NeuroImage: Clinical 42 (2024) 103609

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Thalamic atrophy and dysconnectivity are associated with cognitive impairment in a multi-center, clinical routine, real-word study of people with relapsing-remitting multiple sclerosis

Robert Zivadinov [a,b,*], Niels Bergsland [a], Dejan Jakimovski [a], Bianca Weinstock-Guttman [c], Lorena Lorefice [d], Menno M. Schoonheim [e], Sarah A. Morrow [f,g], Mary Ann Picone [h], Gabriel Pardo [i], Myassar Zarif [j], Mark Gudesblatt [j], Jacqueline A. Nicholas [k], Andrew Smith [k], Samuel Hunter [l], Stephen Newman [m], Mahmoud A. AbdelRazek [n,o], Ina Hoti [n], Jon Riolo [p], Diego Silva [p], Tom A. Fuchs [e], Michael G. Dwyer [a,b], Ralph HB. Benedict [c]

[a] *Buffalo Neuroimaging Analysis Center, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, NY, United States*
[b] *Center for Biomedical Imaging at Clinical and Translational Science Institute, University of Buffalo, State University of New York, NY, United States*
[c] *Jacobs Multiple Sclerosis Center, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York and Kaleida Health, BGH, Buffalo, NY, United States*
[d] *Department of Medical Sciences and Public Health, Multiple Sclerosis Center, Binaghi Hospital, ASL Cagliari, University of Cagliari, Cagliari, Italy*
[e] *MS Center Amsterdam, Anatomy & Neurosciences, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam UMC location VUmc, Amsterdam, the Netherlands*
[f] *Schulich School of Medicine and Dentistry, London Health Sciences Centre, University Hospital, London, Ontario, CA, Canada*
[g] *Department of Clinical Neurological Sciences, Hotchkiss Brain Institute, University of Calgary, Canada*
[h] *Holy Name Medical Center, Teaneck, NJ, United States*
[i] *Oklahoma Medical Research Foundation, Oklahoma City, OK, United States*
[j] *South Shore Neurologic Associates NYU Langone, Patchogue, NY, United States*
[k] *OhioHealth MS Center, Riverside Methodist Hospital, Columbus, OH, United States*
[l] *Advanced Neurosciences Institute, Franklin, TN, United States*
[m] *Island Neurological Association, Plainview, NY, United States*
[n] *Mount Auburn Hospital, Harvard Medical School, United States*
[o] *Atrium Health Neurosciences Institute, Wake Forest University School of Medicine, United States*
[p] *Bristol Myers Squibb, Summit, NJ, United States*

A B S T R A C T

Background: Prior research has established a link between thalamic pathology and cognitive impairment (CI) in people with multiple sclerosis (pwMS). However, the translation of these findings to pwMS in everyday clinical settings has been insufficient.
Objective: To assess which global and/or thalamic imaging biomarkers can be used to identify pwMS at risk for CI and cognitive worsening (CW) in a real-world setting.
Methods: This was an international, multi-center (11 centers), longitudinal, retrospective, real-word study of people with relapsing-remitting MS (pwRRMS). Brain MRI exams acquired at baseline and follow-up were collected. Cognitive status was evaluated using the Symbol Digit Modalities Test (SDMT). Thalamic volume (TV) measurement was performed on T2-FLAIR, as well as on T1-WI, when available. Thalamic dysconnectivity, T2-lesion volume (T2-LV), and volumes of gray matter (GM), whole brain (WB) and lateral ventricles (LVV) were also assessed.
Results: 332 pwMS were followed for an average of 2.8 years. At baseline, T2-LV, LVV, TV and thalamic dysconnectivity on T2-FLAIR ($p < 0.016$), and WB, GM and TV volumes on T1-WI ($p < 0.039$) were significantly worse in 90 (27.1 %) CI vs. 242 (62.9 %) non-CI pwRRMS. Greater SDMT decline over the follow-up was associated with lower baseline TV on T2-FLAIR (standardized $\beta = 0.203$, $p = 0.002$) and greater thalamic dysconnectivity (standardized $\beta = -0.14$, $p = 0.028$) in a linear regression model.
Conclusions: PwRRMS with thalamic atrophy and worse thalamic dysconnectivity present more frequently with CI and experience greater CW over mid-term follow-up in a real-world setting.

* Corresponding author at: Buffalo Neuroimaging Analysis Center, Center for Biomedical Imaging at the Clinical Translational Science Institute, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, 77 Goodell Street, Suite 450, Buffalo, NY 14203, United States.
E-mail address: rzivadinov@bnac.net (R. Zivadinov).

https://doi.org/10.1016/j.nicl.2024.103609
Received 4 January 2024; Received in revised form 29 February 2024; Accepted 22 April 2024
Available online 27 April 2024

1. Introduction

The thalami are important structures located on both sides of the third ventricle in the brain (Minagar et al., 2013). They play a vital role in various neurological functions, such as motor control, sensory perception and integration (Minagar et al., 2013). The thalamus' proximity to cerebrospinal fluid, its unique neurological functions, and its central position as a relay for incoming and outgoing neural signals of both cortical and subcortical origin, make it particularly susceptible to the pathology of multiple sclerosis (MS) (Azevedo et al., 2015; Calabrese et al., 2011; Ontaneda et al., 2021; Rocca et al., 2010; Zivadinov et al., 2013b). As a result of MS, the myelin sheath encasing the axon undergoes progressive injury. This damage disrupts the normal flow of electrical signals in the brain, leading to the diverse and debilitating symptoms of MS (Minagar et al., 2013; Ontaneda et al., 2021). As many of these nerve fibers pass through or are connected to the thalamus, this region is a critical area to study and monitor in MS patients, making it a focus for ongoing research and clinical examination (Amin and Ontaneda, 2020; Azevedo et al., 2018; Eshaghi et al., 2018; Minagar et al., 2013; Ontaneda et al., 2021; Zivadinov et al., 2016a; Zivadinov et al., 2022).

Over recent years, the focus on measurement of thalamic atrophy has become an essential aspect of understanding and monitoring MS (Minagar et al., 2013; Ontaneda et al., 2021). This attention stems from several critical findings. For example, thalamic atrophy can be observed at the very onset of the disease (Azevedo et al., 2015; Calabrese et al., 2011; Eshaghi et al., 2018; Pareto et al., 2019; Rocca et al., 2010; Zivadinov et al., 2013b), including in initial clinical presentations and in young patients with MS (Aubert-Broche et al., 2014; Mesaros et al., 2008). Moreover, thalamic atrophy progresses regardless of MS subtype (Azevedo et al., 2018; Eshaghi et al., 2018; Ontaneda et al., 2021), with its rate surpassing that of overall brain, gray matter (GM), or white matter (WM), emphasizing the thalamus' susceptibility to injury in MS (Azevedo et al., 2018; Eshaghi et al., 2018; Minagar et al., 2013; Zivadinov et al., 2016a). Clinically, thalamic atrophy is associated with the likelihood of developing definitive MS (Calabrese et al., 2011; Zivadinov et al., 2013b) and with the trajectory of disability progression (Azevedo et al., 2018; Eshaghi et al., 2018; Zivadinov et al., 2013a; Zivadinov et al., 2016b). Consequently, it has been established as a pivotal endpoint in MS studies, aiding in the evaluation of disease evolution and therapeutic outcomes (Jakimovski et al., 2023).

Thalamic volume loss has been also linked to cognitive impairment (CI) in people with MS (pwMS) (Amin and Ontaneda, 2020; Bergsland et al., 2021; Bisecco et al., 2015; Houtchens et al., 2007; Schoonheim et al., 2022). Studies have shown that thalamic atrophy is associated with deficits in attention, information processing speed, and working memory in pwMS (Amin and Ontaneda, 2020) along with its severity being correlated with the degree of CI (Amin and Ontaneda, 2020; Benedict et al., 2013; Bergsland et al., 2015; Bergsland et al., 2021; Bisecco et al., 2015; Bisecco et al., 2021; Houtchens et al., 2007).

In pwMS, ongoing lesion development leads to a disconnection syndrome, severing or impairing many vital pathways that converge in the thalamus, thus hampering its ability to coordinate neural communication (Bisecco et al., 2015; Fuchs et al., 2019; Schoonheim et al., 2022). This disruption is believed to precipitate a cascade of structural and functional network alterations that parallel the clinical manifestations of the disease, particularly related to CI (Schoonheim et al., 2022). Measuring structural dysconnectivity typically requires the use of advanced diffusion-weighted imaging protocols, which are rarely a part of the routine clinical practice (Fuchs et al., 2019; Sjogard et al., 2021). However, measuring thalamic dysconnectivity using T2-FLAIR and spatial mapping against reference tractograms has been shown to be a promising alternative approach in understanding the relationship between the location of WM lesions (indicative of the disconnections within the thalamus neural networks) and CI in MS (Fuchs et al., 2018). While this approach was recently applied in the research studies (Fuchs et al., 2018; Fuchs et al., 2020), there is a gap in translating these findings into a real-world clinical practice.

Against this background, this study assessed which global and/or thalamic imaging biomarkers can be used to identify pwMS at risk for CI and cognitive worsening (CW) in a real-world setting, using a broad, heterogeneous group of people with relapsing-remitting MS (pwRRMS). This assessment was conducted through routine clinical imaging across multiple international centers over a mid-term period. We hypothesized that imaging biomarkers reflecting thalamic pathology (thalamic atrophy and dysconnectivity) would better predict CI and CW at the individual and group levels in pwRRMS, in real-word setting, compared to other inflammatory or neurodegenerative imaging outcomes. The rationale for this stems from an increasing body of evidence suggesting that thalamic pathology plays a central role in CI and CW in pwRRMS (Amin and Ontaneda, 2020; Bergsland et al., 2021; Debernard et al., 2015; Houtchens et al., 2007; Minagar et al., 2013; Schoonheim et al., 2012; Steckova et al., 2014; Till et al., 2011). We also hypothesized that measurement of thalamic atrophy on T2-FLAIR images (a common denominator sequence in the clinical routine) would be comparable to that obtained on T1-weghted imaging (WI) (not always obtained in the clinical routine), in terms of the association with cross-sectional and longitudinal cognitive outcomes.

2. Material and methods

2.1. Study design and population

The Cognitive DeepGRAI (Deep Gray Rating via Artificial Intelligence) study is an international, multi-center, longitudinal, observational, retrospective, real-word study of pwRRMS. Demographic and clinical data, along with brain MRIs were retrospectively collected from 11 MS research facilities located throughout the United States (US), Canada and Europe.

Approval for the study was granted by both central and local ethics committees, with the University at Buffalo (IRB STUDY00003141) overseeing the process. Owing to the study's retrospective, observational nature and de-identified collection, the requirement for obtaining informed consent from participants was waived.

2.2. Inclusion/exclusion criteria

The participation requirements for the study were as follows: (1) adult pwRRMS aged between 18 and 85 years, (2) the provision of a specific set of clinical data at the baseline and at follow-up, which included demographic and disease-specific requirements, (3) presence of baseline and follow-up MRI acquired after 12 to 60 months), (4) assessment of Symbol Digit Modalities Test (SDMT) (Benedict et al., 2017) and Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) at baseline and follow-up (both within a six-month window of the respective MRI scan dates), (4) the ability to centrally retrieve the original MRI scan data, and (6) the acquisition of T2-FLAIR. The study did not include individuals if they: (1) had any other neurological condition impacting the structure or function of the CNS besides MS, (2) were pregnant or breastfeeding during the course of the study, (3) were involved in any clinical trials with interventions during the study timeline, and (4) had experienced a relapse or had been treated with steroids within 30 days prior to the MRI scans.

2.3. Data collection

The study involved a retrospective collection of data from patient health records and brain MRI scans from MS centers using a designated electronic form for clinical research. The anonymized data was then consolidated into a centralized database. Details such as age, age at onset, sex, ethnicity, educational background, disease duration, disease course, SDMT score and EDSS evaluation were gathered at the initial

assessment and at the follow-ups. Additionally, information regarding types of disease-modifying treatments (DMT) being used was also documented.

MRI examinations were conducted using either 1.5 T or 3 T scanners without a prerequisite for standardization. Consistency in the type of scanner or magnetic field strength was not required for the patients, reflecting a more typical clinical environment. Both two-dimensional (2D) or three-dimensional (3D) T2-FLAIR images, along with 2D or 3D T1-WI, were collected.

All images were subjected to visual inspection. Quality metrics assessed included slice thickness, and overall quality indicators such as anatomical coverage, the presence of imaging artifacts, the extent of patient movement, noise levels, and image contrast, following previously established guidelines (Zivadinov et al., 2022). Additionally, any variations in the MRI hardware, scanner software, coils, and imaging protocols between the baseline and follow-up scans were documented and assessed, as previously reported (Zivadinov et al., 2022).

2.4. MRI analyses

The centralized coordinating center in Buffalo, NY conducted image analysis without knowledge of the patients' demographic details or clinical conditions.

T2 lesion volume (LV) was quantified using a semi-automated edge detection contouring/thresholding technique (Zivadinov et al., 2012). To enhance the reliability and accuracy of measurements, all regions of interest (ROIs) were drawn on images co-registered into a common imaging space (Jenkinson et al., 2002). For longitudinal lesion activity determination, existing ROIs from prior assessments were superimposed to assist in identifying new/enlarging T2 lesions.

All 2D and 3D T1-WI were lesion filled for all cross-sectional and longitudinal analyses to minimize the effect of T1 hypointensities (Gelineau-Morel et al., 2012). Baseline assessments of brain volume utilized FMRIB's Structural Image Evaluation, using Normalisation, of Atrophy Cross-sectional (SIENAX) software, both on 2D and 3D T1-WI to calculate normalized whole brain volume (WBV) and GM volume (GMV) (Smith et al., 2002). Longitudinal evaluation of brain volume changes involved calculating the percentage of brain volume change (PBVC) using SIENA (Smith et al., 2002), and the percentage GMV change (PGMVC) with SIENAX-multi time-point (SIENAX-MTP) (Dwyer et al., 2014), TV was measured with FIRST software on 2D and 3D T1-WI (Patenaude et al., 2011).

Baseline lateral ventricle volume (LVV) and its percentage change over time were calculated using Neurological Software Tool for Reliable Atrophy Measurement (NeuroSTREAM) based on T2-FLAIR images (Dwyer et al., 2017; Dwyer et al., 2018). TV measurements were also independently conducted on T2-FLAIR scans utilizing the DeepGRAI neural network model, as previously described (Dwyer et al., 2021; Zivadinov et al., 2022), Thalamic dysconnectivity, indicative of WM network disruption within thalamic pathways, was evaluated on T2-FLAIR images, as previously reported using the Network Modification (NeMo) tool (Fuchs et al., 2018; Fuchs et al., 2019; Fuchs et al., 2021). Briefly, NeMo measures WM tract disruption for tract bundles from all cortical and subcortical regions connected to the thalamus. This tool approximates WM tract disruption by comparing lesion masks with an internal database of 73 healthy adult tractograms (Kuceyeski et al., 2013). The proportion of disrupted WM tracts connecting to the thalamus was quantified.

2.5. Statistical analyses

The demographic, clinical, and MRI data were aggregated and standardized using R Statistical Software (v4.2.3; R Core Team 2021). All statistical analyses were executed using R in conjunction with SPSS software, version 26.0 (IBM, Armonk, NY, USA).

Data distribution and the spread of residuals were assessed through the examination of quantile–quantile (Q-Q) plots. Variables that followed a normal distribution were summarized using the mean and standard deviation (SD), while those that did not were summarized in terms of median and interquartile range (IQR).

CI at baseline was defined as a cutoff score of SDMT used to define impaired cognitive processing speed, corresponding to < 1.5 SD below healthy normative values (<44), as previously proposed (Beier et al., 2017). We used a 4- and 8-point SDMT score decline as a definition for cognitive worsening (CW) over the follow-up (Benedict et al., 2017). Disability progression (DP) was defined as an increase from baseline EDSS of at least 1.5 point, if the baseline EDSS was 0, 1.0 if the baseline EDSS was 1.0 to 5.5, or 0.5 if the baseline EDSS score was > 5.5 (Ghione et al., 2020). The pwRRMS with CWDP were defined as those who experienced both the CW and the DP over the follow-up.

Baseline differences between the CI and non-CI groups were assessed using chi-square test Student's *t*-test, the Mann-Whitney rank sum test, as well as ANCOVA models adjusting for baseline age, EDSS scores and time of follow-up. Binary logistic regression was used to predict the presence of CI at baseline by demographic (baseline age, baseline EDSS and disease duration) and MRI measures. In particular, the T2-FLAIR logistic step-wise regression model was built where CI status was the dependent variable, age, disease duration and baseline EDSS were included as covariates, and the 4 T2-FLAIR MRI measures (T2-LV, LVV, TV, thalamic dysconnectivity) were used as independent predictors. Similarly, the T1-WI model employed the same structure and utilized the T1-WI MRI measures (WBV, GMV and TV-FIRST) as independent predictors. Receiver operating characteristic (ROC) curve analyses were performed. The first ROC analysis determined the ability of baseline age, EDSS, TV, and thalamic dysconnectivity to discriminate between CI and non-CI pwRRMS. The area under curve (AUC), 95 % confidence intervals, and p-values were reported as appropriate. The diagnostic accuracy and the proposed cut-offs were described based on pre-selected levels of 80 % and 70 % sensitivity.

The relationships between baseline SDMT performance and absolute change over the follow-up and MRI measures were also assessed using Spearman's rank correlation test. Additional adjusted Spearman's correlations were performed. Linear regression analysis, adjusted for baseline age, disease duration and EDSS, explored which MRI measures were associated with greater SDMT decline over the follow-up.

Differences between those that experienced CW vs. those that did not were assessed using ANCOVA adjusted for age, age of education, baseline EDSS scores and disease duration. Additional ROC curve assessed the ability to discriminate pwRRMS with or without CW over the follow-up using the baseline thalamic atrophy. The diagnostic accuracy and the proposed cut-offs were described based on pre-selected levels of 80 % and 70 % sensitivity.

For all analyses, a p-value lower than 0.05 was considered statistically significant and a p-value lower than 0.1 as a trend.

3. Results

3.1. Study population

The Cognitive DeepGRAI study enrolled 332 pwRRMS from 11 MS centers in US (8), Canada (1) and Europe (2), according to the inclusion and exclusion criteria, with a median of 30 pwRRMS per center.

Table 1 shows demographic and clinical characteristics of the total study population, according to CI status. At baseline, 90 (27.1 %) pwRRMS were defined with CI. The mean time of follow-up was shorter in CI compared to non-CI pwRRMS (2.5 (1.5) vs 2.9 (1.5) years, $p = 0.024$). PwRRMS who had CI at baseline were significantly older (48.9 (12.0) vs 44.3 (10.4) years old, $p = 0.001$), had longer disease duration (16.8 (10.6) vs. 11.9 (9.1) years, $p < 0.001$) and greater EDSS (3.0 (2.0–4.0) vs. 2.0 (1.5–3.0), $p < 0.001$) compared to non-CI. At baseline, there was a trend for lower education in CI pwRRMS. Ten (3 %) pwRRMS transitioned to secondary-progressive MS (SPMS) over the

follow-up period. The annualized relapse rate over the follow-up was low (0.09). Seventy-two (21.7 %) pwRRMS experienced DP over the follow-up, with a similar proportion in the CI (22.2 %) and non-CI (21.5 %) groups. The absolute SDMT score decline was significantly higher in the non-CI compared to the CI group over the follow-up (−1.7 (7.7) vs 1.5 (10.9) points, p = 0.007). The demographic and clinical characteristics based on specific study center are visualized in Supplement Fig. 1.

According to 4-point SDMT decline, 121 (36.4 %) patients experienced CW over the follow-up, whereas the total was 57 (17.2 %) when the 8-point SDMT decline definition was used. PwRRMS who had CW over the follow-up had significantly greater EDSS (p = 0.001) at baseline.

3.2. MRI scanner and sequence characteristics

Table 2 displays MRI scanner hardware and sequence characteristics of the study cohort, at baseline and follow-up. Over the follow-up, 42 (12.7 %) of the pwRRMS were imaged on scanners that underwent hardware changes, while 119 (35.8 %) were imaged in the presence of hardware, software/coil or protocol changes. At baseline and follow-up, more pwRRMS were scanned at 3.0 T (>50 %) compared to 1.5 T.

As per the inclusion criteria, T2-FLAIR images were available for all patients at baseline and follow-up, passed quality control and were analyzed. Of 332 pwRRMS enrolled in the study, 292 (87.4 %) had T1-WI at baseline and 297 (89.5 %) at follow-up, with the majority having 3D T1-WI (78.4 % at baseline and 82.6 % at follow-up, respectively). All T1-WI images passed quality control and were analyzed.

3.3. Feasibility of MRI outcomes

Longitudinally, the highest analysis failure rate for T2-FLAIR was observed for LVV (29, 8.7 %), T2-LV (22, 6.6 %), while only 17 (5.1 %) pwRRMS had thalamic measures that failed quality control. The highest missing scan/analysis failure rate on T1-WI was observed for PBVC and PGMVC (125, 30.1 %) and TV (47, 14.2 %).

The cross-sectional and longitudinal MRI-based changes for each pwMS based on specific study center are visualized in Supplement Fig. 2 (cross-sectional) and Supplement Fig. 3 (longitudinal changes).

3.4. MRI outcomes at baseline and over the follow-up, according to cognitive impairment status at baseline

Table 3 shows T2-FLAIR and T1-WI MRI outcomes in pwRRMS, according to CI status at baseline.

On T2-FLAIR, pwRRMS with CI showed significantly higher T2-LV (13.2 (12.1) vs. 6.9 (8.9) mL, p < 0.001), thalamic dysconnectivity (3.6 (3.1) vs. 2.1 (2.2), p < 0.001), LVV (25.7 (14.7) vs. 19.3 (11.3) mL, p = 0.016), and lower TV (13.6 (2.1) vs. 14.5 (1.7) mL, p = 0.002) at baseline. All comparisons accounted for differences in age, time of follow-up and baseline EDSS scores. No significant longitudinal differences in MRI outcomes were observed between CI and non-CI groups over the follow-up on T2-FLAIR measures. The same significant findings between the CI and non-CI pwRRMS were further confirmed after controlling for sex and MRI center. In particular, MRI center had a significant effect on the comparison for baseline T2-LV (p = 0.018), baseline TV (p = 0.021), and baseline thalamic dysconnectivity (p = 0.001).

On T1-WI, pwRRMS with CI showed significantly lower TV (14.4 (2.6) vs. 15.9 (2.6), mL, p < 0.001), WBV (1455 (114.8) vs. 1510.5 (91.5) mL, p = 0.023) and GMV (540.2 (73.8) vs. 574.4 (68.2) mL, p = 0.039) at baseline. No significant longitudinal differences in MRI outcomes were observed between CI and non-CI groups over the follow-up on T1-WI measures. The MRI center had a significant effect on differences between the groups in longitudinal T1-WI TV % change (p = 0.005).

In a logistic regression model adjusted for age, baseline disease duration and baseline EDSS scores, and of all T2-FLAIR measures,

Table 1
Demographic and clinical characteristics of the study population.

Characteristics	CI pwRRMS (n = 90)	Non-CI pwRRMS (n = 242)	p-value
Sex, n (%)			0.338
Female	63 (70)	182 (75.2)	
Male	27 (30)	60 (24.8)	
Race, n (%)[1]			0.210
White	72 (80)	200 (82.6)	
Hispanic or Latino	1 (1.1)	2 (0.8)	
Black or African American	11 (12.2)	11 (4.5)	
Other	1 (1.1)	4 (1.6)	
Unknown	5 (5.6)	25 (10.3)	
Education at baseline in years[2], mean (SD)	14.3 (2.7)	14.9 (2.7)	0.063
Age at baseline in years, mean (SD)	48.9 (12)	44.3 (10.4)	**0.001**
Age at onset of the first clinical event in years, mean (SD)	33.1 (11.9)	34.3 (10.9)	0.398
Time of follow-up in years, mean (SD)	2.5 (1.5)	2.9 (1.5)	**0.024**
Disease duration at baseline in years, mean (SD)	16.8 (10.6)	11.9 (9.1)	**<0.001**
EDSS at baseline, median (IQR)	3.0 (2.0–4.0)	2.0 (1.5–3.0)	**<0.001**
EDSS at follow-up, median (IQR)	3.5 (2.0–4.0)	2.5 (1.5–3.5)	**<0.001**
EDSS absolute change over the follow-up, median (IQR)	0.0 (0.0–0.5)	0.0 (0.0–0.5)	0.641
SDMT at baseline, mean (SD)	35.3 (6.7)	56.2 (8.7)	**<0.001**
SDMT at follow-up, mean (SD)	36.7 (10.9)	54.5 (11.1)	**<0.001**
SDMT absolute change over the follow-up, mean (SD)	1.4 (9.7)	−1.7 (7.7)	**0.007**
Number of relapses in 12 months before baseline, median (IQR)	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.768
Number of relapses in 24 months before baseline, median (IQR)	0.0 (0.0–1.0)	0.0 (0.0–1.0)	0.500
Annualized relapse rate, mean (SD)	0.119 (0.32)	0.085 (0.29)	0.376
Number of relapses over follow-up, median (IQR)	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.499
Disability status at follow-up, n (%)			0.934
Improved	6 (6.6)	19 (7.9)	
Stable	64 (71.1)	171 (70.7)	
Progressed	20 (22.2)	52 (21.5)	
DMT at baseline, n (%)			0.160
Non-therapy	19 (21.1)	34 (14)	
Interferon-β	15 (16.7)	70 (29)	
Glatiramer acetate	18 (20)	38 (15.7)	
Oral DMTs	13 (14.4)	38 (15.7)	
Monoclonal antibodies	25 (27.8)	62 (25.6)	
Off-label therapy	–	–	
DMT at follow-up, n (%)			0.448
Non-therapy	17 (18.9)	34 (14)	
Interferon-β	13 (14.4)	50 (20.7)	
Glatiramer acetate	10 (11.1)	28 (11.6)	
Oral DMTs	25 (27.8)	62 (25.6)	
Monoclonal antibodies	25 (27.8)	65 (26.9)	
Off-label therapy	–	3 (1.2)	
DMT status at follow-up, n (%)			0.575
Remained on same DMT	45 (50)	144 (59.5)	
Started DMT	11 (12.2)	20 (8.3)	
Non-therapy	7 (7.8)	14 (5.8)	
Switched DMT	18 (20)	44 (18.2)	
Stopped DMT	9 (10)	20 (8.3)	

pwRRMS – people with relapsing-remitting multiple sclerosis; CI-cognitively impaired; n-number; EDSS-Expanded Disability Status Scale; DMT-disease modifying treatment; SDMT – Symbol Digit Modalities Test; SD – standard deviation, IQR – interquartile range.
Monoclonal antibody therapy includes natalizumab, rituximab, ocrelizumab and ofatumumab. Oral DMT group included cladribine, dimethyl fumarate, fingolimod, siponimod, ozanimod, and teriflunomide. [1]Other race group consisted of one Asian and 4 Native American pwRRMS. [2]Available for 296 pwRRMS.
The data were collected from 11 centers that included: 1) University at Buffalo, NY, USA (n = 56), 2) University of Cagliari, Italy (n = 43), 3) Vrije Universiteit Amsterdam, Netherlands (n = 30), 4) University Hospital, Ontario, Canada (n = 32), 5) Holy Name Medical Center, NJ (n = 25), 6) Oklahoma Medical Research Foundation, OK (n = 15), 7) South Shore Neurologic Associates NYU Langone, NY (n = 13), 8) OhioHealth MS Center, OH (n = 22), 9) Advanced Neurosciences Institute, TN (n = 6), 10) Kaleida Health BGH, NY (n = 55) and 11) Mount Auburn Hospital, Harvard Medical School, MA (n = 35).
CI at baseline was defined as a cutoff score of SDMT (<44) used to define impaired cognitive processing, corresponding to < 1.5 SD below healthy normative values, as previously proposed. (Beier et al., 2017).
P-values lower than 0.05 were considered statistically significant and shown in bold.

Table 2
Description of the MRI hardware and MRI sequences utilized in the total study population.

MRI field strength at baseline[a]		MRI field strength at follow-up[a]	
1.5 T	156 (47.0)	1.5 T	145 (43.7)
3.0 T	175 (52.7)	3.0 T	187 (56.3)
MRI field strength combinations[a]			
1.5 T-1.5 T	140 (42.2)	1.5 T-3 T	14 (4.2)
3 T-3 T	173 (52.1)	3 T-1.5 T	4 (1.2)
MRI scanner hardware changes		MRI scanner model, software and protocol changes	
Yes	42 (12.7)	Yes	119 (35.8)
No	290 (87.3)	No	213 (64.2)
MRI manufacturer at baseline		MRI manufacturer at follow-up	
General Electric	219 (66.0)	General Electric	219 (66.0)
Hitachi	13 (3.9)	Hitachi	13 (3.9)
Philips	8 (2.4)	Philips	8 (2.4)
Siemens	92 (27.7)	Siemens	91 (27.4)
Toshiba	–	Toshiba	1 (0.3)
FLAIR type at baseline		FLAIR type at follow-up	
2D	127 (38.3)	2D	98 (29.5)
3D	205 (61.7)	3D	234 (70.5)
Thickness	1.9 (1.1)	Thickness	1.8 (0.9)
T1 type at baseline		T1 type at follow-up	
Availability of T1	292 (87.4)	Availability of T1	297 (89.5)
2D	63 (21.6)	2D	52 (17.4)
3D	227 (78.4)	3D	244 (82.6)
Thickness	1.9 (1.4)	Thickness	1.7 (1.3)

pwRRMS – people with relapsing-remitting multiple sclerosis, MRI – magnetic resonance imaging, FLAIR – fluid-attenuated inversion recovery.
[a] At index, one pwRRMS was scanned with 1.0 T that later switched to 1.5 T scanner.

thalamic dysconnectivity was the only variable that provided additional explanatory variance in predicting CI status at baseline. In particular, the model predictivity improved by 7 % (R^2 change from 0.1 to 0.17) with thalamic dysconnectivity as the only significant predictor (Wald = 14.06p < 0.001).

In a similar T1-WI-based regression analysis, both baseline GMV and TV were significant predictors of the baseline CI status. In particular, GMV improved the model (R^2 change from 0.12 to 0.18, Wald = 5.3, p = 0.021) and the addition of TV yielded additional improvement (R^2 change from 0.18 to 0.20, Wald = 12.6, p < 0.001) in predictivity. In a joint regression model that utilized the significant predictors from the independent T2-FLAIR and T1-WI based models, thalamic dysconnectivity assessed on T2-FLAIR sequence and TV assessed on T1-WI were the significant predictors of CI status (Wald = 8.3, p = 0.004 and Wald = 5.6, p = 0.018).

Table 3
Differences in T2-FLAIR and T1-WI MRI measures between CI and non-CI pwRRMS at baseline.

T2-FLAIR measures	CI pwRRMS at baseline (n = 90)	Non-CI pwRRMS at baseline (n = 242)	p-value[a]	p-value[b]
T2-LV at baseline	13.2 (12.1)	6.9 (8.9)	**<0.001**	**<0.001**
Absolute T2-LV change	0.5 (1.9)	0.3 (1.6)	0.484	0.547
New/enlarging T2 lesion number	0.5 (1.5)	0.39 (1.2)	0.557	0.827
LVV at baseline	25.7 (14.7)	19.3 (11.3)	**0.016**	**0.021**
% LVV change	2.6 (10.6)	2.2 (23.9)	0.773	0.863
Thalamic volume at baseline	13.6 (2.1)	14.5 (1.7)	**0.002**	**0.001**
Thalamic volume % change	−1.3 (5.5)	−1.3 (4.4)	0.999	0.994
Thalamic dysconnectivity at baseline	3.6 (3.1)	2.1 (2.2)	**<0.001**	**<0.001**
Thalamic dysconnectivity absolute change	−0.06 (1.3)	0.12 (1.1)	0.355	0.333
T1-WI measures				
WBV at baseline	1455.0 (114.8)	1510.5 (91.5)	**0.023**	**0.029**
PBVC	−1.1 (1.5)	−1.2 (1.6)	0.451	0.415
GMV at baseline	540.2 (73.8)	574.4 (68.2)	**0.039**	0.053
PGMVC	0.19 (3.0)	−0.9 (2.8)	0.055	0.063
Thalamic volume at baseline	14.4 (2.6)	15.9 (2.6)	**<0.001**	**<0.001**
Thalamic volume % change	−1.0 (9.9)	−2.9 (7.9)	0.118	0.153

pwRRMS-people with relapsing-remitting multiple sclerosis; WI-weighted imaging; FLAIR – Fluid attenuated inversion recovery, CI – cognitively impaired, LVV – lateral ventricular volume, LV – lesion volume, WBV – whole brain volume, GMV – gray matter volume. PBVC – percent brain volume change, PGMVC – percent gray matter volume change, % – percent.
CI at baseline was defined as a cutoff score of SDMT (<44) used to define impaired cognitive processing, corresponding to < 1.5 SD below healthy normative values, as previously proposed. (Beier et al., 2017).
Volumes are shown in in milliliters. P-values lower than 0.05 were considered statistically significant and shown in bold.
[a] Analysis of covariance (ANCOVA) adjusted for age, time of follow-up, and baseline EDSS scores was used.
[b] Analysis of covariance (ANCOVA) adjusted for sex, time of follow-up, baseline EDSS scores and MRI center.

Table 4 and Figure (Panel A) shows ROC analysis for predicting CI vs non-CI status at baseline in pwRRMS at the group level. The AUC analysis determined the ability of baseline thalamic dysconnectivity, EDSS, TV and age (p < 0.001 for all) to discriminate between CI and non-CI pwRRMS. At the individual level, diagnostic accuracy to discriminate between CI and non-CI pwRRMS for thalamic dysconnectivity (cutoff of 0.96) at 80 % sensitivity had 38 % specificity, while for TV (cutoff of 12.2 mL) at 80 % sensitivity had 3.4 % specificity. At 70 % sensitivity the specificity was 62.4 % for thalamic dysconnectivity and 5.4 % for TV.

3.5. Relationship between MRI measures and SDMT performance at baseline and SDMT absolute change over the follow-up

Table 5 shows the Spearman's rank correlations between baseline

MRI measures and SDMT performance at baseline and absolute SDMT change over the follow-up. Lower SDMT at baseline was related to the baseline T2-FLAIR measures of higher T2-LV, LVV and thalamic dysconnectivity, and lower TV ($p < 0.001$ for all) and to the baseline T1-WI measures of lower WBV, GMV and TV ($p < 0.001$ for all). Over the follow-up, greater absolute SDMT decline was related to lower TV on T2-FLAIR ($p = 0.016$) and TV on T1-WI ($p = 0.043$) at baseline. In an additional age, sex, disease duration and baseline EDSS-adjusted Spearman's correlation analysis, the results remained the same. In particular, higher baseline SDMT performance was associated with lower T2-LV ($r = -0.179$, $p = 0.002$), higher T2-FLAIR TV ($r = 0.212$, $p < 0.001$), thalamic dysconnectivity ($r = -0.215$, $p < 0.001$) and T1-WI measures of higher WBV ($r = 0.146$, $p = 0.017$) and TV ($r = 0.217$, $p < 0.001$). After adjusting for all confounders, the absolute SDMT change was not associated with none of the T2-FLAIR and T1-WI MRI measures.

In a linear regression analysis adjusted by baseline age, disease duration and EDSS, greater SDMT decline over the follow-up was significantly predicted by both lower TV (standardized $\beta = 0.203$, $p = 0.002$) and greater thalamic dysconnectivity (standardized $\beta = -0.14$, $p = 0.028$) on T2-FLAIR at baseline.

3.6. MRI outcomes at baseline and over the follow-up, according to cognitive and disability worsening status over the follow-up

In the ROC analysis, lower TV at baseline was a significant predictor of CW (4-point decline) status over the follow-up (AUC 0.569, $p = 0.041$, 95 % CI 0.503 – 0.635 (Figure, Panel B). Using the same diagnostic measures, a baseline TV cut-off of 12.6 mL had 80 % sensitivity and 12.4 % specificity for discriminating between CW pwRRMS from cognitively stable/improving pwRRMS. Similarly, at 70 % sensitivity, baseline TV of 13.1 mL had 18.4 % specificity.

No longitudinal differences in MRI outcomes were observed between CW (4- and 8-point decline) and non-CW groups over the follow-up on T2-FLAIR- and T1-WI measures.

Supplement Table shows T2-FLAIR and T1-WI MRI outcomes in pwRRMS, according to CW (4-point decline) DP and CW-nonDP status at follow-up. On T2-FLAIR, there was significantly lower TV (13.5 (1.7) vs. 14.3 (2.0) mL, $p = 0.042$) at baseline, and a trend for higher TV loss (−2.6 % (3.5) vs. −1.3 % (2.6), $p = 0.093$) over the follow-up in CWDP group. On T1-WI, there was a trend for lower TV (14.0 (2.0) vs. 15.3 (2.8) mL, $p = 0.052$) at baseline, and significantly greater whole brain atrophy (−1.88 % (1.7) vs. −0.98 % (1.5), $p = 0.038$) over the follow-up in CWDP group.

4. Discussion

This international, multi-center study investigated the value of assessing global and thalamic imaging biomarkers (thalamic atrophy and dysconnectivity) in predicting CI at baseline, and CW over the follow-up, compared to other inflammatory and neurodegenerative imaging biomarkers, using a large cohort of pwRRMS followed via clinical routine in 11 centers. The study utilized retrospectively collected data from over 330 pwRRMS who were followed for an average of 2.8 years.

Table 4
Area under receiver operating curve (AUC) for predicting baseline CI vs. non-CI status in pwRRMS.

Test Result Variable(s)	AUC	Std. Error	p-value	95 % Confidence Intervals	
				Lower Bound	Upper Bound
Thalamic volume at baseline	0.644	0.039	**<0.001**	0.568	0.719
Baseline age	0.622	0.037	**0.001**	0.549	0.696
Baseline EDSS	0.650	0.036	**<0.001**	0.580	0.721
Thalamic dysconnectivity at baseline	0.685	0.037	**<0.001**	0.613	0.757

pwRRMS – people with relapsing-remitting multiple sclerosis, AUC – area under curve, EDSS – Expanded Disability Status Scale.

Assessing thalamic atrophy using routine clinical MRI is difficult due to technical limitations in acquisition and measurement methods. Current techniques are influenced by the stability and quality of image acquisition (Zivadinov et al., 2018; Zivadinov et al., 2022). While previous MRI-cognitive studies in MS targeted research-quality MRI (T1-WI), there is a need for accurate measurement of TV for predicting cognitive outcomes with clinical-quality scans. Since T2-FLAIR MRI is widely used in MS imaging, obtaining thalamic volumetry for it would be highly advantageous. Therefore, to increase the accuracy of the measurement and maximize number of analyses without prior standardization of the MRI exams, we decided to measure TV both on T2-FLAIR (using AI) and T1-WI using FIRST, as previously reported (Zivadinov et al., 2022). In line with the previous study, using the same DeepGRAI approach (Zivadinov et al., 2022), we found that it was feasible to measure TV on T2-FLAIR images in 95 % of pwRRMS, a higher feasibility rate than the 85.8 % found for FIRST TV determination on T1-WI, which made comparison of the two approaches more reliable than in the previous real-word study (Zivadinov et al., 2022). Moreover, compared to previous real-word studies (Barnett et al., 2021; Zivadinov et al., 2018; Zivadinov et al., 2022), in the present study, we found a lower rate of hardware/software/protocol changes over the follow-up, which could have increased reliability of the MRI analyses.

We showed that all explored measures on T2-FLAIR and T1-WI at baseline were worse in CI compared to non-CI pwRRMS, confirming previous research findings in the setting of a real-word study (Amin and Ontaneda, 2020; Bergsland et al., 2015; Bisecco et al., 2015; Dwyer et al., 2021; Houtchens et al., 2007; Minagar et al., 2013). However, in a joint logistic regression model that utilized the significant predictors from the independent T2-FLAIR-derived and T1-WI-based analyses, only the thalamic dysconnectivity and thalamic atrophy were the significant predictors of CI status at baseline, confirming previous research findings indicating that the assessment of thalamic pathology is more relevant for detecting CI in pwRRMS, when compared to other MRI measures, including T2-LV, LVV, WBV and GMV (Amin and Ontaneda, 2020; Bergsland et al., 2015; Bisecco et al., 2015; Dwyer et al., 2021; Houtchens et al., 2007; Minagar et al., 2013). These findings were also confirmed in the ROC analysis which determined that thalamic dysconnectivity, EDSS, thalamic atrophy and age were able to significantly discriminate between CI- and no-CI pwRRMS at baseline.

Both, the correlation, and linear regression analyses showed that pwRRMS who had more thalamic atrophy at baseline developed greater SDMT decline over the follow-up. In addition, in linear regression analysis it was shown that thalamic dysconnectivity was also related to greater SDMT decline over the follow-up. These findings support one of the objectives of the study, which was to investigate in a real-word environment, whether both the assessment of thalamic atrophy and dysconnectivity are complementary approaches to detect CI in pwRRMS. Therefore, integrating these biomarkers into standard clinical routine could enhance patient care by identifying pwRRMS at increased risk for CI.

In the ROC analysis, lower TV at baseline was a significant predictor of CW (4-point decline) status over the follow-up at the group level. However, we were unable to differentiate pwRRMS who developed 0CW (defined both by 4- and 8-point decline), over the follow-up by using any of the proposed longitudinal MRI outcomes in this study. This can be attributed to relatively short follow-up and reliability of SDMT (4- or 8-point based definitions of CW), which is a subject of debate in the literature (Castrogiovanni et al., 2023; DeLuca et al., 2021), and its interpretation is especially challenging if not confirmed on subsequent follow-up visits, as was the case in the current study. Second, we used a single cognitive test (SDMT) for determining CW, instead of using a

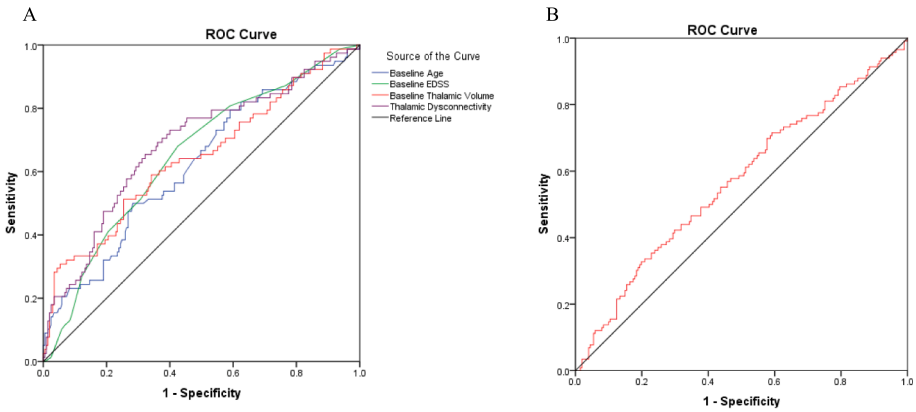

Fig. 1. Receiver operating characteristic (ROC) curves for predicting CI vs. non-CI status at baseline (A) and CW vs. non-CW (B) over follow-up in pwRRMS. pwRRMS – people with relapsing-remitting multiple sclerosis, CI-cognitive impairment; CW-cognitive worsening; AUC – area under curve, EDSS – Expanded Disability Status Scale.

Table 5
Relationship between baseline MRI measures and SDMT performance at baseline and over the follow-up in pwRRMS.

T2-FLAIR measures		SDMT at baseline	Absolute SDMT change	T1-WI based measures		SDMT at baseline	Absolute SDMT change
T2-LV	r-value	−0.326	−0.079	WBV	r-value	0.321	0.048
	p-value	**<0.001**	0.172		p-value	**<0.001**	0.417
LVV	r-value	−0.265	−0.083	GMV	r-value	0.263	0.007
	p-value	**<0.001**	0.154		p-value	**<0.001**	0.9
Thalamic volume	r-value	0.287	0.135	Thalamic volume	r-value	0.312	0.119
	p-value	**<0.001**	**0.016**		p-value	**<0.001**	**0.043**
Thalamic dysconnectivity	r-value	−0.327	−0.048				
	p-value	**<0.001**	0.409				

pwRRMS-people with relapsing-remitting multiple sclerosis; FLAIR – Fluid-attenuated inversion recovery, WI-weighted imaging; SDMT – Symbol Digit Modalities Test, LVV – lateral ventricular volume, LV – lesion volume, WBV – whole brain volume, GMV – gray matter volume.
The relationships were determined using non-parametric Spearman's rank correlation test.
P-values lower than 0.05 were considered statistically significant and shown in bold.

more comprehensive neurocognitive battery, as previously proposed (Benedict et al., 2012). Finally, most of the pwRRMS in this study were on highly effective DMTs and showed extremely low disease activity (as reflected by relapse rate and number of new/enlarging T2 lesions) over the follow-up.

Although this study confirmed the group-level relevance of thalamic pathology and the viability of its measurement on retrospective and prospective clinical datasets, it did not show that measures of thalamic pathology (or other imaging measures) had sufficient accuracy to predict cognitive status or trajectory (CI and CW) on an individual level. Thalamic dysconnectivity showed somewhat better predictivity than TV. While this is a disappointing result, we are not yet at the point where we can predict with clinically sufficient accuracy how MS will affect each individual patient's cognitive outcomes based solely on structural imaging. It is important to note that we did not assess functional MRI outcomes in this study, which provide insight into the neural activity patterns, and are also crucial for predicting CI and CW in pwMS, complementing the structural MRI findings that have traditionally been used (Lin et al., 2019; Ziccardi et al., 2023). While the progress made is promising and moving toward personalized prognosis and treatment, the variability in disease manifestation and response to treatment among pwMS means that there is still a gap between our current capabilities and the goal of individual prediction. This study builds on the progress made thus far, but acknowledging the complexity and unpredictability of the disease is crucial in setting realistic expectations for patients and healthcare providers.

In a previous real-word study using a similar study design and MRI outcomes, it was shown that thalamic atrophy assessed via AI was associated with DP in pwRRMS over mid-term (Zivadinov et al., 2022). In order to investigate whether pwRRMS with CWDP showed greater deterioration of MRI outcomes compared to CW non-DP pwRRMS, we performed additional analyses, and in line with previous findings (Zivadinov et al., 2022), we showed that thalamic atrophy both on T2-FLAIR and T1-WI at baseline was associated with CWDP. In addition, we found that greater WBV loss over the follow-up was also associated with CWDP.

The robustness of this study is notably reinforced by its large cohort, gathered across 11 international MS centers in a real-word setting, providing a substantial data pool for analysis. A noted limitation, however, was the relatively uniform population of pwRRMS, conforming to the study's initial selection criteria. Consequently, there is a clear avenue for future research to investigate thalamic pathology in patients at initial presentation as well as with progressive forms of MS within real-world clinical settings. The transition of only 10 pwRRMS to SPMS during the study period can be attributed to the relatively short duration of follow-up, underlining the need for longer-term studies to fully understand the progression of thalamic pathology over time.

The study design had a limitation in that the SDMT and EDSS were gathered within six months of the MRI scan, which could have increased the variability. Nonetheless, a shorter time frame would likely render

the study impracticable in everyday clinical practice.

This comprehensive study confirms the feasibility of using thalamic atrophy and dysconnectivity as MRI biomarkers for CI in pwRRMS. Despite a short follow-up and a homogenous sample, findings point to these measures' potential in predicting greater SDMT decline and CW, underscoring the need for their integration into routine clinical assessment.

5. Study funding

Study (RF AWARD NUMBER: 84309) was supported by a collaboration grant from Bristol Myers Squibb.

CRediT authorship contribution statement

Robert Zivadinov: Funding acquisition, Supervision, Writing – original draft. **Niels Bergsland:** Conceptualization, Formal analysis, Writing – review & editing. **Dejan Jakimovski:** Formal analysis, Methodology, Writing – review & editing. **Bianca Weinstock-Guttman:** Conceptualization, Writing – review & editing. **Lorena Lorefice:** Data curation, Writing – review & editing. **Menno M. Schoonheim:** Data curation. **Sarah A. Morrow:** Conceptualization. **Mary Ann Picone:** Conceptualization. **Gabriel Pardo:** Conceptualization. **Myassar Zarif:** Conceptualization. **Mark Gudesblatt:** Conceptualization, Writing – review & editing. **Jacqueline A. Nicholas:** Conceptualization, Writing – review & editing. **Andrew Smith:** Conceptualization, Writing – review & editing. **Samuel Hunter:** Conceptualization, Writing – review & editing. **Stephen Newman:** Conceptualization, Writing – review & editing. **Mahmoud A. AbdelRazek:** Conceptualization, Writing – review & editing. **Ina Hoti:** Writing – review & editing, Conceptualization. **Jon Riolo:** Writing – review & editing, Conceptualization. **Diego Silva:** Writing – review & editing, Conceptualization. **Tom A. Fuchs:** . **Michael G. Dwyer:** Methodology, Formal analysis. **Ralph HB. Benedict:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Robert Zivadinov has received personal compensation from Bristol Myers Squibb, EMD Serono, Sanofi, Janssen, Protembis, Filterlex and Novartis for speaking and consultant fees. He received financial support for research activities from Novartis, Bristol Myers Squibb, Octave, Mapi Pharma, Protembis, CorEvitas and V-WAVE Medical. Niels Bergsland and Dejan Jakimovski have nothing to disclose. Myassar Zarif, Samuel Hunter, Stephen Newman, Tom Fuchs and Mary Ann Picone have not declared any conflict of interest. Bianca Weinstock-Guttman received honoraria as a speaker and/or as a consultant for Biogen Idec, Sanofi &Genzyme, Genentech, Novartis, BMS, Bayer, Horizon and Janssen. Dr Weinstock-Guttman received research funds from Biogen Idec, Genentech and Novartis. Lorena Lorefice received honoraria for consultancy and speaking from Biogen, Novartis, Sanofi, Genzyme, Merck and Bristol Myers Squibb. Menno Schoonheim: Serves on the editorial board of Neurology and Frontiers in Neurology, receives research support from the Dutch MS Research Foundation, Eurostars-EUREKA, ARSEP, Amsterdam Neuroscience, MAGNIMS and ZonMW (Vidi grant, project number 09150172010056) and has served as a consultant for or received research support from Atara Biotherapeutics, Biogen, Celgene/Bristol Meyers Squibb, EIP, Sanofi, MedDay and Merck. Sarah A. Morrow, in the last 3 years, has served as an advisory board member or received consulting fees from Biogen Idec; BMS/Celgene; EMDSerono; Novartis; Roche; Sanofi. She has participated in a speaker's bureau for Biogen Idec; BMS/Celgene; EMDSerono; Novartis; Roche; Sanofi. She has received research support from Biogen Idec; EMDSerono, Novartis, Roche; Sanofi Genzyme. She has participated as a site investigator in clinical trials sponsored by Bristol Myers Squibb/Celgene; EMDSerono; Novartis; Roche; Sanofi. Gabriel Pardo received grants (to the institution) from Biogen, EMD Serono, Roche/Genentech, Sanofi Genzyme, Novartis, Abbvie, and BMS; consultant and/or speaker bureau for Biogen, EMD Serono, Roche/Genentech, Sanofi Genzyme, Novartis, Janssen, BMS, TG Therapeutics, PRIME Education, and MSAA. Mark Gudesblatt received honoraria from Biogen and Genentech. Jacqueline Nicholas received research grants from Biogen, Novartis, PCORI, Genentech, University of Buffalo, EMD Serono; Consulting for EMD Serono, Genentech, Greenwich Biosciences, Novartis, TG Therapeutics and Sanofi; Speaking honoraria for BMS, EMD Serono, Horizon, TG Therapeutics. Andrew Smith received honorariums from EMD Serono and Sanofi Genzyme for speaking bureau and combination for Genzyme for an advisory board. Mahmoud A. AbdelRazek received consultant fees from Bristol Myers Squibb. Wachy Vongchucherd, Jon Riolo and Diego Silva are employees of Bristol Myers Squibb. Michael G. Dwyer has received personal compensation from Bristol Myers Squibb and Filterlex for consultant fees. He received financial support for research activities from Novartis, Bristol Myers Squibb, Octave, Mapi Pharma, Protembis, CorEvitas and V-WAVE Medical. Ralph HB. Benedict has received consultation or speaking fees from Bristol Myer Squibb, Biogen, Merck, EMD Serono, Roche, Immune Therapeutics, Novartis, and Sanofi-Genzyme.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2024.103609.

References

Amin, M., Ontaneda, D., 2020. Thalamic Injury and Cognition in Multiple Sclerosis. Front Neurol. 11, 623914.
Aubert-Broche, B., et al., 2014. Onset of multiple sclerosis before adulthood leads to failure of age-expected brain growth. Neurology 83, 2140–2146.
Azevedo, C.J., et al., 2015. Early CNS neurodegeneration in radiologically isolated syndrome. Neurol Neuroimmunol Neuroinflamm. 2, e102.
Azevedo, C.J., et al., 2018. Thalamic atrophy in multiple sclerosis: A magnetic resonance imaging marker of neurodegeneration throughout disease. Ann Neurol. 83, 223–234.
Barnett, M., et al., 2021. Brain atrophy and lesion burden are associated with disability progression in a multiple sclerosis real-world dataset using only T2-FLAIR: The NeuroSTREAM MSBase study. Neuroimage Clin. 32, 102802.
Beier, M., et al., 2017. Proposed cut scores for tests of the Brief International Cognitive Assessment of Multiple Sclerosis (BICAMS). J Neurol Sci. 381, 110–116.
Benedict, R., et al., 2012. Brief International Cognitive Assessment for MS (BICAMS): international standards for validation. BMC Neurol. 12, 55.
Benedict, R.H., et al., 2013. Clinical significance of atrophy and white matter mean diffusivity within the thalamus of multiple sclerosis patients. Mult Scler. 19, 1478–1484.
Benedict, R.H., et al., 2017. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult. Scler. J. 23, 721–733.
Bergsland, N., et al., 2015. Localized atrophy of the thalamus and slowed cognitive processing speed in MS patients. Mult Scler.
Bergsland, N., et al., 2021. Thalamic Nuclei Volumes and Their Relationships to Neuroperformance in Multiple Sclerosis: A Cross-Sectional Structural MRI Study. J Magn Reson Imaging. 53, 731–739.
Bisecco, A., et al., 2015. Connectivity-based parcellation of the thalamus in multiple sclerosis and its implications for cognitive impairment: A multicenter study. Hum Brain Mapp. 36, 2809–2825.
Bisecco, A., et al., 2021. Regional changes in thalamic shape and volume are related to cognitive performance in multiple sclerosis. Mult Scler. 27, 134–138.
Calabrese, M., et al., 2011. The predictive value of gray matter atrophy in clinically isolated syndromes. Neurology 77, 257–263.
Castrogiovanni, N., et al., 2023. Longitudinal Changes in Cognitive Test Scores in Patients With Relapsing-Remitting Multiple Sclerosis: An Analysis of the DECIDE Dataset. Neurology 101, e1–e11.
Debernard, L., et al., 2015. Deep grey matter MRI abnormalities and cognitive function in relapsing-remitting multiple sclerosis. Psychiatry Res. 234, 352–361.
DeLuca, J., et al., 2021. Effect of Ozanimod on Symbol Digit Modalities Test Performance in Relapsing MS. Mult Scler Relat Disord. 48, 102673.
Dwyer, M.G., et al., 2017. Neurological software tool for reliable atrophy measurement (NeuroSTREAM) of the lateral ventricles on clinical-quality T2-FLAIR MRI scans in multiple sclerosis. Neuroimage Clin. 15, 769–779.

Dwyer, M.G., et al., 2018. Establishing pathological cut-offs for lateral ventricular volume expansion rates. Neuroimage Clin. 18, 494–501.
Dwyer, M., et al., 2021. DeepGRAI (Deep Gray Rating via Artificial Intelligence): Fast, feasible, and clinically relevant thalamic atrophy measurement on clinical quality T2-FLAIR MRI in multiple sclerosis. Neuroimage Clin. 30, 102652.
Dwyer, M.G., Bergsland, N., Zivadinov, R., 2014. Improved longitudinal gray and white matter atrophy assessment via application of a 4-dimensional hidden Markov random field model. Neuroimage 90, 207–217.
Eshaghi, A., et al., 2018. Deep gray matter volume loss drives disability worsening in multiple sclerosis. Ann Neurol. 83, 210–222.
Fuchs, T.A., et al., 2018. White matter tract network disruption explains reduced conscientiousness in multiple sclerosis. Hum Brain Mapp. 39, 3682–3690.
Fuchs, T.A., et al., 2019. Preserved network functional connectivity underlies cognitive reserve in multiple sclerosis. Hum Brain Mapp. 40, 5231–5241.
Fuchs, T.A., et al., 2020. Functional connectivity and structural disruption in the default-mode network predicts cognitive rehabilitation outcomes in multiple sclerosis. J Neuroimaging. 30, 523–530.
Fuchs, T.A., et al., 2021. Quantifying disease pathology and predicting disease progression in multiple sclerosis with only clinical routine T2-FLAIR MRI. Neuroimage Clin. 31, 102705.
Gelineau-Morel, R., et al., 2012. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. Hum Brain Mapp. 33, 2802–2814.
Ghione, E., et al., 2020. Disability improvement is associated with less brain atrophy development in multiple sclerosis. AJNR Am J Neuroradiol.
Houtchens, M.K., et al., 2007. Thalamic atrophy and cognition in multiple sclerosis. Neurology 69, 1213–1223.
Jakimovski, D., et al., 2023. Therapy effect on AI-derived thalamic atrophy using clinical routine MRI protocol: A longitudinal, multi-center, propensity-matched multiple sclerosis study. Mult Scler Relat Disord. 74, 104708.
Jenkinson, M., et al., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.
Kuceyeski, A., et al., 2013. The Network Modification (NeMo) Tool: elucidating the effect of white matter integrity changes on cortical and subcortical structural connectivity. Brain Connect. 3, 451–463.
Kurtzke, J.F., 1983. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 33, 1444–1452.
Lin, F., et al., 2019. Altered nuclei-specific thalamic functional connectivity patterns in multiple sclerosis and their associations with fatigue and cognition. Mult Scler. 25, 1243–1254.
Mesaros, S., et al., 2008. Evidence of thalamic gray matter loss in pediatric multiple sclerosis. Neurology 70, 1107–1112.
Minagar, A., et al., 2013. The thalamus and multiple sclerosis: modern views on pathologic, imaging, and clinical aspects. Neurology 80, 210–219.
Ontaneda, D., et al., 2021. Deep grey matter injury in multiple sclerosis: a NAIMS consensus statement. Brain 144, 1974–1984.
Pareto, D., et al., 2019. Brain regional volume estimations with NeuroQuant and FIRST: a study in patients with a clinically isolated syndrome. Neuroradiology 61, 667–674.
Patenaude, B., et al., 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 56, 907–922.
Rocca, M.A., et al., 2010. Thalamic damage and long-term progression of disability in multiple sclerosis. Radiology 257, 463–469.
Schoonheim, M.M., et al., 2012. Subcortical atrophy and cognition: sex effects in multiple sclerosis. Neurology 79, 1754–1761.
Schoonheim, M.M., Broeders, T.A.A., Geurts, J.J.G., 2022. The network collapse in multiple sclerosis: An overview of novel concepts to address disease dynamics. Neuroimage Clin. 35, 103108.
Sjogard, M., et al., 2021. Brain dysconnectivity relates to disability and cognitive impairment in multiple sclerosis. Hum Brain Mapp. 42, 626–643.
Smith, S.M., et al., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 17, 479–489.
Steckova, T., et al., 2014. Thalamic atrophy and cognitive impairment in clinically isolated syndrome and multiple sclerosis. J Neurol Sci. 342, 62–68.
Till, C., et al., 2011. MRI correlates of cognitive impairment in childhood-onset multiple sclerosis. Neuropsychology 25, 319–332.
Ziccardi, S., et al., 2023. Cognitive phenotypes predict response to restorative cognitive rehabilitation in multiple sclerosis. Mult Scler. 13524585231208331.
Zivadinov, R., et al., 2012. Abnormal subcortical deep-gray matter susceptibility-weighted imaging filtered phase measurements in patients with multiple sclerosis: a case-control study. Neuroimage 59, 331–339.
Zivadinov, R., et al., 2013a. Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years. AJNR Am J Neuroradiol. 34, 1931–1939.
Zivadinov, R., et al., 2013b. Thalamic atrophy is associated with development of clinically definite multiple sclerosis. Radiology 268, 831–841.
Zivadinov, R., et al., 2016a. Clinical relevance of brain atrophy assessment in multiple sclerosis. Implications for its use in a clinical routine. Expert Rev Neurother. 1–17.
Zivadinov, R., et al., 2016b. A serial 10-year follow-up study of brain atrophy and disability progression in RRMS patients. Mult Scler. 22, 1709–1718.
Zivadinov, R., et al., 2018. 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. 39, 289–295.
Zivadinov, R., et al., 2022. Thalamic atrophy measured by artificial intelligence in a multicentre clinical routine real-word study is associated with disability progression. J Neurol Neurosurg Psychiatry.