NeuroImage: Clinical 29 (2021) 102554

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

Clinical feasibility of longitudinal lateral ventricular volume measurements on T2-FLAIR across MRI scanner changes

Dejan Jakimovski [a], Robert Zivadinov [a,b], Niels Bergsland [a,c], Deepa P. Ramasamy [a], Jesper Hagemeier [a], Antonia Valentina Genovese [d], David Hojnacki [e], Bianca Weinstock-Guttman [e], Michael G. Dwyer [a,b,*]

[a] Buffalo Neuroimaging Analysis Center, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA
[b] Center for Biomedical Imaging at Clinical Translational Science Institute, University at Buffalo, State University of New York, Buffalo, NY, USA
[c] IRCCS, Fondazione Don Carlo Gnocchi ONLUS, Milan, Italy
[d] Institute of Radiology, Department of Clinical Surgical Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy
[e] Jacobs Comprehensive MS Treatment and Research Center, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences University at Buffalo, Buffalo, NY, USA

ARTICLE INFO

Keywords:
Multiple sclerosis
MRI
Whole brain atrophy
Lateral ventricular volume
Disability progression
MRI field strength

ABSTRACT

Background: Greater brain atrophy is associated with disability progression (DP) in patients with multiple sclerosis (PwMS). However, methodological challenges limit its routine clinical use.
Objective: To determine the feasibility of atrophy measures as markers of DP in PwMS scanned across different MRI field strengths.
Methods: A total of 980 PwMS were scanned on either 1.5 T or 3.0 T MRI scanners. Demographic and clinical data were retrospectively collected, and the presence of DP was determined according to standard clinical trial criteria. Lateral ventricular volume (LVV) change was measured with the NeuroSTREAM technique on clinical routine T2-FLAIR images. Percent brain volume change (PBVC) was measured using SIENA and ventricular cerebrospinal fluid (vCSF) % change was measured using VIENA and SIENAX algorithms on 3D T1-weighted images (WI). Stable vs. DP PwMS were compared using analysis of covariance (ANCOVA). Mixed modeling determined the effect of MRI scanner change on MRI-derived atrophy measures.
Results: Longitudinal LVV analysis was successful in all PwMS. SIENA-based PBVC and VIENA-based changes failed in 37.6% of cases, while SIENAX-based vCSF failed in 12.9% of cases. PwMS with DP ($n = 241$) had significantly greater absolute (20.9% vs. 8.7%, $d = 0.66$, $p < 0.001$) and annualized LVV % change (4.1% vs. 2.3%, $d = 0.27$, $p < 0.001$) when compared to stable PwMS ($n = 739$). In subjects with both analyses available, both 3D-T1 and T2-FLAIR-based analyses differentiated PwMS with DP ($n = 149$). However, only NeuroSTREAM and VIENA-based LVV/vCSF were able to show greater atrophy in PwMS that were scanned on different scanners. PBVC and SIENAX-based vCSF % changes were significantly affected by scanner change (Beta = −0.16, t-statistics = −2.133, $p = 0.033$ and Beta = −2.08, t-statistics = −4.084, $p < 0.001$), whereas no MRI scanner change effects on NeuroSTREAM-based PLVVC and VIENA-based vCSF % change were noted.
Conclusions: LVV-based atrophy on T2-FLAIR is a clinically relevant measure in spite of MRI scanner changes and mild disability levels.

Abbreviations: MS, multiple sclerosis; PwMS, patients with MS; WBV, whole brain volume; FLAIR, fluid-attenuated inversion recovery; SNR, signal-to-noise ratio; CNR, contrast-to-noise ratio; LVV, lateral ventricular volume; CSF, cerebrospinal fluid; PLVVC, percent lateral ventricular volume change; PBVC, percent brain volume change; EDSS, Expanded Disability Status Scale; DMT, disease modifying treatment; RRMS, relapsing-remitting MS; SPMS, secondary-progressive MS; PPMS, primary-progressive MS; PMS, progressive MS; AN(C)OVA, analysis of (co)variance; CI, confidence intervals; SE, standard error; DP, disability progression; IQR, interquartile range; SIENA, Structural Image Evaluation using Normalisation of Atrophy; vCSF, ventricular cerebrospinal fluid.

* Corresponding author at: Buffalo Neuroimaging Analysis Center, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, 100 High Street, Buffalo, NY 14203, USA.
E-mail address: mgdwyer@bnac.net (M.G. Dwyer).

https://doi.org/10.1016/j.nicl.2020.102554
Received 15 July 2020; Received in revised form 24 December 2020; Accepted 29 December 2020
Available online 4 January 2021

1. Introduction

Multiple sclerosis (MS) is both an inflammatory and neurodegenerative disease of the central nervous system that typically presents with intermittent neurological disability followed by either full or partial functional recovery (Thompson et al., 2018). In addition to the classical inflammatory lesions, recent literature describes a relentless development of brain atrophy, which is detectable even at the initial stages of the disease and continues to accelerate (Zivadinov et al., 2013; Ghione et al., 2019). Moreover, the extent of neurodegeneration correlates with clinical MS outcomes, drives cognitive impairment, and better predicts both short- and long-term disability progression (Rocca et al., 2017). Therefore, there is an increasing need for clinical implementation of reliable, easy-to-acquire, and non-invasive biomarkers of neurodegeneration that will facilitate and improve the overall preventive or rehabilitative interventions. Along these lines, prospects such as analysis of serum neurofilament light chain levels and MRI-derived brain volumes/brain volume loss have been extensively studied (Jakimovski et al., 2019).

The use of MRI-based brain atrophy measures in routine clinical care of patients with MS (PwMS) is currently limited by multiple biological and technological factors (Zivadinov et al., 2016). These include daily brain volume fluctuations derived from levels of hydration, diurnal changes, menstrual cycle, and treatment-induced pseudoatrophy (Zivadinov et al., 2016). Similarly, continuous hardware and software updates of the MRI machines can have a significant impact on image comparability and reproducibility. In particular, many imaging centers are currently undergoing a major transition in their MRI hardware inventory with efforts to upgrade old scanners or to move to newer 3.0 T scanners.

Currently, there are multiple semi-automated and automated methods that allow quantification of brain volume changes (Vrenken et al., 2013). Whole brain volume (WBV) longitudinal analyses can be performed by readily available software such as Structural Image Evaluation using Normalisation of Atrophy (SIENA, https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/), Boundary Shift Integral (BSI, https://sourceforge.net/projects/bsintegral/), SepINRIA (http://www-sop.inria.fr/asclepios/software/SepINRIA/), and Statistical Parametric Mapping (SPM, https://www.fil.ion.ucl.ac.uk/spm/). In order to achieve acceptable reliability, these methods typically require 3D T1-weighted sequences, full head coverage, careful parameter optimization, and additional image processing (Nordenskjöld et al., 2013; Popescu et al., 2012). Signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) are significantly affected by changes in scanner field and these in turn can have a substantial impact on most segmentation methods (Jovicich et al., 2009). On the other hand, a recently-described processing pipeline called NeuroSTREAM uses commonly acquired T2 fluid-attenuated inversion recovery (T2-FLAIR) sequences to provide cross-sectional measures as well as longitudinal calculation of changes in lateral ventricular volume (LVV) (Dwyer et al., 2017). Benefits from using LVV as a proxy of brain atrophy and MS neurodegeneration include improved segmentation consistency derived from high CNR between cerebrospinal fluid (CSF) and brain tissue, lower variability due to smaller relative partial volume, and feasibility within scans that were cut-off or contain artifacts (Dwyer et al., 2017).

Against this background, we aimed to determine the feasibility of NeuroSTREAM-based percent LVV changes (PLVVC) on T2-FLAIR when compared to SIENA-based percent brain volume change (PBVC) and VIENA-based ventricular (vCSF) % change measured using the 3D T1-weighted images (WI) in a single, clinical-based setting which interchangeably used both 1.5 T and 3.0 T MRI scanners.

2. Materials and methods

2.1. Study population

The study population used for this analysis was part of a larger and previously described cohort of PwMS (Ghione et al., 2018). In short, PwMS who were recruited and scanned during the period of 2006–2016 were included in a retrospective analysis of their demographic, clinical, and MRI-based characteristics (Ghione et al., 2018). The inclusion criteria for this study were: (a) diagnosis of MS according to the McDonald criteria (Polman et al., 2011), (b) availability of both T2-FLAIR and 3D T1-WI sequences, (c) presence of a baseline and follow-up MRI exam within the same subject over at least a 6 month follow up period, and (d) availability of demographic and clinical information at the baseline and at the follow-up examinations. Exclusion criteria were: (a) having a relapse or steroid treatment in the 30 days preceding the MRI examination (b) pre-existing medical conditions known to be associated with brain pathology (cerebrovascular disease, positive history of alcohol abuse). Additional exclusions included nursing and pregnant mothers. The retrospective electronic medical chart analysis collected basic demographic and clinical information including age, sex, race, disease duration, and type of disease modifying treatment (DMT). All PwMS were seen by an experienced neurologist who assessed Expanded Disability Status Scale (EDSS) scores (Kurtzke, 1983). The definition of disability progression (DP) consisted of 1) an increase of $\geq$ 1.5 EDSS points when baseline EDSS was $\geq$ 0, 2) $\geq$ 1.0 increase if baseline EDSS was $<$ 5.0, or 3) $\geq$ 0.5 increase if baseline EDSS was $\geq$ 5.5.

Based on their clinical presentation, PwMS were classified as either relapsing-remitting MS (RRMS), secondary-progressive MS (SPMS), or primary-progressive MS (PPMS) (Lublin et al., 2014). DMT use was grouped into interferon-β, glatiramer acetate, natalizumab, oral DMTs, off-label medications, and no DMT use at the time of enrollment. Longitudinal DMT changes were categorized as remained on the same medication, switched DMT, discontinued DMT, or started DMT over the follow-up period.

2.2. MRI acquisition and analyses

The study population underwent structural MRI examination using either a 1.5 T or 3.0 T Signa Excite 12 Twin-Speed scanner (GE Healthcare, Milwaukee, MI, USA) with an 8-channel head and neck coil. The 3 T sequences used for this analysis included an axial 2D T2-FLAIR sequence with echo time(TE)/inversion time(TI)/repetition time(TR) of 120 ms/2100 ms/8500 ms, flip angle of 75 degrees, echo train length of 24, phase FOV of 265x192 and 1 $\times$ 1 $\times$ 3 mm matrix. The 3D T1-WI sequence had TE/TI/TR of 5.9 ms/900 ms/2.8 ms flip angle of 10 degrees, FOV of 265 $\times$ 192, and isotropic 1 mm acquisition matrix. Similarly, the sequences acquired on 1.5 T were 2D T2-FLAIR with TE/TI/TR if 129 ms/2000 ms/8000 ms and 3D T1-WI with TE/TI/TR of 3.7 ms/900 ms/7.7 ms. FOV and acquisition matrix parameters remained the same.

WBV and PBVC were measured on 3D T1-WI images using the SIENAX and SIENA algorithms, respectively (Smith et al., 2002). Similarly, SIENAX and VIENA algorithms provided cross-sectional volume of vCSF space and its longitudinal change over the follow-up period, respectively. Inpainting of T1-hypointensities was utilized to minimize tissue mis-classification (Gelineau-Morel et al., 2012). Additionally, NeuroSTREAM software was utilized for calculation of cross-sectional and longitudinal changes in LVV (Dwyer et al., 2017). Total and annualized PBVC, PLVVC, VIENA, and SIENAX-based vCSF % changes were calculated.

2.3. Statistical analyses

All statistical analyses were performed with SPSS (version 25, IBM, Armonk, NY, USA). The distribution of the data and its residuals were

determined by manual inspection of Q-Q plots. Demographic and clinical comparisons were performed using χ^2 for categorical variables, Student's *t*-test and analysis of variance (ANOVA) with Bonferroni-adjusted post-hoc comparisons for normally distributed variables, and Mann-Whitney *U* test and Kruskal-Wallis H test for non-parametric variables. For analysis of MRI variables, both Student's *t*-test and age-adjusted analyses of covariance (ANCOVA) with Bonferroni-adjusted post hoc comparisons were performed. If the assumption of homoscedasticity was violated (Levene's test of equality < 0.05), a Brown-Forsythe test of equality was performed instead. Cohen's d effect size was also utilized. Analyses were performed on the full PwMS population, the respective subgroups of PwMS scanned on each possible MRI scanner combination (1.5 T-1.5 T, 1.5 T-3.0 T, 3.0 T-1.5 T and 3.0 T-3.0 T), and on the total and individual scanner subgroups of PwMS that had successful NeuroSTREAM LVV, SIENA-based PBVC, VIENA-based vCSF and SIENAX-based vCSF MRI measures, respectively.

The effects of sex, age, DP, and changes in MRI scanner strength on longitudinal changes assessed by NeuroSTREAM-based PLVVC, VIENA-based vCSF % change, SIENAX-based vCSF % change, and SIENA PBVC were determined using mixed-effects modeling. Model-based parameter estimates including betas, standard error (SE), t-statistics and 95% confidence intervals (CI) were reported. Interaction effects between DP and time of follow-up and between DP and change in MRI field strength on the atrophy measures were also computed. P-values lower than 0.05 were considered statistically significant. Bar plots demonstrating differences in longitudinal brain atrophy measures between stable and PwMS with DP were created in GraphPad Software (San Diego, CA, USA).

3. Results

3.1. Study population

The clinical and demographic characteristics of the entire study population (n = 980) are shown in Table 1. This study group consisted of mostly Caucasian (n = 872, 89%) female (n = 742, 75.7%) PwMS, on average 45.7 years old, with a mean disease duration of 12.3 years and median baseline EDSS score of 2.5 (IQR 1.5–4.0). At baseline, the group included 794 (81%) RRMS, 163 (16.6%) SPMS and 23 (2.3%) PPMS. Furthermore, almost half of the population (n = 467, 47.7%) was treated with interferon-β, 202 (20.6) were treated with glatiramer acetate, 75 (7.7%) with natalizumab, 14 (1.4%) with oral DMTs and 14 (1.4%) with other off-label medications. At the baseline visit, 144 (14.7%) PwMS were not on any treatment and DMT data was missing on 64 (6.5%) PwMS. The average follow-up time was 4.8 years with approximately 1 MRI per year (average number of MRIs per patient was 5.1 scans). During the follow-up period, the entire group had a mean EDSS change of 0.4 points, and 241 (24.6) individuals developed DP. Lastly, 419 (42.8%) PwMS remained on the same DMT, 275 (28.1%) switched DMT type, 111 (11.3%) started DMT, 74 (7.6%) stopped their DMT, 80 (8.2%) remained on no DMT, and data was missing for 21 (2.1%) PwMS.

Similar clinical and demographic characteristics were seen for each MRI scanner combination group (Table 1). Most importantly, there was no difference in the prevalence of DP between the groups (χ^2 p = 0.62).

3.2. MRI-based brain atrophy measures between DP and stable PwMS in the entire study population

Apart from an age difference between stable and DP PwMS (45.3 vs. 47.1 years old, p = 0.018), there were no other significant demographic differences at baseline. There were no demographic differences within each subgroup analysis between the individual MRI scanner combinations (Supplement Table 1).

The differences in NeuroSTREAM-based LVV and SIENAX-based vCSF measures between stable and DP groups are shown in Table 2 and Fig. 1. T2-FLAIR and NeuroSTREAM-based LVV was available in 100% of cases (n = 980), whereas T1-WI SIENAX-based vCSF was available in 87.1% of cases (n = 854).

In the total sample, PwMS with DP (n = 241) had significantly greater absolute PLVVC (20.9% vs. 8.7%, d = 0.66, p < 0.001) and annualized PLVVC (4.1% vs. 2.3%, d = 0.27, p < 0.001) when compared to stable PwMS. Greater PLVVC and annualized PLVVC in PwMS with DP was also seen in the 1.5 T-1.5 T subgroup (19.8% vs. 6.5%, d = 0.75, p < 0.001 and 4.3% vs. 1.2%, d = 0.49, p = 0.001, respectively) and the 1.5 T-3.0 T group (28% vs. 10.3%, d = 0.88, p < 0.001 and 5.5% vs.

Table 1
Demographic and clinical characteristics of the study population, across scanner combinations.

	MS patients (n = 980)	1.5 T-1.5 T (n = 201)	1.5 T-3.0 T (n = 341)	3.0 T-1.5 T (n = 112)	3.0 T-3.0 T (n = 326)
Female, n (%)	742 (75.7)	156 (77.6)	264 (77.4)	84 (75.0)	238 (73.0)
Age at baseline, mean (SD)	45.7 (10.8)	47.3 (11.5)	46.0 (10.2)	46.3 (10.1)	44.3 (10.9)
Race, n (%)					
- Caucasian	872 (89.0)	183 (91.0)	296 (86.8)	96 (85.7)	297 (91.1)
- African – American	98 (10.0)	17 (8.5)	42 (12.3)	13 (11.6)	26 (8.0)
- Other	10 (1.0)	1 (0.5)	3 (0.9)	3 (2.7)	3 (0.9)
Time of follow-up, mean (SD)	4.8 (2.4)	4.5 (2.4)	5.5 (2.3)	4.1 (3.5)	4.5 (2.7)
Average number of MRIs per year, mean (SD)	5.1 (3.0)	4.9 (3.2)	5.8 (3.3)	4.2 (2.2)	4.8 (2.7)
Disease duration at baseline, mean (SD)	12.3 (9.5)	14.5 (10.1)	12.2 (9.3)	11.1 (8.3)	11.3 (9.7)
Disease subtype, n (%)					
RR	794 (81.0)	142 (70.6)	276 (80.9)	92 (82.1)	284 (87.1)
SP	163 (16.6)	52 (25.9)	56 (16.4)	18 (16.1)	37 (11.3)
PP	23 (2.3)	7 (3.5)	9 (2.6)	2 (1.8)	5 (1.5)
EDSS score at baseline, median (IQR)	2.5 (1.5–4.0)	3.5 (2.0–6.0)	3.0 (1.5–4.5)	2.5 (1.5–3.5)	2.5 (1.5–3.5)
EDSS score change, mean (SD)	0.4 (1.2)	0.4 (1.1)	0.3 (1.2)	0.5 (1.2)	0.3 (1.2)
Disease-modifying therapy at baseline, n (%)					
- Non-therapy	144 (14.7)	22 (10.9)	54 (15.8)	19 (16.9)	49 (15.0)
- Interferon-β	467 (47.7)	97 (48.3)	160 (46.9)	62 (55.4)	148 (45.4)
- Glatiramer acetate	202 (20.6)	47 (23.4)	65 (19.1)	17 (15.2)	73 (22.4)
- Natalizumab	75 (7.7)	13 (6.5)	29 (8.5)	9 (8.0)	24 (7.4)
- Oral DMT	14 (1.4)	2 (0.9)	–	–	12 (3.7)
- Other therapies	14 (1.4)	3 (1.5)	7 (2.1)	3 (2.7)	1 (0.3)
- Unknown	64 (6.5)	17 (8.5)	26 (7.6)	2 (1.8)	19 (5.8)
DP over the follow-up, n (%)	241 (24.6)	52 (25.9)	83 (24.3)	32 (28.6)	74 (22.7)

MS – multiple sclerosis, CIS – clinically isolated syndrome, RR – relapsing-remitting, SP – secondary progressive, PP – primary progressive, EDSS – Expanded Disability Status Scale, CDMS – clinically-definite multiple sclerosis, DP – disability progression, SD – standard deviation, IQR – interquartile range.

Table 2
Use of NeuroSTREAM-based LVV and SIENAX-based vCSF measures in clinical routine across scanner combinations.

	NeuroSTREAM-based LVV	Baseline LVV	LVV % Change	Annualized LVV % Change	SIENAX-based vCSF	Baseline vCSF	vCSF % Change	Annualized vCSF % Change
Total Sample[a]	DP (n = 241)	24.0 (13.4)	20.9 (22.1)	4.1 (7.8)	DP (n = 207)	48.1 (20.0)	16.8 (19.1)	3.4 (4.9)
	Stable (n = 739)	23.2 (13.6)	8.7 (13.6)	2.3 (5.3)	Stable (n = 647)	44.4 (19.3)	9.2 (21.1)	2.2 (6.4)
	Cohen's d	0.06	0.66	0.27	Cohen's d	0.19	0.38	0.21
	P – value	0.889	**<0.001**	**<0.001**	P – value	0.083	**<0.001**	**0.015**
1.5 T-1.5 T	DP (n = 52)	24.4 (13.4)	19.8 (20.7)	4.3 (4.9)	DP (n = 45)	49.6 (18.2)	17.5 (19.2)	3.3 (4.4)
	Stable (n = 149)	23.3 (13.5)	6.5 (14.1)	1.2 (7.4)	Stable (n = 126)	45.8 (20.7)	11.1 (13.6)	2.6 (3.6)
	Cohen's d	0.08	0.75	0.49	Cohen's d	0.19	0.39	0.17
	P – value	0.591	**<0.001***	**0.001***	P – value	0.275	**0.049***	0.283
1.5 T-3.0 T	DP (n = 83)	22.6 (12.5)	28.0 (24.6)	5.5 (7.8)	DP (n = 72)	47.8 (20.2)	16.1 (20.5)	3.2 (5.7)
	Stable (n = 258)	24.4 (15.6)	10.3 (14.4)	2.8 (4.2)	Stable (n = 223)	47.9 (20.9)	4.3 (15.9)	0.6 (3.4)
	Cohen's d	0.127	0.88	0.43	Cohen's d	0.01	0.64	0.55
	P – value	0.352	**<0.001***	**0.004**	P – value	0.943	**<0.001***	**0.001***
3.0 T-1.5 T	DP (n = 32)	26.5 (16.9)	10.0 (18.6)	1.4 (7.1)	DP (n = 27)	52.8 (25.1)	16.2 (14.0)	5.2 (5.4)
	Stable (n = 80)	21.9 (11.5)	3.9 (11.4)	0.5 (4.4)	Stable (n = 71)	40.1 (15.3)	13.5 (10.8)	3.8 (3.9)
	Cohen's d	0.32	0.39	0.15	Cohen's d	0.61	0.22	0.29
	P – value	0.163	0.089*	0.43	P – value	**0.019***	0.325	0.174
3.0 T-3.0 T	DP (n = 74)	24.2 (12.7)	18.4 (18.9)	3.5 (6.0)	DP (n = 62)	45.3 (18.6)	17.3 (19.8)	3.0 (3.6)
	Stable (n = 249)	22.5 (12.1)	9.9 (12.4)	2.9 (4.9)	Stable (n = 227)	41.6 (17.2)	11.5 (20.2)	3.0 (9.6)
	Cohen's d	0.14	0.53	0.11	Cohen's d	0.21	0.29	0.0
	P - value	0.28	**<0.001***	0.405	P - value	0.137	0.143	0.973

LVV – lateral ventricular volume, vCSF – ventricular cerebrospinal fluid volume, DP – disease progression.
All measures are shown as mean (standard deviation).
[a] Age-adjusted analyses of covariance (ANCOVA) with Bonferroni-adjusted post hoc comparisons were performed. Student's *t*-test was performed unless otherwise specified. *- due to heteroskedasticity (Levene's test < 0.05), Brown-Forsythe test of equality was performed instead. P-value lower than 0.05 was considered statistically significant and shown in bold.

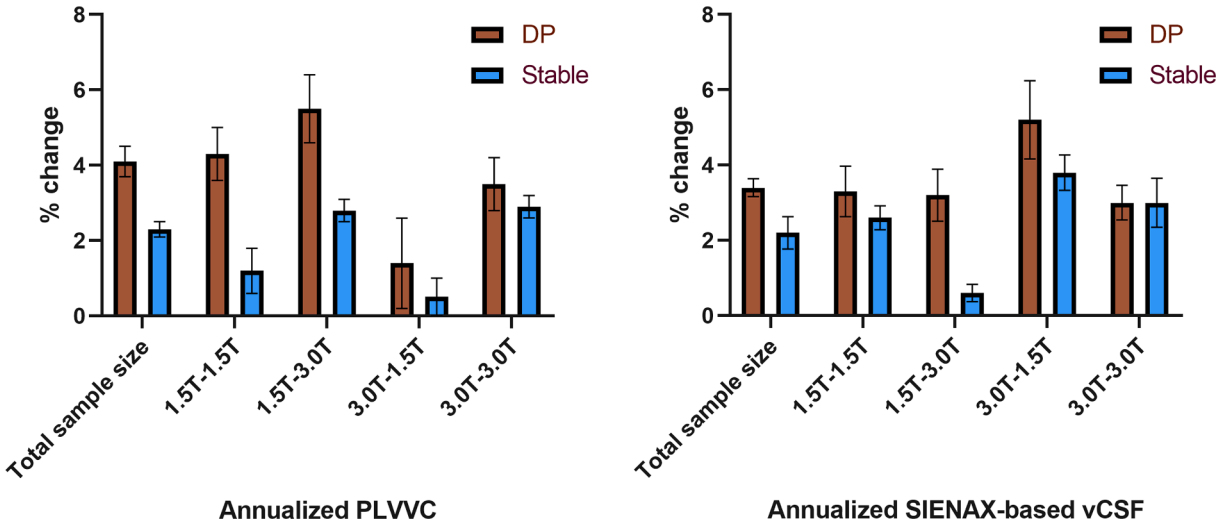

Fig. 1. Bar plot demonstrating annualized NeuroSTREAM-based PLVVC and annualized SIENAX-based vCSF % changes in the entire study population and in each MRI scanner combination. PLVVC – percent lateral ventricular volume change, vCSF – ventricular cerebrospinal fluid. Error bars represent standard error of mean.

2.8%, d = 0.43, p = 0.004, respectively). Although not statistically significant, PwMS with DP had numerically greater PLVVC when compared to stable PwMS (10% vs. 3.9%, d = 0.39, p = 0.089 and 1.4% vs. 0.5%, d = 0.15, p = 0.43, respectively) in the 3.0 T-1.5 T subgroup and in the 3.0 T-3.0 T subgroup (18.4% vs. 9.9%, d = 0.53, p < 0.001 and 3.5% vs. 2.9%, d = 0.11, p = 0.405, respectively). Neither the total population, nor any MRI scanner subgroup had differences in LVV between stable PwMS and PwMS with DP.

Additional head-to-head comparison between NeuroSTEAM-based PLVVC and SIENAX-based vCSF % change was performed (Table 2). Both methods utilize a similar procedure with subtraction of two time points and measure a relatively similar amount of volume (SIENAX-based vCSF additionally includes the 3rd and 4th ventricle). In the total sample size, PwMS with DP had significantly greater SIENAX-based total and annualized vCSF % change when compared to stable PwMS (16.8% vs. 9.2%, d = 0.38, p < 0.001 and 3.4% vs. 2.2%, d = 0.21, p = 0.015). The effect size with SIENAX-based vCSF measures was comparably lower than the analogous NeuroSTREAM-based analyses. SIENAX-based annualized vCSF % change differentiated PwMS with DP and stable PwMS only in the 1.5 T-3.0 T subgroup (3.2% vs. 0.6%, d = 0.55, p = 0.001). Lastly, PwMS with DP had significantly greater total SIENAX-based vCSF % change when compared to stable PwMS in the 1.5 T-1.5 T group (17.5% vs. 11.1%, d = 0.39, p = 0.049).

3.3. MRI-based brain atrophy measures between DP and stable PwMS in the sub-population of subjects with T2-FLAIR and 3D T1-WI available analyses

SIENA-based PBVC and VIENA-based vCSF % change quantification failed in 368 (37.6%) of PwMS. The demographic and clinical characteristics of the subgroup of stable/DP PwMS with both successful NeuroSTREAM, SIENA, VIENA and SIENAX vCSF analyses are shown in Supplement Table 2 and Supplement Table 3. Within the sample size of PwMS with both analyses, PwMS with DP were older when compared to stable PwMS (44.4 vs. 46.7 years old, $p = 0.023$). No other demographic differences were observed. Furthermore, there were no demographic differences between stable PwMS and PwMS with DP in any of the individual MRI scanner subgroups.

The NeuroSTREAM-based PLVVC, SIENA-based PBVC, VIENA-based vCSF % change, and SIENAX-based vCSF % changes in the remaining 612 PwMS are shown in Table 3 and Fig. 2. PwMS with DP ($n = 149$) had significantly greater total and annualized PLVVC (21.3% vs. 8.3%, $d = 0.69$, $p < 0.001$ and 4.7% vs. 2.2%, $d = 0.37$, $p < 0.001$), greater total and annualized PBVC (−4.6% vs. −3.3%, $d = 0.44$, $p < 0.001$ and −0.99% vs. −0.79%, $d = 0.18$, $p < 0.001$), greater total and annualized VIENA-based % change (24.5% vs. 10.8%, $d = 0.81$, $p < 0.001$ and 5.1% vs. 2.6%, $d = 0.48$, $p < 0.001$) and greater total and annualized SIENAX-based vCSF % change (15.2% vs. 7.8%, $d = 0.43$, $p < 0.001$ and 3.4% vs. 1.9%, $d = 0.31$, $p = 0.001$) when compared to stable PwMS ($n = 463$).

Both PwMS groups with no change in MRI strength field (1.5 T-1.5 T and 3.0 T-3.0 T) had similar findings. PwMS with DP who were scanned only on the 1.5 T scanner had significantly greater total and annualized PLVVC (19.8% vs. 6.4%, 0.74, $p = 0.001$ and 5.2% vs. 1.4%, $d = 0.55$, $p = 0.011$), greater total PBVC (-5.0% vs. −3.9%, $d = 0.36$, $p = 0.038$), greater total and annualized VIENA-based % change (20.9% vs. 10.1%, $d = 0.64$, $p < 0.001$ and 4.9% vs. 2.8%, $d = 0.52$, $p = 0.006$) and greater total SIENAX-based vCSF % change (4.9% vs. 2.8%, $d = 0.52$, $p = 0.006$) when compared to stable PwMS. PwMS with DP had only a numerically greater annualized PBVC (-1.3% vs. 0.91%, $d = 0.36$, $p = 0.092$) when comparted to stable PwMS. Similarly, PwMS with DP who were scanned only on the 3.0 T scanner exhibited significantly greater total PLVVC, PBVC, VIENA-based vCSF % change (18.4% vs. 10.1%, $d = 0.47$, $p = 0.015$; −3.7% vs. −2.5%, $d = 0.43$, $p = 0.005$; and 21.6% vs. 10.7%, $d = 0.57$, $p = 0.004$), and total SIENAX-based vCSF % change (15.6% vs. 9.3%, $d = 0.35$, $p = 0.028$) compared to stable PwMS. Only annualized VIENA-based % change was significantly different between stable PwMS and PwMS with DP scanned on 3 T (4.4% vs. 2.9%, $d = 0.36$, $p = 0.037$).

PwMS with DP for whom the MRI scanner changed from 1.5 T at baseline to 3.0 T at follow-up had significantly greater total and annualized PLVVC (29.3% vs. 9.1%, $d = 0.99$, $p < 0.001$ and 6.0% vs. 2.4%, $d = 0.51$, $p = 0.007$), greater total and annualized VIENA-based vCSF % change (31.1% vs. 12.4%, $d = 0.83$, $p < 0.001$ and 6.6% vs. 2.5%, $d = 0.61$, $p = 0.002$), and greater total and annualized SIENAX-based vCSF % change (14.7% vs. 3.3, $d = 0.65$, $p < 0.001$ and 3.3% vs. 0.5%, $d = 0.56$, $p = 0.004$) when compared to stable PwMS. The same PwMS showed greater total PVBC (−5.4% vs. −3.8%, $d = 0.53$, $p = 0.002$), but no differences in annualized PBVC. Lastly, within the small group of PwMS whose scanner changed from 3.0 T at baseline to 1.5 T at follow-up, there were no significant differences in any MRI-based measure between stable PwMS and PwMS with DP.

3.4. MRI field strength effects on ventricular and whole brain atrophy changes

The main and interaction effects of the study variables on the changes in NeuroSTREAM-based LVV, SIENA-based PBVC, and VIENA- and SIENAX-based vCSF % changes are shown in Table 4. Expansion of the LVV was associated with amount of follow-up time (Beta = 5.57, SE = 0.44, t-statistics = 12.686, $p < 0.001$), presence of DP (Beta = 9,5, SE = 2.43, t-statistics = 3.915, $p < 0.001$), and the interaction between those factors (Beta = −4.48, SE = 0.45, t-statistics = −9.962, $p < 0.001$). Based on these estimates, experiencing DP over one year of follow-up would be associated with a 4.48% increase in NeuroSTREAM-LVV.

SIENA-based PBVC was also significantly associated with the follow-up time (Beta = −0.87, SE = 0.085, t-statistics = −10.17, $p < 0.001$), time, DP interaction (Beta = 0.33, SE = 0.087, t-statistics = 3.736, $p < 0.001$), and an interaction between time and change in MRI scanner strength (Beta = −0.16, SE = 0.077, t-statistics = −2.133, $p = 0.033$). In terms of effects on PBVC, experiencing DP over one year of follow-up would be associated with 0.33% in lower WBV. However, change in MRI scanner strength would be associated with an opposite effect of 0.16% over the same amount of follow-up period.

VIENA-based vCSF % change was associated with amount of follow-up time (Beta = 5.69, SE = 0.58, t-statistics = 9.786, $p < 0.001$), sex (Beta = −4.35, SE = 1.46, t-statistics = −2.972, $p = 0.003$), presence of DP (Beta = 6.35, SE = 3.09, t-statistics = 2.058, $p = 0.004$), and interaction between time of follow-up and presence of DP (Beta = −4.13, SE = 0.59, t-statistics = −6.883, $p < 0.001$). Similar to NeuroSTREAM-based LVV, experiencing DP over one year of follow-up was associated with a 4.13% increase in vCSF space.

In multivariate mixed model analysis, SIENAX-based vCSF % change was not associated with the presence of DP (Beta = 5.49, SE = 3.02, t-statistics = −1.823, $p = 0.069$). SIENAX-based vCSF % change was associated with the follow-up time (Beta = 3.17, SE = 0.57, t-statistics = −5.525, $p < 0.001$), the interaction between time of follow-up and presence of DP (Beta = −2.95, SE = 0.59, t-statistics = −4.986, $p < 0.001$), and the interaction between time and change in MRI field strength (Beta = 2.08, SE = 0.51, t-statistics = 4.084, $p < 0.001$). Although presence of DP was associated with a 2.95% yearly increase in SIENAX-based vCSF % change, change in MRI scanner contributed with 2.08% change in the opposite direction.

4. Discussion

In this large, single-center MRI analysis, we demonstrated that NeuroSTREAM-based LVV can distinguish between PwMS with and without DP on a group level. Furthermore, changes in MRI scanner strength over follow-up did not interfere with LVV analysis on T2-FLAIR. On the other hand, both SIENAX-based vCSF % change and SIENA-based PBVC were significantly affected by change in MRI scanner strength and clinically feasible only when both examinations were performed on the same MRI scanner. Moreover, the broader feasibility of NeuroSTREAM-based LVV was observed despite allowing for SIENA/VIENA failure and exclusion from analysis in a large number of PwMS. As such, the seemingly comparable performance between VIENA and NeuroSTREAM should be interpreted in the context of a significant selection bias of high-quality scans for VIENA outcomes.

As seen in our mixed models, SIENAX-based and SIENA-based atrophy measures were significantly impacted by change in MRI scanner field strength. Due to the effect of MRI field strength change, which negated as much as half of the change associated with disability progression (0.16% vs. 0.33% of annualized WBV change), these atrophy measures were unsuccessful in differentiating between stable PwMS and PwMS with DP. In contrast, longitudinal LVV was not significantly impacted by change in MRI field strength and maintained the ability to differentiate the PwMS groups. As noted above, it is also important to mention that this analysis has already taken into account the considerable failure rate for the 3D T1-WI analyses which was present in 37.6% of our cases. If the analysis had included rejected SIENA/VIENA outcomes, which were more frequent with scanner change, the differences in performance would have been much greater and would likely have altered the overall findings in the total PwMS sample by adding a significant scanner influence on both SIENA and VIENA outcomes. In addition, implementation of longitudinal WBV protocols are generally contingent on 3D T1-WI images, a limitation which is not present for the T2 FLAIR-based LVV segmentation (Leigh et al., 2002). A recent multi-

Table 3
Study analysis of cases with both NeuroSTREAM, SIENA, VIENA and SIENAX vCSF analyses across scanner combinations.

	NeuroSTREAM-based LVV	Baseline LVV	LVV % Change	Annualized LVV % Change	SIENAX-based WBV	PBVC	Annualized PBVC	VIENA % Change	Annualized VIENA % Change	SIENAX-based vCSF	SIENAX-vCSF % Change	Annualized vCSF% Change
Total Sample[a]	DP (n = 149)	23.5 (12.8)	21.3 (23.3)	4.7 (7.6)	1518.4 (107.2)	−4.6 (3.3)	−0.99 (1.3)	24.5 (19.9)	5.1 (6.6)	47.9 (19.9)	15.2 (19.8)	3.4 (5.4)
	Stable (n = 463)	22.2 (12.7)	8.3 (13.2)	2.2 (5.6)	1539.5 (98.5)	−3.3 (2.6)	−0.79 (0.9)	10.8 (13.5)	2.6 (3.4)	43.7 (18.7)	7.8 (13.9)	1.9 (4.1)
	Cohen's d	0.07	0.69	0.37	0.19	0.44	0.18	0.81	0.48	0.22	0.43	0.31
	P – value	0.609	**<0.001**	**<0.001**	0.097	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.111	**<0.001**	**0.001**
1.5 T-1.5 T	DP (n = 35)	23.4 (11.9)	19.8 (21.5)	5.2 (5.5)	1495.1 (120.9)	−5.0 (3.4)	−1.3 (0.8)	20.9 (21.0)	4.9 (4.3)	49.5 (18.3)	17.2 (21.3)	3.3 (4.9)
	Stable (n = 103)	23.1 (13.4)	6.4 (14.1)	1.4 (8.1)	1421.5 (105.4)	−3.9 (2.6)	−0.91 (1.3)	10.1 (11.1)	2.8 (3.7)	46.7 (20.3)	10.5 (11.4)	2.7 (3.6)
	Cohen's d	0.02	0.74	0.55	0.65	0.36	0.36	0.64	0.52	0.14	0.39	0.14
	P – value	0.893	**0.001***	**0.011**	0.225	**0.038**	0.092	**<0.001***	**0.006**	0.489	0.098*	0.455
1.5 T-3.0 T	DP (n = 53)	22.4 (12.7)	29.3 (25.2)	6.0 (9.3)	1513.2 (114.9)	−5.4 (3.4)	−0.97 (1.8)	31.1 (28.2)	6.6 (9.0)	47.6 (20.9)	14.7 (20.6)	3.3 (6.5)
	Stable (n = 157)	23.3 (14.7)	9.1 (13.5)	2.4 (3.9)	1527.9 (102.2)	−3.8 (2.6)	−0.74 (1.8)	12.4 (14.5)	2.5 (2.9)	46.8 (20.3)	3.3 (13.5)	0.5 (2.9)
	Cohen's d	0.07	0.99	0.51	0.14	0.53	0.13	0.83	0.61	0.04	0.65	0.56
	P – value	0.702	**<0.001***	**0.007***	0.392	**0.002***	0.227	**<0.001***	**0.002***	0.806	**<0.001***	**0.004***
3.0 T-1.5 T	DP (n = 14)	27.2 (20.7)	4.9 (12.2)	2.4 (7.9)	1521.4 (131.7)	−3.5 (2.2)	−0.95 (0.59)	7.6 (11.3)	2.7 (4.6)	51.8 (28.7)	10.9 (9.6)	4.8 (5.5)
	Stable (n = 35)	18.6 (8.1)	2.1 (9.7)	−0.4 (3.9)	1560.2 (75.2)	−2.9 (2.1)	−1.0 (1.1)	6.3 (7.9)	1.5 (2.1)	35.9 (12.7)	12.3 (9.3)	3.9 (4.5)
	Cohen's d	0.55	0.25	0.45	0.36	0.28	0.06	0.13	0.34	0.72	0.15	0.18
	P – value	0.155*	0.42	0.107	0.315*	0.435	0.779	0.655	0.36*	0.064*	0.639	0.589
3.0 T-3.0 T	DP (n = 47)	23.6 (10.5)	18.4 (21.8)	3.4 (6.6)	1541.7 (71.6)	−3.7 (3.1)	−0.82 (1.1)	21.6 (22.9)	4.4 (4.6)	45.7 (16.7)	15.6 (20.4)	3.1 (4.0)
	Stable (n = 168)	21.2 (10.9)	10.1 (12.4)	3.0 (5.1)	1556.7 (92.0)	−2.5 (2.4)	−0.71 (0.84)	10.7 (14.6)	2.9 (3.8)	40.8 (16.2)	9.3 (15.4)	2.4 (4.7)
	Cohen's d	0.22	0.47	0.07	0.18	0.43	0.11	0.57	0.36	0.29	0.35	0.16
	P - value	0.193	**0.015***	0.67	0.251*	**0.005**	0.483	**0.004**	**0.037**	0.085	**0.028**	0.43

LVV – lateral ventricular volume, SIENA – Structural Image Evaluation, using Normalisation, of Atrophy, WBV – whole brain volume, PBVC – percent brain volume change, DP – disease progression, VIENA – ventricular extension, vCSF – ventricular cerebral spinal fluid.
Student's *t*-test was performed unless otherwise specified. *- due to heteroskedasticity (Levene's test < 0.05), Brown-Forsythe test of equality was performed instead. P-value lower than 0.05 was considered statistically significant and shown in bold. The demographic and clinical characteristics of the study subpopulation with both NeuroSTREAM and SIENA analyses is shown in Supplement Table 1.
[a] Age-adjusted analyses of covariance (ANCOVA) with Bonferroni-adjusted post hoc comparisons were performed. All measures are shown as mean (standard deviation). Baseline NeuroSTREAM LVV, SIENAX-based WBV and vCSF are shown in milliliters (mL) whereas the longitudinal changes are shown in %.

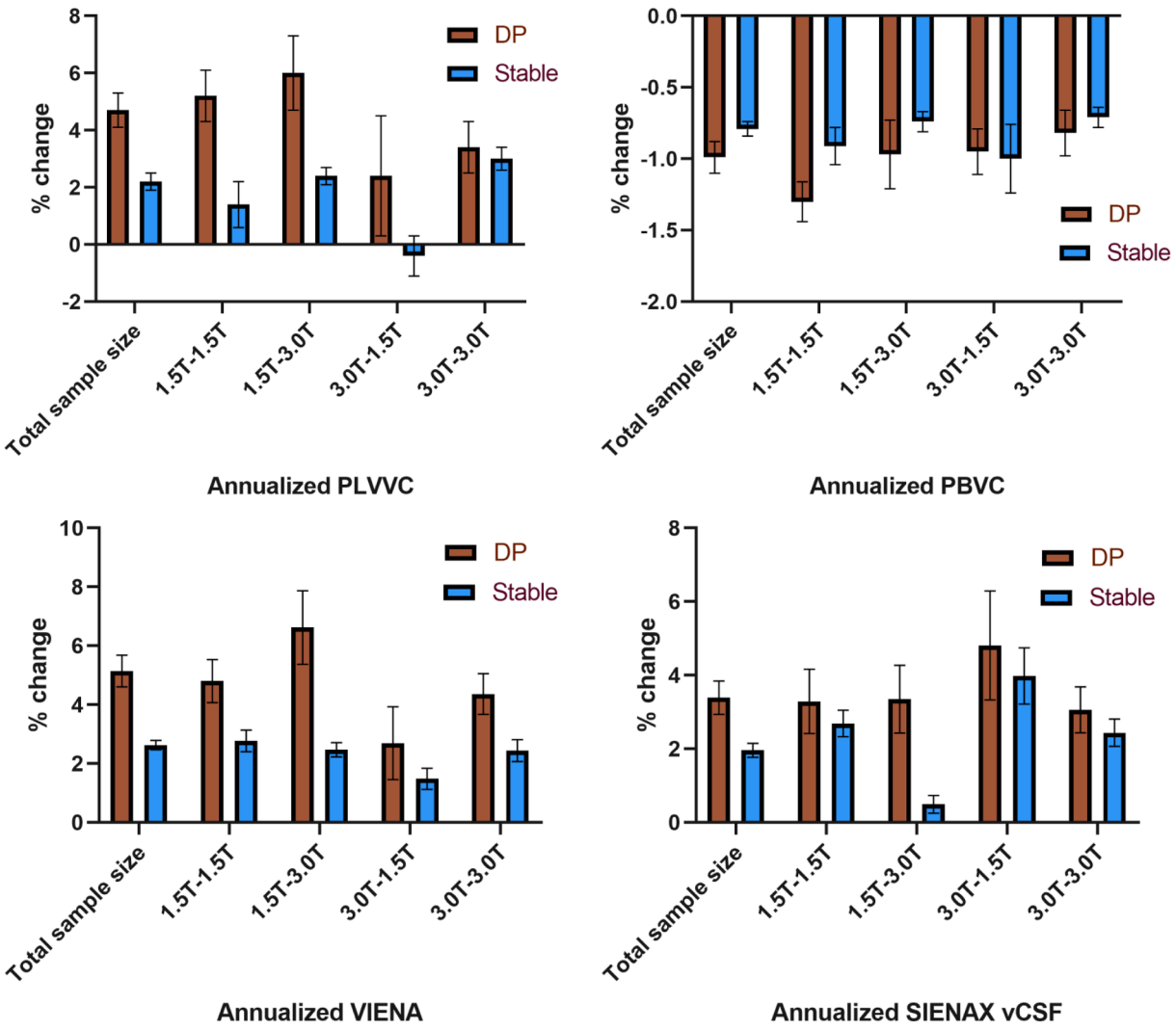

Fig. 2. Bar plot demonstrating NeuroSTREAM-based PLVVC, SIENAX-based vCSF % change, VIENA-based vCSF % change, and SIENA-based PBV changes in the study population. PLVVC – percent lateral ventricular volume change, vCSF – ventricular cerebrospinal fluid, SIENA – Structural Image Evaluation, using Normalisation, of Atrophy, PBV – percent brain volume. Error bars represent standard error of mean.

center, longitudinal, US-based, MRI study showed that up to 99.2% of the scanned PwMS had a viable T2 FLAIR sequence, compared to 72% with 2D T1-WI and only 28.2% with a 3D T1-WI image (Zivadinov et al., 2018). Furthermore, the majority of baseline scans (>70%) were acquired on 1.5 T field strength, whereas greater than 50% were on 3.0 T at the follow-up visit. As phase 3 MS trials include multiple sites from different countries, it is important to implement protocols that allow reliable monitoring of neurodegeneration despite scanner and sequence parameter changes.

Assessment of whole brain atrophy on 3D T1-WI images remains the gold standard for primary and secondary neurodegenerative trial outcomes, and should continue to be recommended outside of trials whenever feasible. However, substantial *a priori* sequence harmonization is need and scanner parameters must remain unchanged through the span of the study. The feasibility and reproducibility of such practices have been recently tested by the North American Imaging in MS (NAIMS) Cooperative where a single MS patient was scanned at 7 different university sites using a uniform 3 T MRI protocol (Shinohara et al., 2017). Even though all scanners were from the same manufacturer (Siemens Skyra, Tim Trio or Verio), the protocol harmonization still showed significant biases and high between-site variation in measured volumes of both whole brain and white and gray matter structures (Shinohara et al., 2017). A similar study using 8 healthy volunteers that were scanned at 8 different sites across North America showed similar findings with imaging site variation of up to 12% (Cannon et al., 2014). They further utilized an Alzheimer's Disease Neuroimaging Initiative (ADNI)-based structural phantom and showed that scanning platforms have significant differences in gradient nonlinearities, inhomogeneity artifacts, and scaling procedures (Cannon et al., 2014). Only after the scans were corrected based on individual phantom-derived gradient non-linearity coefficients did the between-scanner interclass correlation coefficient improve to over 0.987 (Cannon et al., 2014). Such approaches are unlikely to be feasible in clinical routine study. Taken together, due to the aforementioned biological and technological confounding factors, the Magnetic Resonance Imaging in MS (MAGNIMS) group still does not recommend use of brain atrophy in its consensus guidelines (Rovira et al., 2015). Similarly, the Consortium of MS Centers (CMSC) Task Force for standardized MRI protocol only recommends qualitative assessment of the brain volume changes (visual comparison with previous scans) (Traboulsee et al., 2016).

In addition to SIENA, we compared the performance of NeuroSTREAM-based LVV with two more 3D T1-WI analyses that measure the size and change in the vCSF (longitudinal algorithm with VIENA and difference between cross-sectional measures with SIENAX). The presence of DP was significantly associated with greater VIENA-based vCSF % changes and this analysis was not significantly affected

Table 4
The effect of scanner changes on longitudinal brain atrophy measures.

NeuroSTREAM-based PLVVC	Beta	SE	t-statistics	95% CI LB	95% CI UB	p-value
Intercept	−3.54	3.38	−1.045	−10.2	3.1	0.296
Change in MRI strength [No – reference]	−0.09	2.13	−0.04	−4.26	4.1	0.968
Sex [Female - reference]	−1.74	1.08	−1.604	−3.86	0.39	0.387
DP [No – reference]	9.5	2.43	3.915	4.74	14.27	**<0.001**
Age at baseline	−0.02	0.04	−0.53	−0.11	0.06	0.596
Follow-up time	5.57	0.44	12.686	4.71	6.43	**<0.001**
DP and time interaction	−4.48	0.45	−9.962	−5.36	−3.59	**<0.001**
Change in MRI strength and time interaction	−0.01	0.39	−0.033	−0.78	0.76	0.973
SIENA-based % WBV change						
Intercept	0.318	0.66	0.484	−0.97	1.61	0.629
Change in MRI strength [No – reference]	0.64	0.4	1.586	−0.15	1.43	0.113
Sex [Female - reference]	−0.25	0.21	−1.175	−0.67	0.17	0.24
DP [No – reference]	−0.24	0.45	−0.531	−1.12	0.64	0.596
Age at baseline	−0.2	0.01	−1.926	−0.03	0.003	0.055
Follow-up time	−0.87	0.085	−10.17	−1.03	−0.69	**<0.001**
DP and time interaction	0.33	0.087	3.736	0.16	0.49	**<0.001**
Change in MRI strength and time interaction	−0.16	0.077	−2.133	−0.32	−0.01	**0.033**
VIENA-based vCSF % change						
Intercept	−4.23	4.52	−0.936	−13.09	4.64	0.35
Change in MRI strength [No – reference]	−0.92	2.75	−0.335	−6.31	4.48	0.738
Sex [Female - reference]	−4.35	1.46	−2.972	−7.23	−1.48	**0.003**
DP [No – reference]	6.35	3.09	2.058	0.29	12.4	**0.04**
Age at baseline	0.1	0.06	1.749	−0.01	0.22	0.081
Follow-up time	5.69	0.58	9.786	4.55	6.84	**<0.001**
DP and time interaction	−4.13	0.59	−6.883	−5.3	−2.95	**<0.001**
Change in MRI strength and time interaction	0.42	0.53	0.789	−0.62	1.44	0.43
SIENAX-based vCSF % change						
Intercept	2.79	4.34	0.642	−5.75	11.32	0.398
Change in MRI strength [No – reference]	−3.84	2.67	−1.44	−9.08	1.39	0.15
Sex [Female - reference]	−1.96	1.43	−1.372	−4.76	0.85	0.171
DP [No – reference]	5.49	3.02	1.823	−0.43	11.4	0.069
Age at baseline	−0.06	0.06	−1.114	−0.17	0.05	0.266
Follow-up time	3.17	0.57	5.525	2.04	4.29	**<0.001**
DP and time interaction	−2.95	0.59	−4.986	−4.11	−1.79	**<0.001**
Change in MRI strength and time interaction	2.08	0.51	4.084	1.08	3.09	**<0.001**

PLVVC – percent lateral ventricular volume change, vCSF – ventricular cerebral spinal fluid, WBV – whole brain volume, DP – disease progression, SE – standard error, CI – confidence interval, LB – lower bound, UB – upper bound. Bold values denote statistical significance at the $p < 0.05$ level.

by the change in MRI field strength. However, of particular mention is that the quality of scans and failure rates between VIENA and NeuroSTREAM-based LVV introduces a clear selection bias favoring VIENA. The comparable performance is between the 62.4% successful 3D T1-WI-based analysis when compared to 100% of instances with NeuroSTREAM. As previously mentioned, similar limitations apply to SIENA as well. This analysis corroborates previous findings where NeuroSTREAM was shown to perform comparably to well-validated measures such as VIENA, despite being analyzed on 300% coarser voxels (3 mm vs. 1 mm) (Dwyer et al., 2017). On the other hand, the SIENAX-based vCSF % change was significantly affected by the change in MRI field strength with a more than 70% opposite effect. These differences can be attributed to greater variability in cross-sectional vCSF quantification, likely amplified by the greater surface area of the 3rd and 4th ventricle. Lastly, the choroid plexus is visible on the 3D T1-WI (isointense to the brain tissue) and may interfere with automated ventricle segmentation that does not explicitly control for it.

A limitation of the current study is that it is based on data from a single center where both 1.5 T and 3.0 T scanners were from the same manufacturer (GE). To better understand overall real-world feasibility, future studies should investigate NeuroSTREAM-based LVV measures in a multi-centric study that employ different scanner manufacturers and sequence parameters, such as sagittal versus axial FLAIR acquisition, differences in the size of the acquisition matrix ($1 \times 1 \times 3$ vs. $1 \times 1 \times 1$ mm^3), and 3D versus 2D FLAIR images. In a previous investigation, NeuroSTREAM-based LVV quantification on 3 mm, 5 mm and 7 mm thick scans remained consistent with intraclass correlation coefficients of 0.99 (Dwyer et al., 2017). The coefficient of variance ranged from 3.15% for 3 mm thick scans to 4.22% for 7 mm thick scans (Dwyer et al., 2017). Furthermore, the NeuroSTREAM algorithm had similar performance on axial and sagittal scan orientations with coefficients of variation of 3.73% and 3.84%, respectively (Dwyer et al., 2017). In the aforementioned multicenter MS-MRIUS study, NeuroSTREAM-based LVV quantification was also performed on 28 3D T2-FLAIR scans and successfully analyzed together with the 2D T2-FLAIR cases (Zivadinov et al., 2018).

As another limitation to our study, the scanning procedures should preferably be performed for all patients in a back-to-back fashion (scanned on various scanners), at the same time of day. This would allow proper quantification of the scanner effect and minimize biological confounding effects such as hydration status and menstrual cycle. A large multicenter study without any prior sequence harmonization would allow better assessment of the true generalizability of the NeuroSTREAM method. Accumulation of perivascular cerebrovascular pathology in the aging MS population may also present as another confounding factor in the use of LVV-based atrophy measure, and should be explored explicitly (Jakimovski et al., 2019). Furthermore, periventricular pathology and deep gray matter atrophy are MS-specific changes that predominantly occur early in the disease duration. On the other hand, studies have shown that PwMS with progressive disease demonstrate accelerated cortical atrophy which may not be fully captured by this approach (Eijlers et al., 2019). Additional studies should investigate the clinical feasibility of PLVVC in a more disabled MS population, as well as its relation to cognitive decline. Lastly, the results from NeuroSTREAM do not include scaling based on head size, which may be influenced by sex-based differences (Jakimovski et al., 2020).

In conclusion, changes in LVV can differentiate between stable PwMS and PwMS with DP despite change in MRI scanner field strength, and even in a population with mild disability status. Future creation of normative MS databases of age, sex, and disease duration-matched LVV changes may further facilitate the translation of this imaging biomarker to an individual patient level.

CRediT authorship contribution statement

Dejan Jakimovski: Conceptualization, Methodology, Data curation,

Formal analysis, Writing - original draft, Writing - review & editing. **Robert Zivadinov:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition. **Niels Bergsland:** Conceptualization, Methodology, Writing - review & editing, Software. **Deepa P. Ramasamy:** Software, Writing - review & editing. **Jesper Hagemeier:** Formal analysis, Writing - review & editing. **Antonia Valentina Genovese:** Formal analysis, Writing - review & editing. **David Hojnacki:** Data curation, Writing - review & editing. **Bianca Weinstock-Guttman:** Data curation, Writing - review & editing. **Michael G. Dwyer:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2020.102554.

References

Cannon, T.D., Sun, F., McEwen, S.J., Papademetris, X., He, G., van Erp, T.G.M., Jacobson, A., Bearden, C.E., Walker, E., Hu, X., Zhou, L., Seidman, L.J., Thermenos, H.W., Cornblatt, B., Olvet, D.M., Perkins, D., Belger, A., Cadenhead, K., Tsuang, M., Mirzakhanian, H., Addington, J., Frayne, R., Woods, S.W., McGlashan, T.H., Constable, R.T., Qiu, M., Mathalon, D.H., Thompson, P., Toga, A. W., 2014. Reliability of neuroanatomical measurements in a multisite longitudinal study of youth at risk for psychosis: reliability of Multisite, Longitudinal MRI. Hum. Brain Mapp. 35 (5), 2424–2434.

Dwyer, M.G., Silva, D., Bergsland, N., Horakova, D., Ramasamy, D., Durfee, J., Vaneckova, M., Havrdova, E., Zivadinov, R., 2017. Neurological software tool for reliable atrophy measurement (NeuroSTREAM) of the lateral ventricles on clinical-quality T2-FLAIR MRI scans in multiple sclerosis. NeuroImage: Clinical 15, 769–779.

Eijlers, A.J.C., Dekker, I., Steenwijk, M.D., Meijer, K.A., Hulst, H.E., Pouwels, P.J.W., Uitdehaag, B.M.J., Barkhof, F., Vrenken, H., Schoonheim, M.M., Geurts, J.J.G., 2019. Cortical atrophy accelerates as cognitive decline worsens in multiple sclerosis. Neurology 93 (14), e1348–e1359.

Gelineau-Morel, R., Tomassini, V., Jenkinson, M., Johansen-Berg, H., Matthews, P.M., Palace, J., 2012. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. Hum. Brain Mapp. 33 (12), 2802–2814.

Ghione, E., Bergsland, N., Dwyer, M.G., Hagemeier, J., Jakimovski, D., Paunkoski, I., Ramasamy, D.P., Silva, D., Carl, E., Hojnacki, D., Kolb, C., Weinstock-Guttman, B., Zivadinov, R., 2018. Brain atrophy is associated with disability progression in patients with MS followed in a clinical routine. AJNR Am. J. Neuroradiol. 39 (12), 2237–2242.

Ghione, E., Bergsland, N., Dwyer, M.G., Hagemeier, J., Jakimovski, D., Paunkoski, I., et al., 2019. Aging and brain atrophy in multiple sclerosis. J. Neuroimaging 29 (4), 527–535.

Jakimovski, D., Gandhi, S., Paunkoski, I., Bergsland, N., Hagemeier, J., Ramasamy, D.P., Hojnacki, D., Kolb, C., Benedict, R.H.B., Weinstock-Guttman, B., Zivadinov, R., 2019. Hypertension and heart disease are associated with development of brain atrophy in multiple sclerosis: a 5-year longitudinal study. Eur. J. Neurol. 26 (1), 87.

Jakimovski, D., Kuhle, J., Ramanathan, M., Barro, C., Tomic, D., Hagemeier, J., Kropshofer, H., Bergsland, N., Leppert, D., Dwyer, M.G., Michalak, Z., Benedict, R.H. B., Weinstock-Guttman, B., Zivadinov, R., 2019. Serum neurofilament light chain levels associations with gray matter pathology: a 5-year longitudinal study. Ann. Clin. Transl. Neurol. 6 (9), 1757–1770.

Jakimovski, D., Zivadinov, R., Bergsland, N., Ramasamy, D.P., Hagemeier, J., Weinstock-Guttman, B., Kolb, C., Hojnacki, D., Dwyer, M.G., 2020. Sex-specific differences in life span brain volumes in multiple sclerosis. J. Neuroimaging 30 (3), 342–350.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., 2009. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. NeuroImage 46 (1), 177–192.

Kurtzke, J.F., 1983. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 33 (11), 1444–1452.

Leigh, R., Ostuni, J., Pham, D., Goldszal, A., Lewis, B.K., Howard, T., Richert, N., McFarland, H., Frank, J.A., 2002. Estimating cerebral atrophy in multiple sclerosis patients from various MR pulse sequences. Mult. Scler. 8 (5), 420–429.

Lublin, F.D., Reingold, S.C., Cohen, J.A., Cutter, G.R., Sorensen, P.S., Thompson, A.J., Wolinsky, J.S., Balcer, L.J., Banwell, B., Barkhof, F., Bebo, B., Calabresi, P.A., Clanet, M., Comi, G., Fox, R.J., Freedman, M.S., Goodman, A.D., Inglese, M., Kappos, L., Kieseier, B.C., Lincoln, J.A., Lubetzki, C., Miller, A.E., Montalban, X., O'Connor, P.W., Petkau, J., Pozzilli, C., Rudick, R.A., Sormani, M.P., Stuve, O., Waubant, E., Polman, C.H., 2014. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology 83 (3), 278–286.

Nordenskjöld, R., Malmberg, F., Larsson, E.-M., Simmons, A., Brooks, S.J., Lind, L., Ahlström, H., Johansson, L., Kullberg, J., 2013. Intracranial volume estimated with commonly used methods could introduce bias in studies including brain volume measurements. NeuroImage 83, 355–360.

Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F.D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A.J., Waubant, E., Weinshenker, B., Wolinsky, J.S., 2011. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Ann. Neurol. 69 (2), 292–302.

Popescu, V., Battaglini, M., Hoogstrate, W.S., Verfaillie, S.C.J., Sluimer, I.C., van Schijndel, R.A., van Dijk, B.W., Cover, K.S., Knol, D.L., Jenkinson, M., Barkhof, F., de Stefano, N., Vrenken, H., 2012. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. NeuroImage 61 (4), 1484–1494.

Rocca, M.A., Battaglini, M., Benedict, R.H.B., De Stefano, N., Geurts, J.J.G., Henry, R.G., Horsfield, M.A., Jenkinson, M., Pagani, E., Filippi, M., 2017. Brain MRI atrophy quantification in MS: from methods to clinical application. Neurology 88 (4), 403–413.

Rovira, À., Wattjes, M.P., Tintoré, M., Tur, C., Yousry, T.A., Sormani, M.P., De Stefano, N., Filippi, M., Auger, C., Rocca, M.A., Barkhof, F., Fazekas, F., Kappos, L., Polman, C., Miller, D., Montalban, X., 2015. MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. Nat. Rev. Neurol. 11 (8), 471–482.

Shinohara, R.T., Oh, J., Nair, G., Calabresi, P.A., Davatzikos, C., Doshi, J., Henry, R.G., Kim, G., Linn, K.A., Papinutto, N., Pelletier, D., Pham, D.L., Reich, D.S., Rooney, W., Roy, S., Stern, W., Tummala, S., Yousuf, F., Zhu, A., Sicotte, N.L., Bakshi, R., 2017. Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. AJNR Am. J. Neuroradiol. 38 (8), 1501–1509.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17 (1), 479–489.

Thompson, A.J., Baranzini, S.E., Geurts, J., Hemmer, B., Ciccarelli, O., 2018. Multiple sclerosis. Lancet 391 (10130), 1622–1636.

Traboulsee, A., Simon, J.H., Stone, L., Fisher, E., Jones, D.E., Malhotra, A., Newsome, S. D., Oh, J., Reich, D.S., Richert, N., Rammohan, K., Khan, O., Radue, E.-W., Ford, C., Halper, J., Li, D., 2016. Revised recommendations of the consortium of MS centers task force for a standardized MRI protocol and clinical guidelines for the diagnosis and follow-up of multiple sclerosis. AJNR Am. J. Neuroradiol. 37 (3), 394–401.

Vrenken, H., Jenkinson, M., Horsfield, M.A., Battaglini, M., van Schijndel, R.A., Rostrup, E., Geurts, J.J.G., Fisher, E., Zijdenbos, A., Ashburner, J., Miller, D.H., Filippi, M., Fazekas, F., Rovaris, M., Rovira, A., Barkhof, F., de Stefano, N., 2013. Recommendations to improve imaging and analysis of brain lesion load and atrophy in longitudinal studies of multiple sclerosis. J. Neurol. 260 (10), 2458–2471.

Zivadinov, R., Havrdová, E., Bergsland, N., Tyblova, M., Hagemeier, J., Seidl, Z., Dwyer, M.G., Vaneckova, M., Krasensky, J., Carl, E., Kalincik, T., Horáková, D., 2013. Thalamic atrophy is associated with development of clinically definite multiple sclerosis. Radiology 268 (3), 831–841.

Zivadinov, R., Jakimovski, D., Gandhi, S., Ahmed, R., Dwyer, M.G., Horakova, D., Weinstock-Guttman, B., Benedict, R.R.H., Vaneckova, M., Barnett, M., Bergsland, N., 2016. Clinical relevance of brain atrophy assessment in multiple sclerosis. Implications for its use in a clinical routine. Expert Rev. Neurother. 16 (7), 777–793.

Zivadinov, R., Bergsland, N., Korn, J.R., Dwyer, M.G., Khan, N., Medin, J., Price, J.C., Weinstock-Guttman, B., Silva, D., 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 (2), 289–295.