>
<td>.55</td>
</tr>
<tr>
<td>Caudate</td>
<td>0.540</td>
<td>1.716</td>
<td>85</td>
<td>32</td>
<td>.41</td>
<td>0.670</td>
<td>0.508</td>
<td>85</td>
<td>55</td>
<td>.31</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>0.535</td>
<td>−1.076</td>
<td>85</td>
<td>36</td>
<td>.30</td>
<td>0.670</td>
<td>−0.818</td>
<td>40</td>
<td>91</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>By value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−0.52</td>
<td>NA</td>
<td>NA</td>
<td>95/85 ⋆</td>
<td>49/38 ⋆</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>80°</td>
<td>41°</td>
<td>NA</td>
</tr>
<tr>
<td>−0.46</td>
<td>NA</td>
<td>NA</td>
<td>90/75 ⋆</td>
<td>56/43 ⋆</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>80°</td>
<td>46°</td>
<td>NA</td>
</tr>
<tr>
<td>−0.40</td>
<td>NA</td>
<td>NA</td>
<td>80/75 ⋆</td>
<td>65/48 ⋆</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>75°</td>
<td>50°</td>
<td>NA</td>
</tr>
<tr>
<td>−0.37 ⋆</td>
<td>NA</td>
<td>NA</td>
<td>80/75 ⋆</td>
<td>67/51 ⋆</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>70°</td>
<td>53°</td>
<td>NA</td>
</tr>
</tbody>
</table>


* 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.

** Best cutoff from De Stefano et al. 28

* 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.
completely 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, astroglisis, and myelin loss,34 particularly in the white matter. Myelin loss contributes to BVL, while new inflammatory lesions and astroglisis 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,22 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 disability11 and a weaker effect of pseudoatrophy35 (SIENA). Although the gray matter displays stronger association with disability11 and a weaker effect of pseudoatrophy35 (SIENA-X),39 sequences other than MPRAGE, or different 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 (MSmetrics) and FSL (SIENA).23 In addition, we did not evaluate other methods such as Freesurfer,27 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 re-scan 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.
Acquisition, analysis, or interpretation of data: Andorra, Nakamura, Lampert, Pulido-Valdeolivas, Llurfi, Martinez-Heras, Solà-Valls, Sepulveda, Blanco, Saiz, Villoslada, Martínez-Lapiscina, Zubizarreta. 

Drafting of the manuscript: Andorra, Villoslada, Martínez-Lapiscina. 

Critical revision of the manuscript for important intellectual content: All authors. 

Statistical analysis: Andorra, Pulido-Valdeolivas, Martínez-Lapiscina. 

Obtained funding: Villoslada, Martínez-Lapiscina. 

Administrative, technical, or material support: 

\[\text{General de Evaluación, and the Fondo Europeo de Desarrollo Regional ("Otra Manera de Hacer Europa")}.\]

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 Martínez-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 


