



OPEN ACCESS

EDITED BY

Sairam Geethanath,
Johns Hopkins University, United States

REVIEWED BY

Amasha Konar Shridhar,
Columbia University, United States
Pavan Poojar,
Icahn School of Medicine at Mount Sinai,
United States

*CORRESPONDENCE

Miguel D'haeseleer
✉ miguel.dhaeseleer@amscenter.be

RECEIVED 29 May 2025

ACCEPTED 19 September 2025

PUBLISHED 16 October 2025

CITATION

Temmerman J, Vanremoortel A, Bjerke M, D'Hooghe MB, Nagels G, Pauwels A, Sima D, Smeets D, Schependom JV, Wittens MMJ, Engelborghs S and D'haeseleer M (2025) Real-world brain volumetry in multiple sclerosis: importance of methodological consistency and clinical relevance of gray matter atrophy. *Front. Neurol.* 16:1637835. doi: 10.3389/fneur.2025.1637835

COPYRIGHT

© 2025 Temmerman, Vanremoortel, Bjerke, D'Hooghe, Nagels, Pauwels, Sima, Smeets, Schependom, Wittens, Engelborghs and D'haeseleer. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Real-world brain volumetry in multiple sclerosis: importance of methodological consistency and clinical relevance of gray matter atrophy

Joke Temmerman^{1,2,3}, Ann Vanremoortel⁴, Maria Bjerke^{1,3,5,6}, Marie B. D'Hooghe^{1,4}, Guy Nagels^{1,2,7}, Ayla Pauwels^{1,2,4}, Diana Sima^{1,7}, Dirk Smeets^{1,7}, Jeroen Van Schependom^{1,8}, Mandy M. J. Wittens^{1,2,3,7}, Sebastiaan Engelborghs^{1,2,3} and Miguel D'haeseleer^{1,2,4*}

¹Vrije Universiteit Brussel, Center for Neurosciences, NEUR and/or AIMS Research Groups, Brussels, Belgium, ²Department of Neurology, Universitair Ziekenhuis Brussel, Brussels, Belgium, ³Department of Biomedical Sciences, Universiteit Antwerpen, Antwerp, Belgium, ⁴Nationaal Multiple Sclerose Centrum, Melsbroek, Belgium, ⁵Department of Clinical Biology, Laboratory of Clinical Neurochemistry, Universitair Ziekenhuis Brussel, Brussels, Belgium, ⁶Division of Clinical Geriatrics, Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden, ⁷Icometrix, Leuven, Belgium, ⁸Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium

Background: Brain volume loss (BVL) is a marker of neurodegeneration associated with clinical disability in multiple sclerosis (MS). However, its application in routine clinical practice is limited due to measurement errors introduced by the use of different magnetic resonance imaging (MRI) scanners across and within centers.

Objective: To confirm the existence and clinical relevance of longitudinal BVL in a real-world MS cohort with scanner variability, employing a dedicated quantification pipeline combined with post-acquisition harmonization.

Methods: We analyzed MRI data from 72 MS patients scanned across multiple Belgian centers over 48–60 months. Clinical disability was assessed using the Expanded Disability Status Scale, Timed 25-Foot Walk Test, 9-Hole Peg Test (9HPT), and Symbol Digit Modalities Test. Percentage volume change (PVC) in whole brain (WB), total gray matter (TGM), cortical gray matter (CGM), and deep gray matter was quantified using the icobrain ms pipeline. A similarity index was applied to account for scanner differences. Twenty-seven healthy volunteers served as controls.

Results: No significant differences in annualized PVC were observed between MS patients and controls. Within the MS group, 9HPT performance correlated with TGM ($\rho = -0.30$, $p = 0.017$) and CGM ($\rho = -0.31$, $p = 0.015$) volume loss. Modified MS Functional Composite scores correlated with WB ($R = 0.28$, $p = 0.03$), TGM ($\rho = 0.31$, $p = 0.014$), and CGM ($\rho = 0.31$, $p = 0.013$) volume loss and could be independently predicted by these measures.

Conclusion: Using automated brain volumetry with post-acquisition harmonization to address scanner variability, we did not detect accelerated BVL in this real-world MS cohort compared to healthy individuals. Nonetheless, GM volume loss was found to be clinically relevant in MS.

KEYWORDS

multiple sclerosis, brain volume loss, gray matter, real-world, disability progression, post-acquisition harmonization

1 Introduction

Multiple sclerosis (MS), a chronic inflammatory demyelinating and degenerative disorder of the central nervous system (CNS), affects nearly three million people worldwide. It is the most common cause of non-traumatic neurological disability in young to middle-aged adults (1). Clinical deterioration is essentially driven by neuronal loss, which may be the consequence of (i) acute damage in newly-formed demyelinating lesions, resulting from recurrent autoimmune responses mediated by the peripheral immune system, and/or (ii) a more gradually installing neurodegeneration. The latter is believed to arise from a (non-exclusive) combination of mitochondrial dysfunction in chronically demyelinated axons, submeningeal lymphocytic clustering with damage to the underlying cortex, and pathogenic microglial activity around slowly expanding focal lesions and/or diffusely throughout the white matter (2, 3). MS is traditionally categorized into relapsing–remitting (RR), secondary progressive (SP) and primary progressive (PP) subtypes based on clinical presentation. In RR MS, tissue injury is primarily attributed to acute demyelinating lesions, whereas chronic neurodegeneration predominates in the progressive phenotypes. Recent literature endorses the stance that those phenotypes should not be seen as strictly separated entities but rather as a spectrum in which both key processes often occur together, albeit in varying proportions (3–6).

T1 contrast-enhancing and T2 hyperintense lesions are well-established magnetic resonance imaging (MRI) biomarkers of acute inflammation in MS (7). These markers are widely used in clinical practice to aid diagnosis and monitor disease activity; the latter refers to the occurrence of relapses or new focal lesions on MRI. However, their correlation with clinical outcomes is modest and often inconsistent, a phenomenon commonly named as the clinico-radiological paradox (7).

Brain volume loss (BVL) has emerged as a complementary MRI biomarkers, reflecting neurodegeneration in MS. BVL correlates with concurrent and future disability, both physical and cognitive, even in early stage of the disease (8, 9). Histological studies have demonstrated a strong association between cortical thickness measured on MRI and post-mortem brain samples, supporting the validity of BVL as a reliable indicator of actual brain atrophy (10).

BVL has increasingly been incorporated as primary or secondary endpoint in immunomodulating disease-modifying treatment (DMT) trials (11–14). A large meta-analysis evaluated over 13,000 patients across 13 pivotal studies. Brain volume changes were measured starting 6–12 months after treatment initiation to account for potential pseudo-atrophy bias, which is an apparent reduction in brain volume that can occur early after anti-inflammatory treatment due to resolution of oedema rather than neurodegeneration (15). The results showed that the effect of a therapy on BVL significantly correlates with its impact on disability outcomes, independent of its anti-inflammatory properties (16). To distinguish disease-related from age-related brain changes, a cut-off of -0.4% annual percentage volume change (PVC) has been proposed to define “pathological” BVL, offering 80% specificity and 65% sensitivity in categorizing patients (17).

Despite its research relevance, BVL has not yet been widely implemented in routine clinical practice (18). Technical factors, such as the use of different scanners, can introduce variability that often exceeds the magnitude of actual brain volume changes in uncontrolled settings (19). Only a few real-world studies in MS have assessed the value of longitudinal BVL using different scanners and primarily relied on statistical adjustments (20–22), which may limit the clinical applicability of their findings.

Our study addresses this gap by investigating the clinical relevance of longitudinal BVL over 4–5 years in a real-world MS cohort, explicitly accounting for scanner variability using a similarity index as a post-acquisition harmonization method. We employed the registration-based icobrain ms algorithm (icomatrix, Leuven, Belgium), a CE-marked and FDA-cleared automated method to quantify BVL in MS (23). By combining robust real-world data with reliable volumetric quantification while also controlling for scanner variation, our study advances previous works and provides actionable insights for the potential integration of BVL into clinical monitoring.

2 Methods

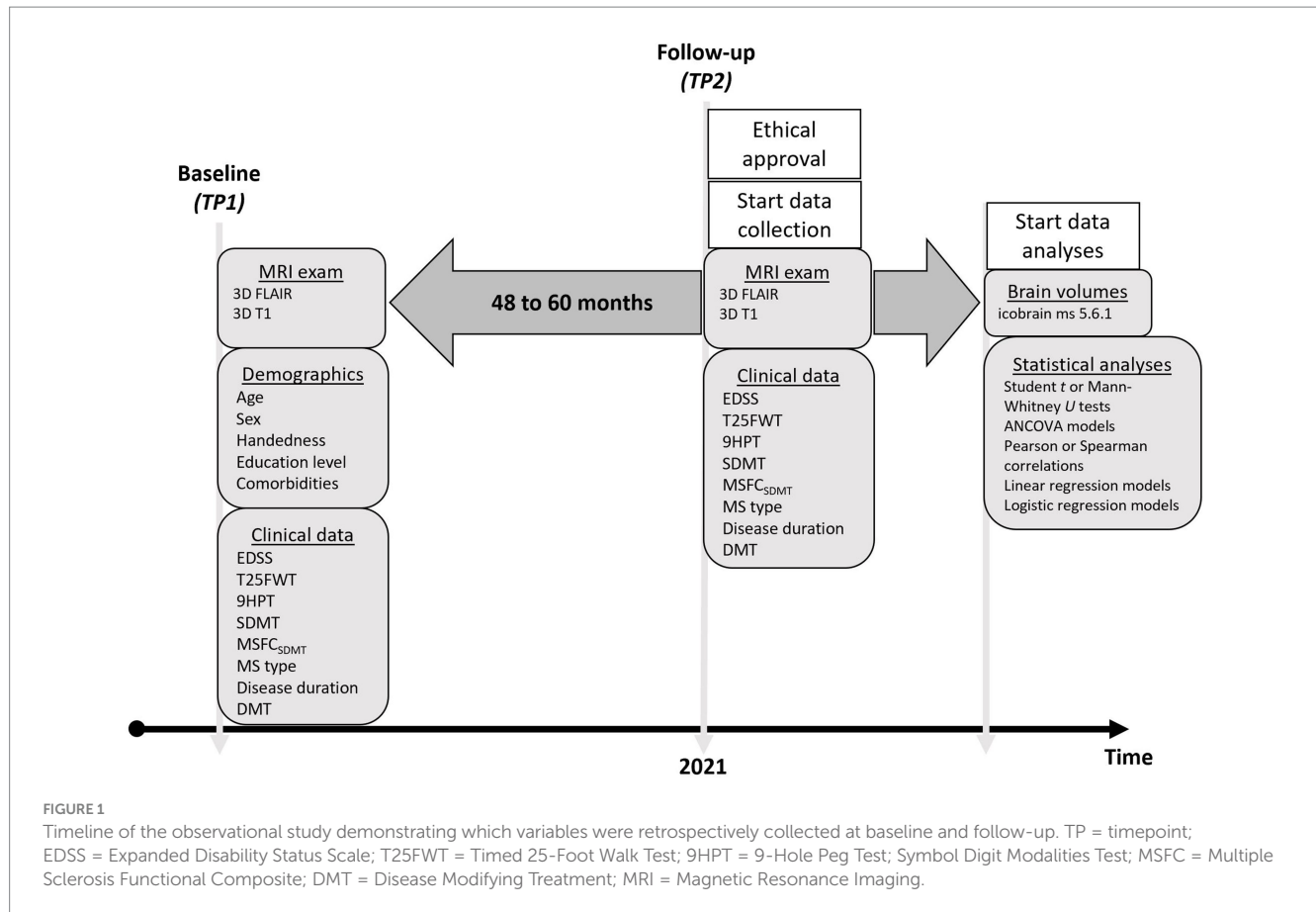
2.1 Objectives and study design

This study addressed three specific research objectives. The primary objective was (1) to assess whether annualized BVL differs between patients with MS and healthy controls (HC). The secondary objectives were (2) to investigate whether BVL is associated with the evolution of clinical outcome parameters in patients with MS, and (3) to evaluate whether baseline clinical characteristics can predict “pathological” BVL (defined as $> -0.4\%$ per year) over the observation period (17). Figure 1 provides an overview of the design of this retrospective longitudinal study. The research was conducted at the Nationaal Multiple Sclerose Centrum (NMSC) Melsbroek, a tertiary center specialized in the neurological and multidisciplinary care of patients with MS, and at the Universitair Ziekenhuis (UZ) Brussel, a university hospital, both located in Belgium. Ethical approval was granted by the Ethics Committee of the NMSC Melsbroek on June 1st 2021 (institutional authorization number: OG 033; internal reference number: EC21/06). According to the Belgian law, retrospective studies do not require participant consent. All data were de-identified prior to analysis: database and electronic health record identifiers were linked only for MRI retrieval, after which each patient was assigned a unique BRAVOLO code. All other identifiers were deleted, with a securely stored encrypted decoding file kept solely for contingency.

2.2 MS cohort

2.2.1 MRI data

The selection process for our MS cohort is shown in Figure 2. We identified all subjects between 2012 and 2021 with a diagnosis of clinically definite MS, according to the McDonald 2017 criteria (24),



of whom at least two MRI examinations were available in the clinical database of NMSC Melsbroek. MRI scans were obtained from routine clinical practice across various centers in Belgium. We selected the most recent MRI scan as our starting point (TP2). From there we went back in time, between 4 and 5 years (rounding allowed), to determine the baseline MRI (TP1). Automated brain volume quantification was performed using the icobrain MS software (version 5.6.1) for which the method and validation have been described earlier (23, 25). To be eligible for volumetric analysis using this pipeline, scans were required to include 3D fluid-attenuated inversion recovery (FLAIR) and T1-weighted sequences. Other MRI acquisition parameters were not pre-specified.

Patients were excluded if they had experienced a relapse or received a pulse steroid treatment within 30 days prior to either MRI examination, or if gadolinium-enhancing lesions were present on any scan. This was done to avoid the pseudo-atrophy effect, in which apparent brain volume reduction does not reflect actual neurodegeneration (15). The paired MRI scans were pseudonymized and transferred to the icobrain ms research server to calculate the longitudinal PVC for whole brain (WB), total gray matter (TGM), cortical gray matter (CGM), and deep gray matter (DGM) (23). We used the following formula to annualize the PVC values, which was designed in collaboration with researchers at icometrix:

$$\frac{PVC}{\left(\frac{\text{days between TP1 and TP2}}{365.25} \right)}$$

To address scanner-related variability in MRI data, we applied the similarity index as a post-acquisition harmonization approach. The similarity index is a global quality metric summarizing technical and anatomical differences between two MRI acquisitions into one variable, based on the calculation of the normalized mutual information by linear registration of those two images, with a high value indicating a higher level of similarity (26). While a threshold of 0.20 might have been better for assessing reliability at the individual level, based on a previous scan-rescan study by Sima and colleagues (26), applying this cutoff would have excluded too many participants. Therefore, in agreement with researchers from icometrix, we adopted a threshold of 0.15 to retain a sufficient number of participants while still providing robust and reliable results at the group level. The quality of MRI images and the resulting volumes were assessed using an automated quality control (QC) system designed by icometrix to flag scans that need further visual inspection for issues (27). MRI scans flagged as “approved with remarks” were then manually reviewed for potential issues, including scan artifacts (i.e., wrapping, ringing, striping, blurring, ghosting, spiking, and susceptibility artifacts), incomplete head coverage, low tissue contrast, suboptimal image alignment and high noise levels. Manual review was performed by an experienced researcher (MMJW) to determine whether they could be included in the analysis.

Matching MRI pairs were retrieved for 162 patients with MS, which were subsequently processed by the icobrain ms software. Only 44% ($n = 72$) of the initial cohort was found eligible for final inclusion based on the similarity index with cut-off 0.15 and the QC systems. In this study group, both 1.5 T and 3 T MRI were used, obtained from

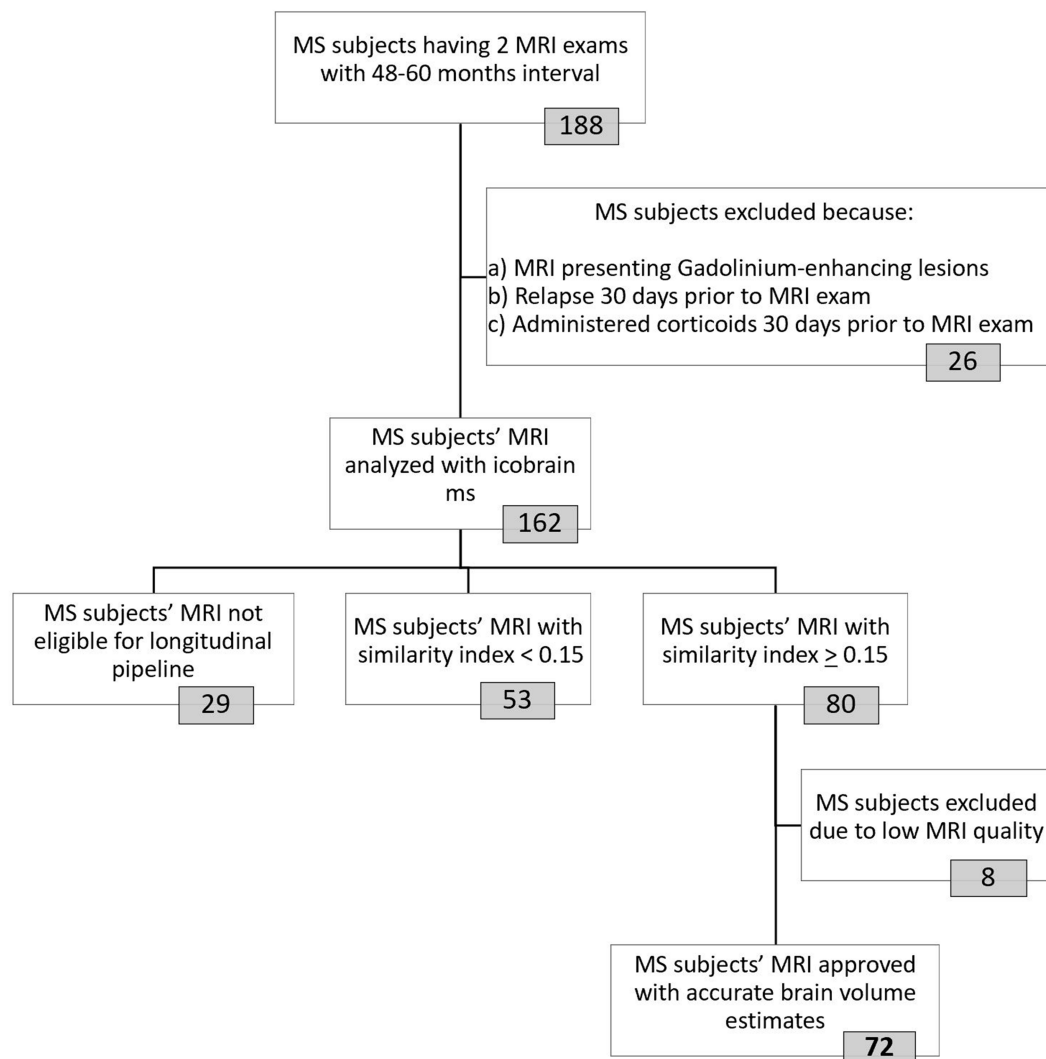


FIGURE 2

Selection procedure of final MS cohort used for data-analyses. The icobrain ms software was used to analyze MRI images from 162 included MS subjects. This fully automated registration-based method works in two steps: first, a cross-sectional pipeline enables preprocessing and segmentation (gray matter, white matter, cerebrospinal fluid) of the 3D T1-weighted MRI images at each timepoint. In the second step, the longitudinal pipeline, affine registration, non-rigid registration in both directions (from TP1-TP2 and vice versa) and Jacobian integration of deformation field allows for brain volume measurement. Images from 29 MS patients were not eligible for longitudinal analyses due to invalid/corrupt input data ($n = 13$) or differences in contrast enhancement (pre-contrast versus post-contrast T1 images at the two timepoints, $n = 16$). To account for the use of different MRI machines, a similarity index with a cut-off of 0.15 was implemented. Icometrix provides detailed reports of their findings, including results from their automated quality control that tag the images as “approved,” “approved with remarks” or “rejected.” MRI images from patients with the tag “approved with remarks” were manually double-checked to determine whether they could be included. Eight MS patients were excluded because their (cross-sectional and/or longitudinal) volume estimates were inaccurate due to low tissue contrast ($n = 6$), suboptimal alignment between images ($n = 2$) and/or failed coverage of the entire brain ($n = 1$). MRI = Magnetic Resonance Imaging; MS = Multiple sclerosis; TP = Timepoint.

four different manufacturers: Philips, Siemens, GE and Olea Medical (Supplementary Table 1). The majority of participants had no change in field strength (76%) or manufacturer (71%) between TP1 and TP2.

2.2.2 Clinical data

Demographic patient data at the time of TP1 were collected from the medical records and included age, sex, disease onset, disease duration, disease-modifying treatments (DMT; first-line: interferon beta-1A, interferon beta-1B, dimethyl fumarate, glatiramer acetate, and teriflunomide; second-line: natalizumab, fingolimod, and ocrelizumab), MS phenotype (secondary and primary progressive MS was combined as progressive MS, PMS), education level and the

presence of cardiovascular risk factors/comorbidities (smoking, diabetes mellitus, arterial hypertension, hypercholesterolemia, and obesity; the latter defined as a body mass index of 30 or more). Clinical disability was assessed using multiple outcome measures: Expanded Disability Status Scale (EDSS) for general disability (28), Timed 25-Foot Walk Test (T25FWT) for walking function (29), dominant hand 9-Hole Peg Test (9HPT) for dexterity (30), and Symbol Digit Modalities Test (SDMT) scores for cognition (31). These variables were extracted when evaluated around the same time as TP1 and TP2, not more than 6 months before or after each MRI scan. We used a composite measure at baseline and follow-up based on the Z-scores of the T25FWT, 9HPT, and SDMT, to generate a modified Multiple

Sclerosis Functional Composite (MSFC_{SDMT}) Score (32). A Z-score indicates how far a patient's score deviates from the mean value of a reference population. We derived mean values and standard deviations from the National MS Society Task Force database, which represents a broad spectrum of MS patients, to calculate these Z-scores (27). Additionally, we calculated the change in clinical scores by determining the difference between follow-up and baseline measurements, resulting in δ EDSS, δ T25FWT, δ 9HPT, δ SDMT, and δ MSFC_{SDMT} values. For clinical interpretation, a higher score on the EDSS, T25FWT, and 9HPT reflects a higher level of disability, while a higher score on the SDMT and MSFC_{SDMT} reflects a lower level of disability. Thus, positive changes (meaning higher scores at TP2 compared to TP1) indicate worsening disability according to the EDSS, T25FWT, and 9HPT, while suggesting improvement according to the SDMT and MSFC_{SDMT}. Of the 72 MS patients in our study group, 62 had complete clinical data available.

2.3 HC cohort

We used a historic HC cohort that was established via a different study, which included appropriate institutional board approval and written informed consent, from the same research team (33). In brief, volunteers had undergone two MRI exams, with an interval of at least 1 year, between 2015 and 2020, in a controlled study environment set up at UZ Brussel. Imaging was done with a 3 T scanner (Achieva, Philips Medical Systems) that included 3D T2-weighted FLAIR and T1-weighted sequence with the following parameters: 310 sagittal slices, TR = 4.939 ms, FOV = 230 × 230 mm², voxel resolution 0.53 × 0.53 × 0.5 mm³. Demographic variables were re-used, but BVL was *de novo* analyzed using the same version of the processing pipeline from icometrix as the one employed in the MS cohort.

2.4 Statistical analyses

The three study objectives are outlined above (section 2.1). For the primary endpoint of this study (1), annualized PVC for WB, TGM, CGM and DGM, were compared between patients with MS and HC. Following a Shapiro–Wilk test for normality, group differences were recorded with unpaired Student *t* or Mann–Whitney *U* tests, where appropriate. ANCOVA models were used to check for influences of potential confounders. All analyses addressing the primary objective were performed in the complete MS study group (*n* = 72). For secondary objectives requiring complete clinical data, analyses were restricted to a secondary analysis cohort (*n* = 62). Within this secondary analysis cohort, we first stratified patients with and without “disability worsening,” based on the established cut-offs for deterioration in EDSS, T25FWT, 9HPT and SDMT scores, respectively, as defined in Table 1 (34). We then investigated differences in PVC between patients with MS who showed “disability worsening” and those that did not. All subsequent secondary analyses were conducted in the entire secondary analysis cohort unless otherwise specified. For secondary objective (2), the relationship between PVC and the evolution of clinical outcome parameters was analyzed using Pearson or Spearman's rank correlation coefficients and linear regression models. The linear regression models were constructed using the change in clinical scores over time as outcome

TABLE 1 Cut-offs used to categorize MS patients as “disability worsening.”

Clinical score	Criteria for disability worsening
EDSS	Increase of 1.5 points if baseline score is 0
	Increase of 1.0 point if baseline score is between 1.0 and 5.5
	Increase of 0.5 point if baseline score is higher than 5.5
T25FWT/9HPT	Significant change of $\geq 20\%$
SDMT	Reduction of ≥ 4 points or a 10% worsening

9HPT = 9-Hole Peg Test; EDSS = Expanded Disability Status Scale; SDMT = Symbol Digit Modalities Test; T25FWT = Timed 25-Foot Walk Test. Based on Meca-Lallana et al. (34).

variables and the respective PVC measures as predictors. Scatter plots illustrating the normality of the residuals and homoskedasticity were visually checked and multicollinearity was avoided. With stepwise forward inclusion potential confounders (age, sex, cardiovascular risk factors/comorbidities, education, DMT, MS phenotype, and disease duration) were added in significant models to evaluate their influence on the relation between clinical and MRI parameters. Categorical confounders were decoded as followed: sex (female; male), presence of cardiovascular health comorbidities (none; 1; ≥ 2), education (≥ 12 years starting from elementary school = “higher”; < 12 years starting from elementary school = “lower”), DMT (none; first-line; second-line) and MS phenotype (RR; PMS). Likewise, we performed logistic regression for (secondary) objective (3), to assess whether baseline demographics or measures of clinical status (i.e., EDSS, T25FWT, 9HPT, SDMT, MSFC_{SDMT}) could be predictive for “pathological” WB volume loss (see definition above).

Statistical analyses were performed in R (version 4.0.5; Auckland, New Zealand). All reported *p* values are two-tailed with statistical significance set at 0.05. This study was conducted according to “The Strengthening the Reporting of Observational Studies in Epidemiology” (STROBE) statement for reporting observational cohort studies (Supplementary Table 2) (35).

3 Results

3.1 BVL in patients with MS versus HC

Following the selection process, our MS group consisted of 72 patients (59 RR MS and 13 PMS). Seventy-two percent was on DMT: interferon beta-1A (*n* = 4), interferon beta-1B (*n* = 10), dimethyl fumarate (*n* = 9), glatiramer acetate (*n* = 8), teriflunomide (*n* = 5), natalizumab (*n* = 10), fingolimod (*n* = 5), and ocrelizumab (*n* = 1). We observed less BVL for TGM, CGM and DGM in patients on second-line agents as compared to those on first-line. However, these differences were not statistically significant (Table 2). The HC cohort consisted of 27 subjects with a median similarity index of 0.29%. We found significant differences in median similarity index and interscan interval between both groups, but not in age, sex or annualized PVC of WB, TGM, CGM or DGM (Table 3).

Several additional *post hoc* analyses were performed to see whether potential confounders influenced the between group comparisons. We first evaluated the effect of DMT by comparing BVL between HC and MS subjects not receiving DMT (28%), but results were similar to those observed for the entire MS cohort

TABLE 2 Differences in BVL between different DMT modalities.

Brain volume measures	No DMT (<i>n</i> = 20)	First-line (<i>n</i> = 36)	Second-line (<i>n</i> = 16)	<i>p</i> -value
Annualized PVC WB	−0.18 ± 0.43	−0.17 ± 0.26	−0.18 ± 0.26	0.360
Annualized PVC TGM	−0.33 ± 0.37	−0.27 ± 0.30	−0.19 ± 0.31	0.407
Annualized PVC CGM	−0.32 ± 0.38	−0.27 ± 0.31	−0.10 ± 0.30	0.473
Annualized PVC DGM	−0.47 [0.82]	−0.50 [0.68]	−0.30 [0.97]	0.701

Data presented in mean ± SD or median [range]. CGM = Cortical Gray Matter; DGM = Deep Gray Matter; DMT = Disease-modifying treatment; PVC = Percentage Volume Change; SD = Standard Deviation; TGM = Total Gray Matter; WB = Whole Brain. First-line DMT includes dimethyl fumarate (*n* = 9), glatiramer acetate (*n* = 8), interferon beta-1A (*n* = 4), interferon beta-1B (*n* = 10) and teriflunomide (*n* = 5). Second-line treatment includes fingolimod (*n* = 5), natalizumab (*n* = 10) and ocrelizumab (*n* = 1).

TABLE 3 Demographics of MS and HC cohorts.

Characteristics	MS	HC	MS vs HC <i>p</i> -value
Number of subjects	72	27	NA
Age (years)	45 ± 9	49 ± 13	0.180
Sex (F/M)	57/15 (79.2%)	16/11 (59.3%)	0.071
MS phenotype (RR/PMS)	59 / 13	NA	NA
DMT (None/First-line/Second-line)	20/36/16	NA	NA
Interscan interval (months)	53 [43–62]	32 [21–49]	< 0.001
Similarity-index	0.21 [0.15–0.34]	0.29 [0.24–0.44]	< 0.001
Annualized PVC WB	−0.17 ± 0.31	−0.29 ± 0.27	0.055
Annualized PVC TGM	−0.27 ± 0.32	−0.35 ± 0.29	0.238
Annualized PVC CGM	−0.27 ± 0.33	−0.34 ± 0.29	0.271
Annualized PVC DGM	−0.45 [−2.30–1.32]	−0.55 [−2.22–0.11]	0.430

Continuous data presented in mean ± SD or median [range], categorical data. CGM = Cortical Gray Matter; DGM = Deep Gray Matter; DMT = Disease-modifying treatment; F = Female; HC = Healthy controls; M = Male; MS = Multiple Sclerosis; NA = Not Applicable; PMS = Progressive Multiple Sclerosis; PVC = Percentage Volume Change; RR = Relapsing–Remitting; SD = Standard Deviation; TGM = Total Gray Matter; WB = Whole Brain. First-line DMT includes dimethyl fumarate (*n* = 9), glatiramer acetate (*n* = 8), interferon beta-1A (*n* = 4), interferon beta-1B (*n* = 10) and teriflunomide (*n* = 5). Second-line treatment includes fingolimod (*n* = 5), natalizumab (*n* = 10) and ocrelizumab (*n* = 1). Significant *p*-values are shown in bold.

(Supplementary Table 3). Although the mean age of patients with MS and HC cohorts was comparable (Table 3), visual inspection of the boxplots did suggest an unequal distribution. When categorizing both cohorts by age, we observed that the HC had the highest fraction of individuals over 55 years of age, whereas this age category was a minority in the MS cohort (Supplementary Figure 1). This figure also illustrates accelerated BVL with increasing age in both groups. Additionally, the similarity index was significantly higher in the HC (Table 3), who had no scanner changes, as compared to the MS cohort, where 29% experienced a manufacturer change and 21% experienced a field strength change. To account for these confounders, an ANCOVA was performed to analyze differences in annualized PVC of WB between MS and HC cohorts, controlling for age and similarity index. Age emerged as a significant predictor of BVL ($p = 0.002$), and the similarity index had a borderline significant effect ($p = 0.053$). These findings suggest that our rather unexpected observation of similar BVL in patients with MS and HC may have been due to differences in age distribution and, to a lesser extent, in similarity index.

3.2 Brain volume change and clinical disability progression in patients with MS

We found a significant decrease in TGM and CGM PVC over time in patients demonstrating disability worsening based on the

9HPT test, as compared with those that did not (Table 4). For the complete secondary analysis cohort, we found a negative correlation between the δ 9HPT scores and PVC for TGM ($\rho = -0.30$, $p = 0.017$) and CGM ($\rho = -0.31$, $p = 0.015$). There was a positive correlation between δ MSFC_{SDMT} scores and PVC WB ($R = 0.28$, $p = 0.03$), TGM ($\rho = 0.31$, $p = 0.014$) and CGM ($\rho = 0.31$, $p = 0.013$). Regression modeling revealed that clinical worsening according to the δ MSFC_{SDMT} could be predicted by changes in WB ($\beta = 0.05$, SE = 0.02, $p = 0.03$), TGM ($\beta = 0.07$, SE = 0.02, $p = 0.002$), and CGM ($\beta = 0.07$, SE = 0.02, $p = 0.003$) volumes without being influenced by potential confounders (age, sex, education level, presence of cardiovascular comorbidities, disease duration, DMT or MS phenotype) (Table 5; Figure 3; Supplementary Table 4).

3.3 Baseline clinical variables predictive of pathological brain volume change

Annualized WB PVC was dichotomized, with “pathological” BVL below −0.4%. Twelve patients with MS (19%) demonstrated “pathological” annualized WB volume loss, whereas the remaining 50 patients (81%) had “physiological” WB volume loss. There were no significant differences observed between these two groups (Table 6) and none of the baseline characteristics could significantly predict the probability of having “pathological” BVL (Table 7).

TABLE 4 Comparison of brain volume measures between “disability worsening” and “stable disability” MS patients according to the EDSS, T25FWT, 9HPT and SDMT change over time.

Worsening based on EDSS change	Brain volume measures	Disability worsening (n = 24)	Stable disability (n = 38)	Worsening vs Stable p-value
	PVC WB	−0.85 ± 1.75	−0.68 ± 1.24	0.635
	PVC TGM	−1.67 [−3.67–1.32]	−0.86 [−6.02–1.46]	0.053
	PVC CGM	−1.59 [−3.91–1.44]	−0.82 [−6.37–1.43]	0.063
	PVC DGM	−2.44 ± 2.67	−1.40 ± 2.30	0.123
Worsening based on T25FWT change	Brain volume measures	Disability worsening (n = 36)	Stable disability (n = 26)	Worsening vs Stable p-value
	PVC WB	−0.61 ± 1.63	−0.92 ± 1.15	0.386
	PVC TGM	−0.86 [−6.02–1.32]	−1.01 [−3.60–1.46]	0.368
	PVC CGM	−0.82 [−6.37–1.44]	−1.01 [−3.61–1.43]	0.492
	PVC DGM	−1.71 ± 2.65	−1.92 ± 2.28	0.737
Worsening based on 9HPT change	Brain volume measures	Disability worsening (n = 40)	Stable disability (n = 22)	Worsening vs Stable p-value
	PVC WB	−0.89 ± 1.32	−0.47 ± 1.64	0.312
	PVC TGM	−1.10 [−6.02–1.46]	−0.37 [−2.97–1.32]	0.005*
	PVC CGM	−1.07 [−6.37–1.44]	−0.30 [−2.83–1.31]	0.003*
	PVC DGM	−1.75 [−6.63–1.59]	−1.48 [−6.82–5.64]	0.752
Worsening based on SDMT change	Brain volume measures	Disability worsening (n = 19)	Stable disability (n = 43)	Worsening vs Stable p-value
	PVC WB	−0.97 ± 1.46	−0.64 ± 1.44	0.417
	PVC TGM	−1.33 [−6.02–1.32]	−0.76 [−3.67–1.46]	0.132
	PVC CGM	−1.30 [−6.37–1.31]	−0.78 [−3.91–1.44]	0.140
	PVC DGM	−1.82 ± 2.98	−1.80 ± 2.27	0.980

Data presented in mean ± SD or median [range]. 9HPT = 9-Hole Peg Test; CGM = Cortical Gray Matter; DGM = Deep Gray Matter; EDSS = Expanded Disability Status Scale; MS = Multiple Sclerosis; PVC = Percentage Volume Change; SD = Standard Deviation; SDMT = Symbol Digit Modalities Test; T25FWT = Timed 25-Foot Walk Test; TGM = Total Gray Matter; WB = Whole Brain. * $p < 0.01$. Significant p -values are shown in bold.

TABLE 5 Linear regression models with clinical scores as dependent variables and MRI measures as independent variables.

Brain volume measures	ΔEDSS	ΔT25FWT	Δ9HPT	ΔSDMT	ΔMSFC _{SDMT}
PVC WB	−0.03 ± 0.10	−0.05 ± 0.21	−0.04 ± 0.44	0.68 ± 0.70	0.05 ± 0.02*
PVC TGM	−0.07 ± 0.10	−0.04 ± 0.22	−0.49 ± 0.45	1.25 ± 0.71	0.07 ± 0.02**
PVC CGM	−0.06 ± 0.10	−0.04 ± 0.21	−0.48 ± 0.43	1.23 ± 0.69	0.07 ± 0.02**
PVC DGM	−0.003 ± 0.05	−0.01 ± 0.12	0.06 ± 0.26	0.15 ± 0.41	0.001 ± 0.01

Data is presented as estimates of regression coefficient (β) ± SE. 9HPT = 9-Hole Peg Test; CGM = Cortical Gray Matter; DGM = Deep Gray Matter; EDSS = Expanded Disability Status Scale; MSFC_{SDMT} = Multiple Sclerosis Functional Composite Score using SDMT; PVC = Percentage Volume Change; SDMT = Symbol Digit Modalities Test; SE = Standard Error; T25FWT = Timed 25-Foot Walk Test; TGM = Total Gray Matter; WB = Whole Brain. * $p < 0.05$; ** $p < 0.01$. Significant p -values are shown in bold.

The number of patients showed an unequal distribution amongst both groups, prompting our hypothesis whether another cut-off could have been more informative. When using a cut-off based on the observed mean WB PVC in the MS cohort, (i.e., below −0.16% PVC), there were 28 individuals with “pathological” and 34 with “physiological”

WB volume loss. A significant difference in baseline T25FWT score was observed between these two groups (Supplementary Table 5, $p = 0.034$). However, regression analyses once again failed to detect baseline variables that independently predict the probability of reaching “pathological” BVL (Supplementary Table 6).

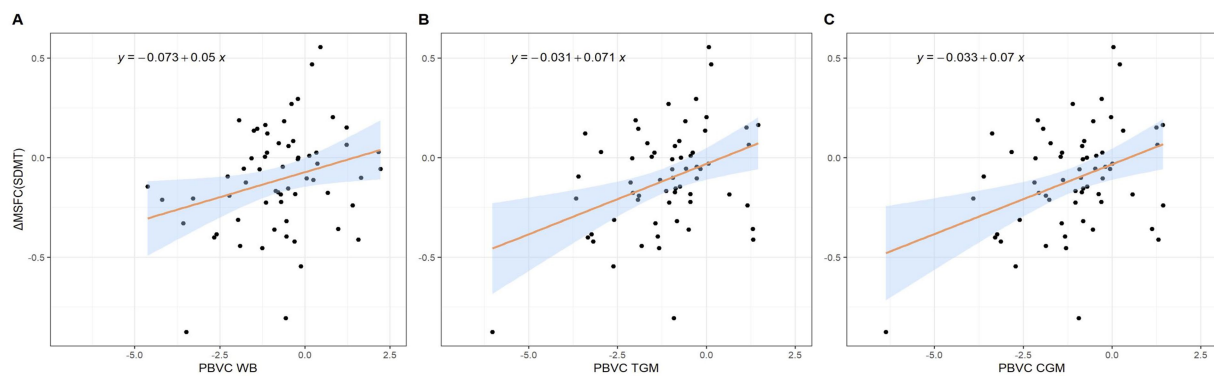


FIGURE 3

Scatterplots representing the relation between the change in MSFC_{SDMT} and (A) percentage whole brain volume change, (B) percentage total gray matter volume change and (C) percentage cortical gray matter volume change. CGM = Cortical Gray Matter; MSFC_{SDMT} = Multiple Sclerosis Functional Composite using SDMT; PVC = Percentage Volume Change; TGM = Total Gray Matter; WB = Whole Brain.

4 Discussion

BVL has recently gained attention as an MRI-derived proxy for neurodegeneration in patients with MS. The implementation of brain volumetry into routine clinical practice, though, has been hindered, mainly due to the risk of measurements becoming less precise once they are conducted outside strictly standardized research settings. In this longitudinal study, we aimed to validate the clinical relevance of BVL in a real-world cohort of patients with MS while explicitly taking into scanner differences using a similarity index as post-acquisition harmonization approach. We did not observe significant differences in annualized BVL between MS and HC essentially failing the primary objective of this study. However, worsening functional ability in patients with MS, as measured by 9HPT scores, was linked to increased atrophy in both TGM and CGM. Additionally, clinical decline on the composite MSFC_{SDMT} outcome measure was associated with WB, TGM, and CGM volume loss. We could not predict “pathological” BVL based on baseline demographics or clinical status.

It has been widely accepted that MS patients present with accelerated BVL compared to healthy individuals (9, 20, 21). The majority of studies on BVL, both in standardized and real-world settings, use the structural image evaluation with normalization of atrophy, or SIENA, method for PVC quantification, whereas the effect of change in MRI scanner strength (1.5 or 3.0 T) over time is usually taken into account with mixed-effect regression models (20–22). We have used the icobrain ms software to quantify BVL and applied a similarity index as post-acquisition harmonization method to consider scanner switches. We found PVC rates below the typical values of brain atrophy in MS (36), with only 15 of our 72 participants (21%) actually demonstrating annualized WB volume loss below 0.4%. However, this discrepancy cannot be due to the use of another tool for PVC quantification, as consistency between both methods for real-world BVL analysis has already been demonstrated (37). The negative outcome of our first objective (primary endpoint) does not seem to be influenced by a treatment effect, but may have been due to the unequal distribution of age between the MS and HC cohorts. Recent research suggest that age-related BVL accelerates significantly around 60 years of age (38). As shown in [Supplementary Figure 1](#), the greatest mean BVL was observed in the oldest age category

(55–70 years) for both HC and MS cohorts, where there is an important imbalance between the number of people with MS and HC. This imbalance may contribute to the apparent lack of difference in brain atrophy between the groups at the overall group level. Limited statistical power may have also contributed to our negative finding to some extent. Our sample sizes were sufficient to detect medium-to-large effects, but smaller differences in BVL could have gone undetected, increasing the risk of a type II error. Finally, survivor bias may have played a role: patients with more aggressive disease courses may have been less likely to remain in long-term follow-up, resulting in an overrepresentation of more stable individuals in the MS cohort and attenuating observed differences compared to HC. Nonetheless, some other studies have also reported low brain atrophy rates in MS (39, 40), further supporting the reliability of our BVL measures. Despite having a cohort with less global BVL than expected, these patients with MS still exhibited signs of disability worsening, which we were able to connect to GM volume loss.

MS was classically considered a disease of white matter (WM), but the intensity and relevance of GM involvement has become increasingly evident over the past two decades. GM volume loss, particularly in the deep nuclei, can occur early in the disease course and independent of focal WM lesions (41, 42). Moreover, several longitudinal studies have demonstrated that the volumetric decline of GM structures is not only more pronounced as compared to WM, but also more strongly associated with clinical outcomes (43, 44). Interestingly, the increased rate of WB volume loss observed in MS patients with progression independent of relapse activity, as compared to those that remained stable, could mainly be attributed to changes in the cerebral cortex (45). Our result are in line with previous findings, reporting significant associations between (C)GM atrophy and disease worsening, according to the 9HPT score and MSFC_{SDMT} (46, 47). BVL in GM can manifest according to different spatial patterns, which is relevant to specific clinical manifestations and even phenotypes (48–50). Regional GM atrophy may more accurately reflect the status of certain individual clinical measures than global estimates. For example, thalamic atrophy seems to be associated with cognitive impairment reflected by the SDMT in patients with MS, while overall GM atrophy does not (51). While we did not find statistically significant DMT effects on BVL, possibly due to a lack of

TABLE 6 Demographics of MS cohorts used for logistic regression analyses, using a cut-off of -0.4% annualized whole brain volume loss.

Characteristics	Physiological BVL ($\geq -0.4\%/year$)	Pathological BVL ($< -0.4\%/year$)
Number of subjects	50	12
Age (years)	44 $\pm$ 9	49 $\pm$ 12
Sex (F/M)	40/10	10/2
CVD (None/1/>2)	27/16/7	5/5/2
MS phenotype (RR/PMS)	44/6	9/3
DMT (None/First-line/Second-line)	11/27/12	5/5/2
DMT per agent		
None	11	5
Dimethyl fumarate	7	2
Glatiramer acetate	6	1
Interferon beta-1A	4	0
Interferon beta-1B	6	2
Teriflunomide	4	0
Fingolimod	4	0
Natalizumab	7	2
Ocrelizumab	1	0
Education level (Lower/Higher)	24/26	8/4
Interscan interval (months)	53 $\pm$ 5	54 $\pm$ 5
Disease duration (months)	150 [36–396]	192 [24–408]
Baseline EDSS	3.0 [1.0–6.5]	4.0 [1.5–6.5]
Baseline T25FWT (seconds)	5.3 [3.2–16.5]	6.0 [4.9–15.0]
Baseline 9HPT Dominant (seconds)	20.6 [14.0–38.3]	20.9 [16.2–37.1]
Baseline SDMT	52.2 $\pm$ 13.4	48.9 $\pm$ 14.6
Baseline MSFC _{SDMT}	0.4 $\pm$ 0.7	0.2 $\pm$ 0.7

Data presented in mean $\pm$ SD or median [range]. 9HPT = 9-Hole Peg Test; BVL = Brain volume loss; CVD = Cardiovascular comorbidities; DMT = Disease-modifying Treatment; EDSS = Expanded Disability Status Scale; F = Female; M = Male; MSFC_{SDMT} = Multiple Sclerosis Functional Composite Score using SDMT; RR = Relapsing–Remitting; PMS = Progressive Multiple Sclerosis; RR = Relapsing–Remitting; SD = Standard Deviation; SDMT = Symbol Digit Modalities Test; T25FWT = Timed 25-Foot Walk Test.

statistical power, others have recently reported an increase of thalamic volume with natalizumab, suggesting a potential neuroprotective role.⁵² So, even though we found an association between (C)GM atrophy and overall disease progression, more locally defined anatomic regions, such as the thalamus, may even be more informative about certain clinical as well as therapeutic aspects of MS. Unfortunately, parcellation of the CGM and segmentation of the DGM is not part of the default icobrain ms pipeline (and thus not validated for use in data from clinical routine) but may still be an interesting objective for future studies. Next to the importance of spatial patterns in brain atrophy, an important temporal factor might

TABLE 7 Logistic regression models for annualized whole brain volume loss with a cut-off of -0.40% per year as dependent variables and baseline clinical and demographic measures as independent variables.

Baseline measures	Pathological BVL ($< -0.4\%$ per year)	
EDSS	0.59 $\pm$ 0.43	
T25FWT	-0.21 ± 0.24	
9HPT	-0.15 ± 0.23	
SDMT	0.09 $\pm$ 0.14	
MSFC	-2.38 ± 4.03	
Age	0.03 $\pm$ 0.05	
Sex (Male)	-0.81 ± 1.10	
Disease duration	-0.001 ± 0.06	
MS phenotype (RR MS)	0.20 $\pm$ 1.15	
Education level (Higher)	-1.00 ± 0.88	
Comorbidities (One/Two or more)	0.26 $\pm$ 0.97	0.17 $\pm$ 1.13
DMT (First-line/Second-line)	-0.65 ± 0.92	-1.77 ± 1.32

Data is presented as estimates of regression coefficient (β) $\pm$ SE. 9HPT = 9-Hole Peg Test; BVL = Brain volume loss; DMT = Disease-modifying Treatment; EDSS = Expanded Disability Status Scale; F = Female; M = Male; MSFC_{SDMT} = Multiple Sclerosis Functional Composite Score using SDMT; RR = Relapsing–Remitting; SDMT = Symbol Digit Modalities Test; SE = Standard Error; T25FWT = Timed 25-Foot Walk Test. First-line DMT includes dimethyl fumarate ($n = 9$), glatiramer acetate ($n = 8$), interferon beta-1A ($n = 4$), interferon beta-1B ($n = 10$) and teriflunomide ($n = 5$). Second-line treatment includes fingolimod ($n = 5$), natalizumab ($n = 10$) and ocrelizumab ($n = 1$).

be at play when considering the intricate nature of the relationship between WM and GM pathology. A recent systematic review concluded that global GM atrophy appears to be secondary to focal WM lesions in early stages of the disease, and only later in the disease course will adopt a more independent character due to other neurodegenerative processes that are still not fully understood (52). This may correspond to the gradual change in inflammatory pathways throughout the disease course, with a predominant role of invading T lymphocytes in the formation of WM lesions in relapsing MS (1), and meningeal inflammation, amongst others, likely acting as a driving force for cortical atrophy in the progressive phase, which seemingly develops according to a clear gradient of neuronal loss that turns inwards from the boundary between GM and cerebrospinal fluid (53, 54). Such shift may even happen very early in the disease, as CIS patients that progress toward RR MS already show a 3.4 fold increase in GM atrophy, but no change in WM when compared to HC (46).

We acknowledge several limitations in the present study. First, despite our efforts to minimize the potential confounding impact of acute inflammatory activity or edema on BVL quantification, a pseudo-atrophy effect may not be limited to the presence of active lesions only. Recent literature has shown (i) that up to 25% of acute clinical events identified as relapse do not appear to be associated with lesional changes on MRI (55), (ii) contrast-enhancement within acute focal lesions typically lasts for approximately 4 weeks, whereas the pseudo-atrophy effect seems to reach a plateau only after 16 to 20 weeks (56, 57), (iii) inflammation in MS is not restricted to focal lesions and may occur diffusely throughout the normal-appearing white matter as well (58), and (iv) the exact pathophysiological mechanism behind pseudoatrophy is still not fully understood and may involve other processes (besides accelerated water

loss/fluid shifts), such as changes in glial cells (15, 56). Second, during our selection process, we experienced important loss of participants that drastically reduced the number of patients in our final sample. The subgroup with pathological BVL was particularly small ($n = 12$), which may have further limited our statistical power for the final objective. Previous studies have reported drop-out rates ranging from 37 to 82% (20–22, 59), and notably, one of these indicated a higher failure rate when focusing exclusively on 3D (82%) versus 2D (56%) imaging (59). Our drop-out rate of 56% aligns with the existing literature and we included 3D acquisition only. Notably, data loss in our study more frequently stemmed from applying the similarity index with a cut-off of 0.15 (33%), than from failing to meet key technical standards (16%). Even though we decided to continue with a lower similarity index cut-off (0.15 rather than 0.20) to retain as much patients as possible, we still experienced significant data loss due to this measure. Third, the similarity index does not account for various biological factors that can influence longitudinal volume changes, such as timing of scans, hydration state, lifestyle factors, and comorbidities. Nonetheless, a recent study suggests that technical causes contribute more significantly to variations in brain volume measures than physiological factors (60), reinforcing the value of using a similarity index as a post-acquisition harmonization approach in our analysis. Fourth, we used retrospective data for the MS cohort and relied on clinical data from two timepoints only (baseline and follow-up), limiting our ability to track confirmed disability worsening. On the other hand, we did employ a range of disability measures. This comprehensive approach allows us to capture various dimensions of disability, which is crucial in a heterogeneous condition like MS. Fifth, we used the SDMT for calculating MSFC scores, whereas this metric normally involves the Paced Auditory Serial Addition Test (PASAT) (61). This does not allow us to accurately compare our findings with others from the field. Still, both scores are representatives of cognition (particularly information processing speed) and using the SDMT score instead of the PASAT may even be an improvement due to its slightly better predictive validity, particularly when considering longitudinal data (62–64). Finally, we have only included the total EDSS score without its individual functional systems (i.e., the visual, brainstem, pyramidal, cerebellar, sensory, bowel and bladder, and cerebral functions) (28). In this real-world study, collecting such detailed information during routine medical visits is often challenging due to time constraints. To overcome this limitation, we have tried to take a multifaceted approach, by incorporating various measures of disability in MS.

5 Conclusion

Using the automated icobrain ms algorithm with similarity index-based post-acquisition harmonization to address scanner variability, we did not detect accelerated BVL in this real-world MS cohort compared to healthy individuals. Nonetheless, GM volume loss remains clinically relevant in MS, as it was associated with disability worsening according to the 9HPT and the MSFC_{SDMT}. To enable routine clinical use of BVL measurements, future research should prioritize robust post-acquisition correction methods that explicitly account for scanner differences. In the meantime, a practical recommendation for clinical settings is to perform follow-up scans on the same scanner whenever feasible, to enhance the reliability of longitudinal BVL assessments.

Data availability statement

Anonymized data supporting the findings of our study will be shared upon reasonable request from a qualified investigator and after approval by the ethics committees of NMSC Melsbroek and UZ Brussel.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the NMSC Melsbroek on June 1st 2021 (institutional authorization number: OG 033; internal reference number: EC21/06). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because According to the Belgian law, retrospective studies do not require participant consent.

Author contributions

JT: Writing – original draft, Investigation, Visualization, Data curation, Formal analysis, Validation, Project administration, Methodology, Writing – review & editing. AV: Data curation, Methodology, Conceptualization, Writing – review & editing. MB: Writing – review & editing, Supervision. MD'H: Conceptualization, Writing – review & editing, Supervision, Methodology. GN: Methodology, Data curation, Writing – review & editing, Formal analysis. AP: Writing – review & editing, Conceptualization, Investigation, Methodology. DSi: Software, Writing – review & editing, Methodology, Formal analysis, Data curation. DS_m: Data curation, Software, Writing – review & editing, Methodology. JS: Writing – review & editing, Formal analysis, Data curation, Methodology. MW: Writing – review & editing, Formal analysis, Methodology, Data curation. SE: Supervision, Writing – review & editing. MD'h: Investigation, Supervision, Writing – original draft, Writing – review & editing, Conceptualization.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

GN, MD'H, and MD'h have received consultancy fees, research support and/or travel grants from Merck Serono, Biogen, Sanofi Genzyme, Teva, Janssen-Cilag, Roche and/or Novartis. GN and DS_m (Dirk Smeets) are shareholders of icometrix and MMJW was a part-time employee of icometrix. SE has received consultancy fees and/or research support from ADx Neurosciences, Biogen, Danone, Eisai, icometrix, Janssen Pharmaceutica, Novartis, Nutricia and Roche.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Dobson R, Giovannoni G. Multiple sclerosis - a review. *Eur J Neurol.* (2019) 26:27–40. doi: 10.1111/ENE.13819
- Hauser SL, Cree BAC. Treatment of multiple sclerosis: a review. *Am J Med.* (2020) 133:1380. doi: 10.1016/J.AMJMED.2020.05.049
- Hellwig K, Lacobaeus E, Jensen MB. Smouldering multiple sclerosis: the 'real MS'. *Ther Adv Neurol Disord.* (2022);15. doi: 10.1177/17562864211066751
- Krieger S, Cook K, Hersh CM. Understanding multiple sclerosis as a disease spectrum: above and below the clinical threshold. *Curr Opin Neurol.* (2024) 37:189–201. doi: 10.1097/WCO.0000000000001262
- Scott LJ. Siponimod: a review in secondary progressive multiple sclerosis. *CNS Drugs.* (2020) 34:1191–200. doi: 10.1007/S40263-020-00771-Z/TABLES/2
- Portaccio E. Group on behalf of the IMSRC, Bellinva a, et al. progression is independent of relapse activity in early multiple sclerosis: a real-life cohort study. *Brain.* (2022) 145:2796–805. doi: 10.1093/BRAIN/AWAC111
- Hemond CC, Bakshi R. Magnetic resonance imaging in multiple sclerosis. *Cold Spring Harb Perspect Med.* (2018) 8:a028969. doi: 10.1101/cshperspect.a028969
- Sastre-Garriga J, Pareto D, Rovira À. Brain atrophy in multiple sclerosis: clinical relevance and technical aspects. *Neuroimaging Clin N Am.* (2017) 27:289–300. doi: 10.1016/J.NIC.2017.01.002
- Sokratous M, Siokas E, Tsouris Z. Brain atrophy in multiple sclerosis: mechanisms, clinical relevance and treatment options. *Autoimmunity Highlights* 2019 10:1. (2019) 10:1–25. doi: 10.1186/S13317-019-0117-5
- Voorn P, Twisk JWR, Barkhoff F. Postmortem validation of MRI cortical volume measurements in MS. *Hum Brain Mapp.* (2016) 37:2223–33. doi: 10.1002/HBM.23168
- Ringgaard S, Jensen HB, Nielsen HH. Efficacy of high-intensity aerobic exercise on brain MRI measures in multiple sclerosis. *Neurology.* (2021) 96:e203–13. doi: 10.1212/WNL.00000000000011241
- Tansy AP, Li T, Krieger S. A randomized, placebo-controlled, phase 2 trial of laquinimod in primary progressive multiple sclerosis. *Neurology.* (2020) 95:E1027–40. doi: 10.1212/WNL.00000000000010284
- Parker RA, Plantone D, Doshi A. Efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis (MS-SMART): a phase 2b, multiarm, double-blind, randomised placebo-controlled trial. *Lancet Neurol.* (2020) 19:214–25. doi: 10.1016/S1474-4422(19)30485-5
- Plantone D, Doshi A, John N. Multiple sclerosis-secondary progressive multi-arm randomisation trial (MS-SMART): a multiarm phase IIb randomised, double-blind, placebo-controlled clinical trial comparing the efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis. *BMJ Open.* (2018) 8:e021944. doi: 10.1136/BMJOPEN-2018-021944
- Minagar A, Stüve O, Lassmann H. Mechanisms of action of disease-modifying agents and brain volume changes in multiple sclerosis. *Neurology.* (2008) 71:136–44. doi: 10.1212/01.WNL.0000316810.01120.05
- Sormani MP, Arnold DL, De Stefano N. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Ann Neurol.* (2014) 75:43–9. doi: 10.1002/ANA.24018
- Bartolozzi ML, Battaglini M, Baldini M. Establishing pathological cut-offs of brain atrophy rates in multiple sclerosis. *J Neurol Neurosurg Psychiatry.* (2016) 87:93–9. doi: 10.1136/jnnp-2014-309903
- Banwell B, Stefano NDe, Enzinger C. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. *Lancet Neurol.* (2021) 20:653–70. doi: 10.1016/S1474-4422(21)00095-8
- Rocca MA, Ciccarelli O, Enzinger C. MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. *Nat Rev Neurol.* (2020) 16:171–82. doi: 10.1038/S41582-020-0314-X
- Hagemer J, Jakimovski D, Paunkoski I. Brain atrophy is associated with disability progression in patients with MS followed in a clinical routine. *Am J Neuroradiol.* (2018) 39:2237–42. doi: 10.3174/ajnr.A5876
- Hagemer J, Jakimovski D, Ramasamy DP. Disability improvement is associated with less brain atrophy development in multiple sclerosis. *Am J Neuroradiol.* (2020) 41:1577–83. doi: 10.3174/ajnr.A6684
- Ramasamy DP, Hagemer J, Genovese AV. Clinical feasibility of longitudinal lateral ventricular volume measurements on T2-FLAIR across MRI scanner changes. *Neuroimage Clin.* (2021) 29:102554. doi: 10.1016/J.NICL.2020.102554
- Cambron M, Horakova D, Jain S. Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain Behav.* (2016) 6:1–12. doi: 10.1002/brb3.518
- Carroll WM, Coetzee T, Comi G. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* (2018) 17:162–73. doi: 10.1016/S1474-4422(17)30470-2
- Finkelsztejn A, Fragoso YD, Bastos EA, Duarte JA, Varela JS, Houbrechts R, et al. Intercontinental validation of brain volume measurements using MSmetrix. *Neuroradiol J.* (2018) 31:147–9. doi: 10.1177/1971400917721521
- Sima DM, Horáková D, Nguyen AL, Van Hecke W, Kalincik T, Barnett MH, et al. Assessing the reliability of longitudinal MRI examinations in multiple sclerosis follow-up. *Mult Scler.* (2019) 25:547.
- Havrdova EH, Barnett MH, Stefano NDe. Utility of icobrain for brain volumetry in multiple sclerosis clinical practice. *Mult Scler Relat Disord.* (2024) 92:106148. doi: 10.1016/j.msard.2024.106148
- Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology.* (1983) 33:1444–52. doi: 10.1212/WNL.33.11.1444
- Phillips G, LaRocca N, Hudson LD. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler.* (2017) 23:704–10. doi: 10.1177/1352458517690823
- Benedict R, Phillips G, LaRocca N. The nine-hole peg test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler.* (2017) 23:711–20. doi: 10.1177/1352458517690824
- Benedict RHB, Deluca J, Phillips G, LaRocca N, Hudson LD, Rudick R. Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler.* (2017) 23:721–33. doi: 10.1177/1352458517690821
- Fischer JS, Rudick RA, Cutter GR, Reingold SC. The multiple sclerosis functional composite measure (MSFC): an integrated approach to MS clinical outcome assessment. *Mult Scler J.* (1999) 5:244–50. doi: 10.1177/135245859900500409
- Guldorf K, Nagels G, Smeets D. Brain volume loss can occur at the rate of Normal aging in patients with multiple sclerosis who are free from disease activity. *J Clinical Med.* (2022) 11:523. doi: 10.3390/JCM111030523
- Eichau-Madueño S, Ferrer-Lozano J, Forero L, et al. Deciphering multiple sclerosis progression. *Front Neurol.* (2021) 12:608491. doi: 10.3389/FNEUR.2021.608491/BIBTEX
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* (2008) 61:344–9. doi: 10.1016/j.jclinepi.2007.11.008

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2025.1637835/full#supplementary-material>

36. Mattle HP, O'Riordan J, Oreja-Guevara C. Clinical relevance of brain volume measures in multiple sclerosis. *CNS Drugs*. (2014) 28:147–56. doi: 10.1007/S40263-014-0140-Z/METRICS
37. Beadnall HN, Wang C, Van Hecke W, Ribbens A, Billiet T, Barnett MH. Comparing longitudinal brain atrophy measurement techniques in a real-world multiple sclerosis clinical practice cohort: towards clinical integration? *Ther Adv Neurol Disord*. (2019) 12:1756286418823462. doi: 10.1177/1756286418823462
38. Buddenkotte T, Spies L, Gocke C. Higher effect sizes for the detection of accelerated brain volume loss and disability progression in multiple sclerosis using deep-learning. *Comput Biol Med*. (2024) 183:109289. doi: 10.1016/J.COMPBIOMED.2024.109289
39. Chu R, Kim G, Tauhid S, Khalid F, Healy BC, Bakshi R. Whole brain and deep gray matter atrophy detection over 5 years with 3T MRI in multiple sclerosis using a variety of automated segmentation pipelines. *PLoS One*. (2018) 13:e0206939. doi: 10.1371/JOURNAL.PONE.0206939
40. Seidl Z, Ribbens A, Hecke WV. Long-term effectiveness of natalizumab on MRI outcomes and no evidence of disease activity in relapsing-remitting multiple sclerosis patients treated in a Czech Republic real-world setting: a longitudinal, retrospective study. *Mult Scler Relat Disord*. (2020) 46:102543. doi: 10.1016/J.MSARD.2020.102543
41. Agosta F, Luca MDe, Bartolozzi ML. Evidence of early cortical atrophy in MS: relevance to white matter changes and disability. *Neurology*. (2003) 60:1157–62. doi: 10.1212/01.WNL.0000055926.69643.03
42. Arnold DL, Dwyer MG, Gauthier SA. Deep grey matter injury in multiple sclerosis: a NAIMS consensus statement. *Brain*. (2021) 144:1974. doi: 10.1093/BRAIN/AWAB132
43. Geurts JGG, Calabrese M, Fisher E, Rudick RA. Measurement and clinical effect of grey matter pathology in multiple sclerosis. *Lancet Neurol*. (2012) 11:1082–92. doi: 10.1016/S1474-4422(12)70230-2
44. Vaneckova M, Ramasamy DP, Tyblova M. A serial 10-year follow-up study of brain atrophy and disability progression in RRMS patients. *Mult Scler*. (2016) 22:1709–18. doi: 10.1177/1352458516629769
45. Benkert P, Todea R-A, Rahmzadeh R. Association of Brain Atrophy with Disease Progression Independent of relapse activity in patients with relapsing multiple sclerosis. *JAMA Neurol*. (2022) 79:682–92. doi: 10.1001/JAMANEUROL.2022.1025
46. Fisher E, Lee JC, Nakamura K, Rudick RA. Gray matter atrophy in multiple sclerosis: a longitudinal study. *Ann Neurol*. (2008) 64:255–65. doi: 10.1002/ana.21436
47. Papadopoulou A, Barro C, Chakravarty MM. Central nervous system atrophy predicts future dynamics of disability progression in a real-world multiple sclerosis cohort. *Eur J Neurol*. (2021) 28:4153–66. doi: 10.1111/ENE.15098
48. Uher T, Vaneckova M, Tyblova M. Gray matter atrophy patterns in multiple sclerosis: a 10-year source-based morphometry study. *Neuroimage Clin*. (2018) 17:444. doi: 10.1016/J.NICL.2017.11.002
49. Margoni M, Vizzino C, Mistri D. Correspondence among gray matter atrophy and atlas-based neurotransmitter maps is clinically relevant in multiple sclerosis. *Molecular Psychiatry*. (2023) 28:1770–82. doi: 10.1038/s41380-023-01943-1
50. Altmann DR, Tur C, Cardoso MJ. Deep gray matter volume loss drives disability worsening in multiple sclerosis. *Ann Neurol*. (2018) 83:210. doi: 10.1002/ANA.25145
51. Weinstock-Guttman B, Lorence L, Schoonheim MM. Thalamic atrophy and dysconnectivity are associated with cognitive impairment in a multi-center, clinical routine, real-world study of people with relapsing-remitting multiple sclerosis. *Neuroimage Clin*. (2024) 42:103609. doi: 10.1016/J.NICL.2024.103609
52. Lie JA, Weeda MM, Mattiesing RM, et al. Relationship between white matter lesions and Gray matter atrophy in multiple sclerosis. *Neurology*. (2022) 98:E1562–73. doi: 10.1212/WNL.0000000000200006/ASSET/1147B425-829E-40F9-B9FB-0B0FD B570E44/ASSETS/GRAPHIC/17TTU1.GIF
53. Carassiti D, Radotra B, Gentleman SM. Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. *Brain*. (2011) 134:2755–71. doi: 10.1093/BRAIN/AWR182
54. Roncaroli F, Nicholas R, Serafini B. A gradient of neuronal loss and meningeal inflammation in multiple sclerosis. *Ann Neurol*. (2010) 68:477–93. doi: 10.1002/ANA.22230
55. Kerbrat A, Page ELE, Bigaut K. Acute clinical events identified as relapses with stable magnetic resonance imaging in multiple sclerosis. *JAMA Neurol*. (2024) 81:814–23. doi: 10.1001/JAMANEUROL.2024.1961
56. Stromillo ML, Cortese R, Gasperini C. Dynamics of pseudo-atrophy in RRMS reveals predominant gray matter compartmentalization. *Ann Clin Transl Neurol*. (2021) 8:623–30. doi: 10.1002/ACN3.51302
57. Fisher E, Rudick RA, Kooijmans M. Rate of brain atrophy in relapsing MS decreases during treatment with IFN β -1a. *Neurology*. (2005) 64:236–40. doi: 10.1212/01.WNL.0000149516.30155.B8
58. Zeis T, Graumann U, Reynolds R, Schaeren-Wiemers N. Normal-appearing white matter in multiple sclerosis is in a subtle balance between inflammation and neuroprotection. *Brain*. (2008) 131:288–303. doi: 10.1093/BRAIN/AWM291
59. Zivadinov R, Bergsland N, Korn JR, Dwyer MG, Khan N, Medin J. Feasibility of brain atrophy measurement in clinical routine without prior standardization of the MRI protocol: Results from MS-MRIUS, a longitudinal observational, Multicenter real-world outcome study in patients with relapsing-remitting MS. *AJNR Am J Neuroradiol*. (2018) 39:289–95. doi: 10.3174/AJNR.A5442
60. Maranzano J, Caramanos Z, Giacomini PS. Brain volume loss in individuals over time: source of variance and limits of detectability. *NeuroImage*. (2020) 214:116737. doi: 10.1016/J.NEUROIMAGE.2020.116737
61. Gronwall DMA. Paced auditory serial addition task: a measure of recovery from concussion. *Percept Mot Skills*. (1977) 44:367–73. doi: 10.2466/PMS.1977.44.2.367;PAGE: STRING:ARTICLE/CHAPTER
62. Drake AS, Weinstock-Guttman B, Morrow SA, Hojnacki D, Munschauer FE. Psychometrics and normative data for the multiple sclerosis functional composite: replacing the PASAT with the symbol digit modalities test. *Mult Scler*. (2010) 16:228–37. doi: 10.1177/1352458509354552
63. López-Góngora M, Querol L, Escartín A. A one-year follow-up study of the symbol digit modalities test (SDMT) and the paced auditory serial addition test (PASAT) in relapsing-remitting multiple sclerosis: an appraisal of comparative longitudinal sensitivity. *BMC Neurol*. (2015) 15:1–8. doi: 10.1186/S12883-015-0296-2/FIGURES/4
64. Sonder JM, Burggraaf J, Knol DL, Polman CH, Uitdehaag BMJ. Comparing long-term results of PASAT and SDMT scores in relation to neuropsychological testing in multiple sclerosis. *Mult Scler*. (2014) 20:481–8. doi: 10.1177/1352458513501570
65. Rekik A, Aissi M, Rekik I, Mhiri M, Frih MA. Brain atrophy patterns in multiple sclerosis patients treated with natalizumab and its clinical correlates. *Brain Behav*. (2022) 12:e2573. doi: 10.1002/BRB3.2573;PAGEGROUP:STRING:PUBLICATION