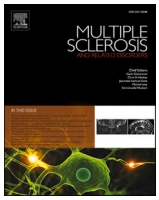
Multiple Sclerosis and Related Disorders 92 (2024) 106148

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

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



Original article

Utility of icobrain for brain volumetry in multiple sclerosis clinical practice

Ai-Lan Nguyen [a,b,*], Maria Pia Sormani [c], Dana Horakova [d], Eva H Havrdova [d], Michael H Barnett [e], Nicola De Stefano [f], Marco Battaglini [f], Manuela Vaneckova [g], Elaine Lui [h,i], Frank Gaillard [h,i], Patricia M Desmond [h,i], Hayden Prime [j], Mineesh Datta [j], Anneke Van der Walt [k], Vilija G Jokubaitis [k], Femke Podevyn [l], Robert Zivadinov [m], Bianca Weinstock-Guttman [m], Marie B D'hooghe [n], Guy Nagels [l,o,p], Vincent Van Pesch [q], Guy Laureys [r], Liesbeth Van Hijfte [r], Jeannette Lechner-Scott [s,t,u], Francesco Patti [v], Edgardo Cristiano [w], Juan I Rojas [w], Diana M Sima [l], Wim Van Hecke [l], Tomas Kalincik [a,b,1], Helmut Butzkueven [k,1]

[a] Neuroimmunology Centre, Royal Melbourne Hospital, University of Melbourne, Melbourne, Australia
[b] CORe, Department of Medicine, University of Melbourne, Melbourne, Australia
[c] Department of Health Sciences (DISSAL), Biostatistics Unit, University of Genoa, Genoa, Italy
[d] Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czechia
[e] Brain and Mind Centre, University of Sydney, Sydney, Australia
[f] Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy
[g] Department of Radiology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czechia
[h] Department of Medical Imaging, Royal Melbourne Hospital, Melbourne, Australia
[i] Department of Radiology, University of Melbourne, Melbourne, Australia
[j] Department of Radiology, Box Hill Hospital, Melbourne, Australia
[k] Department of Neuroscience, School of Translational Medicine, Monash University, Melbourne, Australia
[l] Icometrix, Research and Development, Leuven, Belgium
[m] Department of Neurology, State University of New York at Buffalo, Buffalo, United States
[n] National MS Center Melsbroek, Melsbroek, Belgium
[o] AI supported modelling in clinical sciences (AIMS), Vrije Universiteit Brussel, Brussels, Belgium
[p] Department of Engineering, University of Oxford, Oxford, United Kingdom
[q] Department of Neurology, Université Catholique de Louvain, Brussels, Belgium
[r] Department of Neurology, Ghent University Hospital, Gent, Belgium
[s] Department of Neurology, John Hunter Hospital, Newcastle, Australia
[t] School of Medicine and Public Health, University of Newcastle, Newcastle, Australia
[u] Centre for Brain and Mental Health, Hunter Medical Research Institute, Newcastle, Australia
[v] Department "GF Ingrassia", Section of Neurosciences, University of Catania, Catania, Italy
[w] Centro de Esclerosis Múltiple de Buenos Aires (CEMBA), Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Brain
Volumetry
Variability
Icometrix
Real-world

ABSTRACT

Background: Few studies on multiple sclerosis (MS) have explored the variability of percentage brain volume change (PBVC) measurements obtained from different clinical MRIs. In a retrospective multicentre cohort study, we quantified the variability of annualised PBVC in clinical MRIs.
Methods: Clinical MRIs of relapse-onset MS patients were assessed by icobrain. Volumetric data were analysed on same-scanner and different-scanner MRI pairs if they passed quality control criteria. Alignment similarity between two images had to be comparable to same-scanner scan-rescan images.
Results: Of 6826 MRIs, 85 % had appropriate volumetric sequences and 4446 serial MRI pairs were analysed. 3334 (75 %) MRI pairs from 1207 patients met the inclusions. The PBVC of included MRI pairs showed variance of 0.78 % for same-scanner pairs and 0.80 % for different-scanner pairs. Further selection of included MRI pairs

* Corresponding author at: Department of Neurology, Royal Melbourne Hospital, Parkville 3050, Australia.
E-mail address: ai-lan.nguyen@unimelb.edu.au (A.-L. Nguyen).
[1] authors contributed equally to the manuscript

https://doi.org/10.1016/j.msard.2024.106148
Received 16 August 2024; Received in revised form 24 October 2024; Accepted 27 October 2024
Available online 8 November 2024

with the best variance resulted in 1885 (42 %) MRI pairs with PBVC variance of 0.34 %. Excluded MRI pairs with poor alignment similarity had variances of 2.97 % for same-scanner pairs and 20.79 % for different-scanner pairs. *Conclusion:* Icobrain should be utilised for PBVC determination only on selected MRIs with the best alignment similarity. Applying strict selection criteria for the included MRI pairs and longitudinal imaging on the same scanner remain mandatory to reduce PBVC variability.

1. Introduction

Excessive brain volume loss in MS is a biomarker of disease progression (Popescu et al., 2013; Uher et al., 2015) and an indication of insufficient treatment response (Kappos et al., 2016). Unfortunately, serial brain volumetry in clinical MS practice is not commonly unavailable. There are logistic and technical limitations of measuring brain tissue loss, including but not limited to: cost, scan-rescan variability, short-term effect of immunotherapy (such as pseudoatrophy, resolution of oedema with steroid therapy), variability in normal aging, comorbidities and hydration state (Sastre-Garriga et al., 2020).

An important factor is the broad range of postprocessing algorithms, with no algorithm presently recognised as the gold standard. The most widely used software is SIENA (Structural Image Evaluation using Normalisation of Atrophy), an automated, registration-based software that estimates longitudinal percentage brain volume change (PBVC) (Smith et al., 2001; Smith et al., 2002). However, despite being fully automated, SIENA still requires a small degree of trained user intervention which may not be possible at many centres (Sastre-Garriga et al., 2020; Giorgio and De Stefano, 2013).

Icometrix is a commercial provider that developed icobrain, a software for volumetric MRI analysis (Jain et al., 2015; Smeets et al., 2016; Steenwijk et al., 2017). It offers a centralised analysis with data analysed remotely, similar to SIENA for clinical trials. However, icobrain's performance in clinical MS practice has not been widely assessed (Beadnall et al., 2019; D'hooghe et al., 2019; Fragoso et al., 2017). Additionally, patients are often serially scanned on different MRI scanners in real-world settings, and few MS studies have explored the use of longitudinal brain volumetry using different scanners (Durand-Dubief et al., 2012; Biberacher et al., 2016; Zivadinov et al., 2017; Zivadinov et al., 2018; Storelli et al., 2018).

In this study, we tested the utility of icobrain in real-world settings. The goal was to 1) quantify the variability of annualised PBVC in clinically acquired MRIs from relapse-onset MS patients and 2) explore the demographic and clinical characteristics that predict PBVC in this cohort.

2. Methods

2.1. Participants

Our retrospectively analysed cohort of patients attended 4 centres in 2 countries and underwent brain MRIs as part of their routine clinical care. The sites were: General University Hospital (Prague, Czech Republic), Royal Melbourne Hospital (Melbourne, Australia), Box Hill Hospital (Melbourne, Australia) and Brain and Mind Centre (Sydney, Australia). These centres all document clinical outcomes using the MSBase Registry, an international MS registry (Butzkueven et al., 2006).

Eligible patients were enrolled from December 18, 2017 to October 13, 2018. Inclusion criteria were age $\geq$18 years, diagnosis with relapse-onset MS, known date of MS onset and consent to providing access to a stored minimum dataset in the MSBase Registry (date of birth, sex, minimum of one Expanded Disability Status Scale assessment) (Butzkueven et al., 2006). Participants were excluded if there were no suitable brain MRI scans stored in their hospital MRI depository.

Ethics approval was granted at each site's institutional review board and informed consent was obtained from each patient. MSBase (registered with WHO International Clinical Trials Registry Platform ID ACTRN12605000455662) was approved by the Melbourne Health Human Research Ethics Committee and by the local ethics committees in participating centres.

2.2. Data processing

Brain MRI scans were performed as part of routine clinical care and patients were scanned on combinations of MRIs at their site (Table S1). The minimum inclusion criteria for cerebral MRI scans were 3D T1-weighted (T1) sequence (maximum slice thickness of 1.6 mm) and 3D flair-attenuated inversion recovery (FLAIR) sequence (or 2D FLAIR with maximum slice thickness of 3 mm). Scans missing the appropriate sequences were labelled 'Unusable' (Fig. 1).

All MRIs were processed using the icobrain software version 3.1. Usable images initially undergo cross-sectional processing (icobrain cross). The icobrain cross pipeline results in segmentation and quantitative measurements for whole brain volume (BV) and grey matter volume (GMV) normalised for head size, FLAIR lesion volume and T1 hypointense lesion volume.

For longitudinal analyses, consecutive MRI scans were selected. The longitudinal processing (icobrain long) provides measurements for annualised PBVC and absolute change in FLAIR lesion volume. PBVC used a registration-based approach applying Jacobian integration (Smeets et al., 2016).

Both icobrain cross and icobrain long pipelines undergo automated quality control (QC) $\pm$ manual/visual QC analysis. The automated QC can flag datasets that need to be checked by an experienced assessor, leading to a visual or manual QC. The results of the automatic and visual quality analysis are then recorded as a 'QC status', accompanied by potential explanations in the form of 'quality remarks.' The QC status labels are: 1) Approved (data usable and no remarks given), 2) Approved with remarks (the remarks contain information on which measurements are considered reliable or not) and 3) Rejected (data not usable due to bad quality). The quality remarks provide additional information about images, processing and registration issues and are grouped into 'consequences.' The consequences are listed in Fig. S2.

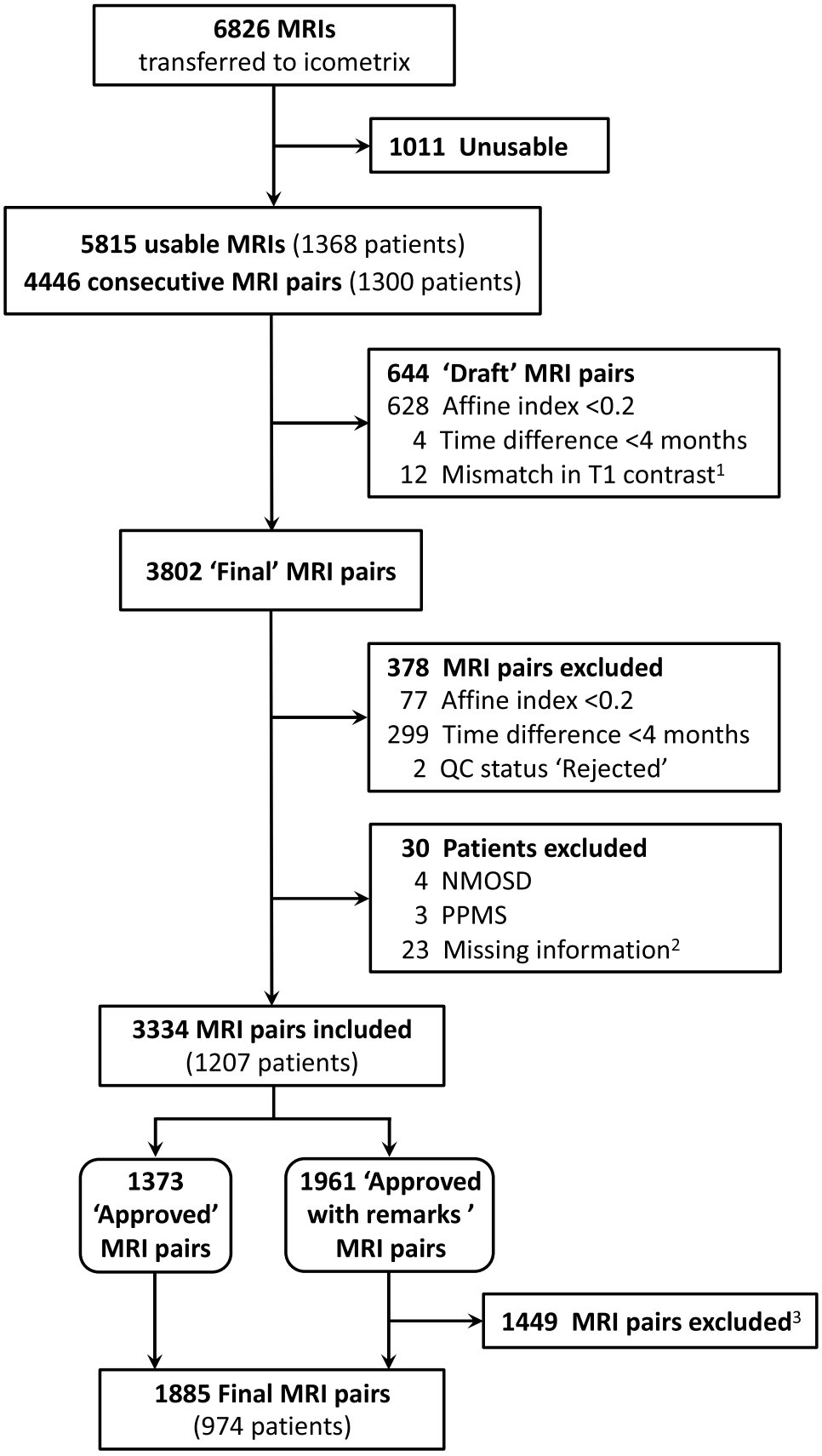
Reports are finally labelled as 'Draft' or 'Final' based on whether the data are considered unacceptable or acceptable for use. Longitudinal scan pairs are labelled 'Draft' if there is a mismatch in the T1 images (precontrast vs postcontrast), if the time difference between MRI pairs (scan interval) is <4 months (122 days), or if the affine similarity index is <0.2 (Fig. 1).

2.3. Affine similarity index

The automated QC includes computing an 'affine similarity index' that measures the similarity between two scans after being affinely transformed to each other (Studholme et al., 1999). This affine similarity index is proportional to image similarity. The affine similarity index is the average between 2 normalised mutual information (NMI) values obtained from T1 images after performing affine registrations between the two images in two directions (timepoint 1 to timepoint 2 and vice versa) (Studholme et al., 1999). NMI is calculated in each direction as:

$$\mathrm{NMI}(A,B) = \frac{H(A)+H(B)}{H(A,B)}$$

where *H(A)* and *H(B)* indicate entropy of the two T1 images and *H(A,B)* indicates joint entropy after affine co-registration and resampling in one



(caption on next page)

Fig. 1. Consort chart for MRI inclusion
Abbreviations: MRI, magnetic resonance imaging; NMOSD, neuromyelitis optica spectrum disorder; PPMS, primary progressive MS; QC, quality control.
[1] Mismatch in T1 pre-contrast or T1 post-contrast
[2] Includes missing date of RRMS onset, MS subtype
[3] 'Approved with remarks' pairs excluded if quality remark or 'consequence' (Figure S2) present.

of the directions. The mean value of NMI(A,B) and NMI(B,A) indicates the final affine similarity index.

Volumetric information was included in further analyses if affine similarity index was ≥0.2. This threshold was selected on the basis of a scan-rescan study (Sima et al., 2019) where intrascanner and interscanner MRI scan-rescan pairs were available from 60 MS patients. It was observed that the affine similarity index for intrascanner comparisons was always above 0.2, and that the PBVC measurement error for interscanner comparisons with affine similarity index above 0.2 did not exceed the measurement error for intrascanner comparisons.

2.4. Study design and outcome measures

MRI pairs were created from consecutive MRIs of individual patients. They were included if they had suitable sequences, were classified as 'Final', had affine index ≥0.2, scan interval ≥4 months, and a QC status of 'Approved' or 'Approved with remarks'. Intrascanner and interscanner pairs were defined as consecutive scans from the same MRI or different MRIs respectively. MRI-1 and MRI-2 refer to the first and second MRI in each MRI pair. A single MRI could be included in up to two MRI pairs, as MRI-1 of one pair and MRI-2 of an earlier pair.

Patient demographic and clinical information were obtained from the MSBase registry, including MS course, Expanded Disability Status Scale (EDSS) scores, relapse dates, dates of corticosteroid use, type and duration of disease modifying therapy (DMT). MRI-1 was considered the baseline for the acquisition of visit course, EDSS, BV, GMV, FLAIR and T1 lesion volumes. Patient age and disease duration were taken as the midpoint between MRI-1 and MRI-2.

DMTs were classified as low-efficacy (glatiramer acetate, interferon beta, teriflunomide) or high-efficacy (alemtuzumab, autologous stem-cell transplant, daclizumab, dimethyl-fumarate, fingolimod, natalizumab, ocrelizumab, rituximab). The primary outcomes were 1) variance of PBVC, and 2) PBVC between MRI-1 and MRI-2.

2.5. Statistical analysis

The database was locked for analysis on November 1, 2018. All PBVC values reported were annualised. The variance of PBVC was compared between the MRI groups as follows: intrascanner vs interscanner MRI pairs, affine index ≥0.2 vs <0.2, Approved vs Approved with remarks, as well as the different 'consequence' statement groups (Fig. S2).

Mixed effects models were used to identify the variables that predict PBVC; with demographic, MRI and disease factors as fixed effects and random intercept for subject. The variables tested included: sex, age at MRI, age at MS diagnosis, MS disease duration, site, scan interval, MRI field strength, MRI vendor, icobrain software version, intrascanner vs interscanner MRI pair, BV, GMV, FLAIR lesion volume, T1 hypointense lesion volume, FLAIR lesion volume change, relapse number during MRI pair, relapse in month pre MRI-1 & 2, corticosteroid use in month pre MRI-1 & 2, DMT use during MRI pair, duration of time on current DMT pre MRI-1, proportion of time spent on DMT during MRI pair (high and low efficacy DMT), EDSS and MS course.

Variables associated with PBVC at α=0.05 were included in a multivariable mixed effects model to obtain predictors of PBVC. R^2 was calculated, representing the proportion of total variance explained by the multivariable model. All statistical analyses were performed in R, version 3.6.0.

3. Results

3.1. Cohort characteristics

A total of 6826 MRIs were transferred to icometrix, and 5815 (85 %) contained suitable sequences. 18 different MRI machines were used (Table S1). There were 4446 consecutive MRI pairs from 1300 patients: 3536 (80 %) intrascanner pairs and 910 (20 %) interscanner pairs. In total, 3334 (75 %) MRI pairs, containing 4830 unique scans from 1207 patients met the inclusion criteria (Fig. 1). Of these, 2976 (89 %) were intrascanner pairs and 358 (11 %) were interscanner pairs. Thus, from the 4446 consecutive pairs, 84 % (2976/3536) intrascanner pairs and 39 % (358/910) interscanner pairs were included. Of the 3334 included MRI pairs, there was a higher proportion of scans having a QC status of 'Approved with remarks' (59 %) compared to 'Approved' (41 %) (Fig. 1).

The demographic and clinical characteristics of the included vs excluded cohort are summarised in Table 1. A key difference was the proportion of intrascanner and interscanner pairs, with a much higher proportion of interscanner pairs in the excluded group (50 %). In the included cohort, the mean (SD) age at MS diagnosis was 32.6 (9.7) years and the median (IQR) MS disease duration was 6.8 years (2.8, 11.6). Most patients (94 %) had relapsing-remitting MS, with a median (IQR) EDSS of 2.5 (1.5, 4). The cohort was almost all treated (93 %), predominantly with high-efficacy DMTs, and had been on treatment for a median (IQR) of 1.6 years (0.4, 2.8) prior to MRI-1. The median (IQR; range) scan interval was 0.9 years (0.6, 1.1; 0.3–4.9).

3.2. PBVC variability

3.2.1. Intrascanner vs interscanner MRI pairs

The mean (SD) annualised PBVC for the 3334 included MRI pairs was −0.21 % (0.88) and variance was 0.78 %. The variance of PBVC was similar between included intrascanner pairs (0.78 %) and interscanner pairs (0.80 %) (Fig. 2). As these pairs had been filtered for affine index ≥0.2 and the majority of interscanner pairs (61 %) were from the excluded group, the PBVC of pairs with affine index <0.2 was also assessed (while retaining the other inclusion criteria). Of the 662 MRI pairs with affine index <0.2, the PBVC variance was 2.97 % in the intrascanner group and 20.79 % in the interscanner group (Fig. 2). This corresponds to a minimum detectable effect size of 0.70 % annualised PBVC in the intrascanner group and 1.84 % annualised PBVC in the interscanner group (for a hypothetical cohort of 100 patients, power 0.95 and $\alpha = 0.05$).

3.2.2. QC status & 'consequences'

We assessed variance of PBVC of the included 3334 MRI pairs stratified by QC status. The 'Approved' pairs had a PBVC variance of 0.35 %, and the 'Approved with remarks' pairs had a variance of 1.08 % (Fig. 3). The 'Approved with remarks' group was further divided by the 'consequences' or statements (Fig. S2). MRI pairs labelled with ≥1 'consequence' were excluded and a subgroup with no statement or comment was included in the analysis (Table S3). The 512 MRI pairs with no comments had the smallest variance, 0.29 %, of all the 'consequences' (where $n \geq 10$) and were comparable with the variance of the 'Approved' group (Table S3).

We therefore combined the 'Approved' and 'Approved with remarks – no comment' groups to create a cohort of 1885 MRI pairs (3051 unique MRIs) from 974 patients which had the lowest PBVC variance (Fig. 1). Of

Table 1
Demographic & clinical characteristics of patient cohort.

CHARACTERISTICS	VALUES OF *INCLUDED* MRI PAIRS[a]	VALUES OF *EXCLUDED* MRI PAIRS[b]
DEMOGRAPHIC		
Number of participants	1207	686
Sex, F (%)	910 (74 %)	488 (71 %)
Age, y, mean (SD)		
at CIS diagnosis	30.2 (9.6)	29.9 (9.6)
at MS diagnosis	32.6 (9.7)	32.3 (9.7)
at MRI[c]	40.7 (10.8)	40.9 (10.6)
MS disease duration[d], y, mean/ median (IQR)	8.1; 6.8 (2.8, 11.6)	8.7; 7.2 (3.8, 12.3)
MRI PAIRS		
Number of MRI pairs		
Total	3334	1112
Per participant, median (IQR), range	3 (2, 4), 1–10	1 (1, 2), 1–11
Scan interval between MRI pairs, y, median (IQR)	0.9 (0.6, 1.1)	0.7 (0.3, 1.1)
Scanner differences, n (%)		
Interscanner pair (different scanners)	358 (11 %)	552 (50 %)
Intrascanner pair (same scanner)	2976 (89 %)	560 (50 %)
Field strength differences, n (%)		
1.5T→3T	23 (0.7 %)	388 (35 %)
3T→1.5T	9 (0.3 %)	13 (1 %)
1.5T both	1696 (51 %)	224 (20 %)
3T both	1606 (48 %)	487 (44 %)
Vendor differences, n (%)		
GE both	391 (12 %)	106 (9 %)
Philips both	1059 (32 %)	151 (14 %)
Siemens both	1871 (56 %)	534 (48 %)
Different vendors	13 (0.4 %)	321 (29 %)
CLINICAL (MS)		
EDSS at visit closest to MRI-1[e], median (IQR), range	2.5 (1.5, 4), 0–7.5	2.5 (1.5, 4), 1–8
MS course at visit closest to MRI-1[f], n (%)		
CIS	51 (2 %)	55 (5 %)
RRMS	3116 (94 %)	926 (83 %)
SPMS	150 (4 %)	47 (4 %)
PPMS	0	6 (1 %)
Unknown	0	78 (7 %)
Relapse number during MRI pair[g], n (%)		
0	2787 (84)	947 (85 %)
1	460 (14 %)	133 (12 %)
2	80 (2 %)	29 (3 %)
3	6 (0.2 %)	1
4	1	2
Relapses in month pre MRIs, n (%)		
None	3080 (92 %)	1014 (91 %)
Pre both MRI-1 and MRI-2	7 (0.2 %)	6 (1 %)
Pre MRI-1	147 (4 %)	47 (4 %)
Pre MRI-2	100 (3 %)	45 (4 %)
Corticosteroid use in month pre MRIs, n (%)		
None	3122 (94 %)	1033 (93 %)
Pre both MRI-1 and MRI-2	6 (0.2 %)	2 (0.2 %)
Pre MRI-1	120 (4 %)	37 (3 %)
Pre MRI-2	86 (2 %)	39 (4 %)
DMT use during MRI pair, n (%)	3099 (93 %)	958 (86 %)
Duration of current DMT pre MRI-1, y, median (IQR)	1.6 (0.4, 2.8)	1.4 (0.2, 3.2)
Time spent on DMT during MRI pair,%, median (IQR)		
Any	100 (97, 100)	100 (84, 100)
High-efficacy[h]	100 (54, 100)	100 (0, 100)
Low-efficacy[i]	0 (0, 0)	0 (0, 0)

Abbreviations: CIS, clinically isolated syndrome; DMT, disease modifying therapy; EDSS, expanded disability status scale; F, female; GE, General Electric; IQR, interquartile range; MRI, magnetic resonance imaging; MRI-1, first MRI in pair; MRI-2, second MRI in pair; MS, multiple sclerosis; SD, standard deviation; T, tesla; y, year.

[a] Included MRI pairs: consecutive pairs, 'final', affine index ≥0.2, scan interval ≥4 m, no rejected, no contrast mismatch, no PPMS/NMO.
[b] Excluded MRI pairs: missing corticosteroid and DMT information from 1 patient.
[c] Age at MRI is taken as the midpoint of the MRI pair.
[d] Disease duration at MRI is taken as the midpoint of the MRI pair.
[e] If visit within 12 months of MRI-1. This visit was available for 3194 MRI pairs (140 missing).
[f] Information available for 3317 MRI pairs (17 missing).
[g] Excluding 1 month pre MRI-2.
[h] High-efficacy DMTs include: alemtuzumab, autologous stem-cell transplant, daclizumab, dimethyl-fumarate, fingolimod, natalizumab, ocrelizumab, rituximab.
[i] Low-efficacy DMTs include: glatiramer acetate, interferon beta, teriflunomide.

these 1885 MRI pairs, 1685 (89 %) were intrascanner pairs and 200 (11 %) were interscanner pairs. This final cohort of 1885 MRI pairs had a mean (SD) PBVC of −0.26 % (0.58) and variance of 0.34 % (Fig. 3). This corresponds to a minimum detectable effect size of 0.24 % annualised PBVC (for a hypothetical cohort of 100 patients, power 0.95 and $\alpha = 0.05$).

3.2.3. Scan interval results

We further assessed the variance of the final 1885 MRI pairs by the time difference between MRI-1 & 2, noting that all included pairs were already ≥4 months apart. Limiting scan intervals to ≥6 months (183 days) resulted in 1659 MRI pairs with mean (SD) PBVC −0.27 % (0.53) and variance 0.28 %. 1379 MRI pairs with scan intervals ≥8 months (244 days) had mean (SD) PBVC −0.25 % (0.49) and variance 0.24 %. The mean (SD) and variance of 753 MRI pairs with scan intervals ≥12 months (365 days) were −0.22 % (0.45) and 0.21 % respectively.

3.3. Prediction of annualised PBVC

The final cohort of 1885 MRI pairs with QC status 'Approved' and 'Approved with remarks' (without a quality remark or comment) were subsequently used to identify predictors of annualised PBVC. Demographic, clinical, scanner and radiological variables (Table 2) were assessed using a two-step mixed model analysis. The final multivariable model identified age, sex, scanner field strength, MRI vendor, GMV at MRI-1 and relapses prior to MRI-1 as the six independent predictors for PBVC ($R^2 = 0.05$) (Table 2).

4. Discussion

In this study, we explored quantification of PBVC in a multicentre real-world setting using icobrain. We found that exclusion of consecutive MRI pairs with affine index <0.2 and scan interval <4 months led to an attrition of 23 % pairs. Interscanner pairs made up the majority of MRI pairs with affine index <0.2. We found comparable variance in PBVC between the remaining intrascanner and interscanner MRI pairs (0.78 % and 0.80 % respectively), indicating good performance of the affine index method. Further filtering of MRI pairs by image quality remarks or "consequences" listed in the 'Approved with remarks' group resulted in selection of a final cohort of MRI pairs which were 42 % of the original number of consecutive MRI pairs. This final cohort of MRI pairs had the lowest variance, 0.34 %, again illustrating both the effectiveness and need for these QC steps.

Additional examination of MRI pairs at one site showed that a large majority of pairs excluded for image quality reasons had low contrast. This was only an issue when using the standard scanner sequence for brain volumetry and ceased to be a problem after modification of the sequence. The most important factors that influence the affine similarity index are related to substantial differences in MRI acquisition protocols, which influence the contrast and noise in the image. These technical factors are more influential than volume differences for the time

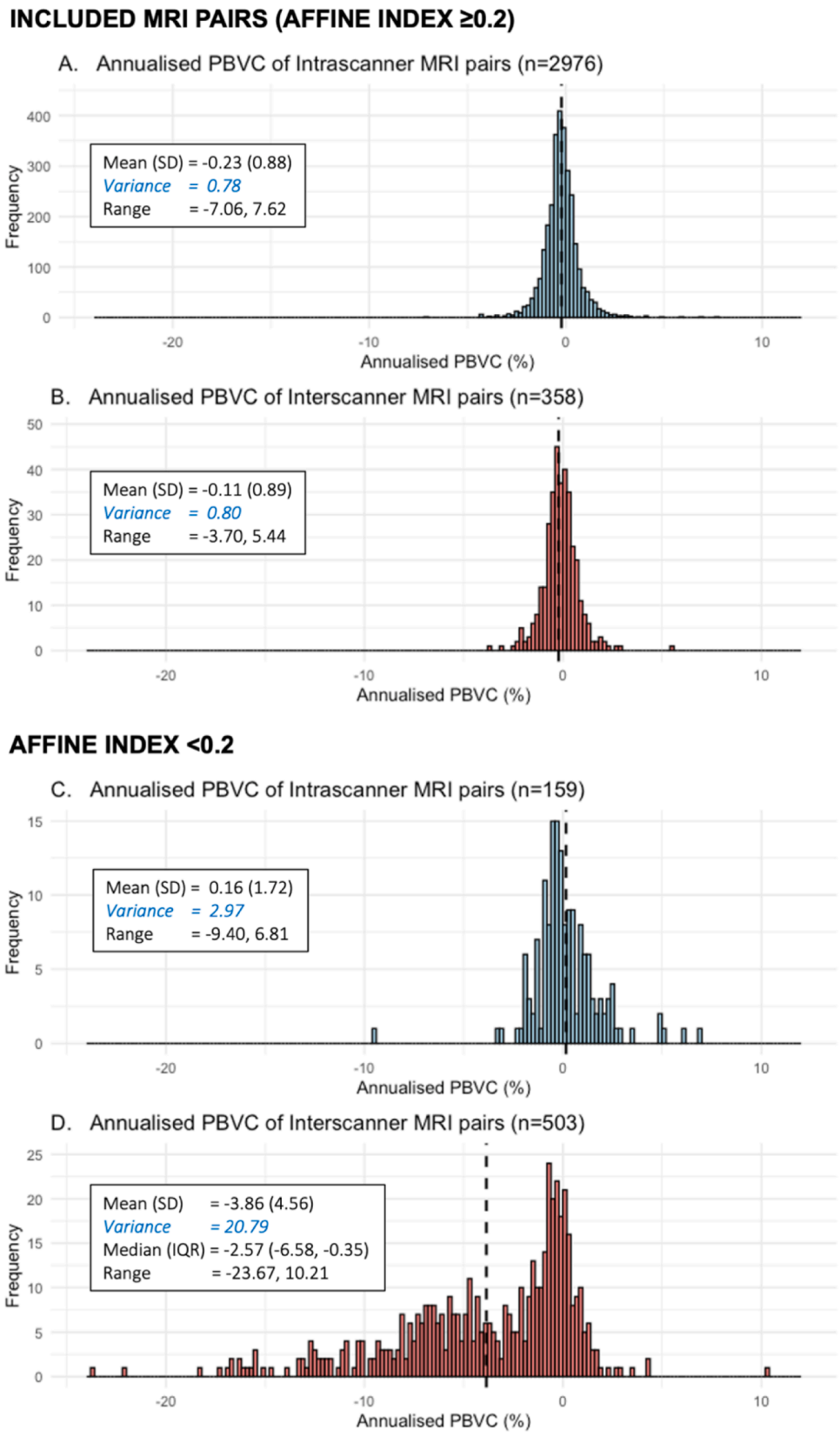


Fig. 2. Distribution of PBVC values for intrascanner and interscanner MRI pairs based on affine index cutoff. PBVCs (%) are shown for intrascanner and interscanner MRI pairs where affine index ≥0.2 and <0.2. The dotted line represents the mean.
Abbreviations: IQR, interquartile range; MRI, magnetic resonance imaging; n, number; PBVC, percentage brain volume change; SD, standard deviation.

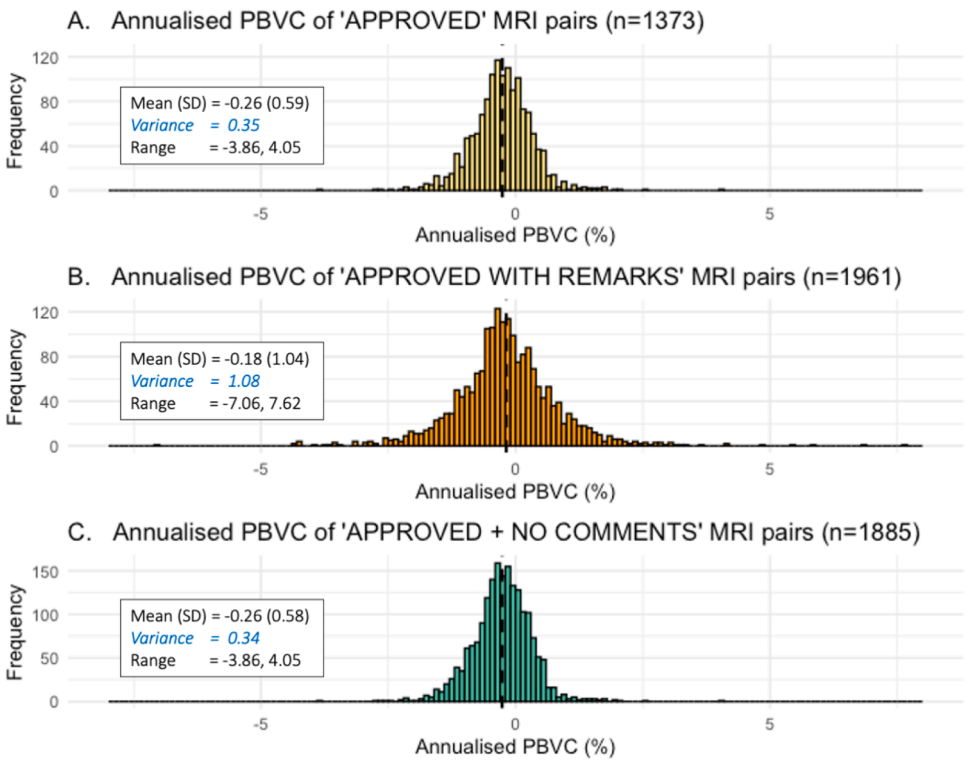


Fig. 3. Distribution of PBVC values based on QC status
PBVCs (%) are shown based on QC status: 'Approved' (A), 'Approved with remarks' (B) and the combination of 'Approved' & 'Approved with remarks' where there is no qualifying remark or comment for the latter (C). The dotted line represents the mean.
Abbreviations: MRI, magnetic resonance imaging; n, number; PBVC, percentage brain volume change; QC, quality control; SD, standard deviation.

intervals seen in our cohort. It highlights the importance of selecting an optimal MRI protocol for volumetric software before implementation on routine clinical MRIs.

Brain atrophy has been used as an outcome measure in MS clinical trials for over a decade but its implementation in clinical practice has been difficult. Analysis of brain volumetry in trial data is relevant at the group level. However, interpretation of this data on an individual level has several potential sources of error. Interscanner variability is the major limitation in real-world longitudinal measurements, and we have confirmed this as a major roadblock for quantitative serial MRI analysis (Huppertz et al., 2010).

Prior studies that explored the effect of interscanner variability on volumetric MRI analysis suggested that it could be mitigated during longitudinal follow-up with multiple recorded time-points. Takao and colleagues followed 208 healthy participants on two 3.0T MRIs over 2 years (Takao et al., 2011) and found no differences in SIENA PBVCs between the four combinations of scanners, with the standard deviation of PBVC reported as 0.71 %. A contemporary study by Durand-Dubief et al. assessed 9 MS patients on two 1.5T MRI scanners at three time-points over 1 year (Durand-Dubief et al., 2012). They also found that registration-based algorithms were less sensitive to MRI scanner changes and showed acceptable reproducibility across the 2 scanners. The reported mean difference in PBVC between scanners was 0.45 % using Jacobian integration and 0.12 % using SIENA.

Conversely, Biberacher and colleagues found significant scanner effects on PBVC (Biberacher et al., 2016). Two patients with relapsing-remitting MS were scanned 6 times on three 3.0T MRIs over 3 weeks. The PBVCs estimated by SIENA fluctuated between −4.45 % and +4.48 %. A recent study by Lee et al. assessed 237 subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study who were scanned longitudinally on the same or different 1.5T MRIs (Lee et al., 2019). They found that different interscanner combinations led to different offsets in SIENA PBVCs that ranged from −1.81 % to +0.99 %. The mean absolute error of PBVC in these studies was higher than in the original validation study in healthy controls (Smith et al., 2002). Our study adds to this knowledge by validating the affine index as a software tool to reduce scanner effects on the variance of PBVC.

We further assessed the feasibility of our real-world data to support studies by estimating the sample size required of a multicentre study aimed at a hypothetical treatment effect on annualised PBVC of 50 %. Using the mean and SD of the 3334 included MRI pairs with affine index >0.2, there would need to be at least 1155 MRI pairs per arm (using α=0.05 and power of 1-β=0.8), excluding the 25 % of pairs discarded due to affine index <0.2. Using the 1885 final MRI pairs (discarding the 58 % of scan pairs with remarks on quality control), we would require a minimum of 314 MRI pairs per arm. To obtain such a usable sample we should include at least 750 patients in our real-life study, expecting that 42 % of their MRI pairs will pass the quality control. This estimated sample size is significantly lower than that required if we did not exclude the 'Approved with remarks' group, and emphasises the importance of careful selection of MRI pairs in multi-site or multi-scanner analyses. However, it is still higher than the sample size that would be required in a single-scanner setting, (Beadnall et al., 2019; Rocca et al., 2017) and higher than the sample size estimates reported for multicentre non-trial MRI data on untreated MS patients (Takao et al., 2013) and healthy controls (Battaglini et al., 2019).

We showed that a greater annualised BV loss was associated with older age, male sex, smaller GMV at MRI-1, presence of a relapse in the month prior to MRI-1, MRI pairs of different field strength and transition between General Electric and Philips scanners (compared to Siemens and Philips transitions). Our models explained only 5 % of the variance of PBVC, despite selecting the 'best' MRI pairs. Thus, almost all variability in PBVC remains unexplained. To some extent, this could be explained by lack of availability of known contributors of BV loss such as scanner upgrades, protocol or coil changes, physiological diurnal variation, lifestyle (hydration state, alcohol intake, smoking), comorbidities

Table 2
Variables tested as potential predictors of annualised PBVC.

Variables	Univariable models Beta (SE), p-value (1885 MRI pairs, 974 pts)	Multivariable model Beta (SE), p-value (1777 MRI pairs, 935 pts)
DEMOGRAPHIC		
Sex (male)	−0.08 (0.03), 0.01	−0.10 (0.03), 0.01
Age at MRI, y	-3.1×10^{-3} (1.3×10^{-3}), 0.02	-4.2×10^{-3} (1.9×10^{-3}), 0.03
Age at MS diagnosis, y	-3.0×10^{-3} (1.4×10^{-3}), 0.04	- [g]
MS disease duration, y	-1.6×10^{-3} (2.1×10^{-3}), 0.46	
Site (reference: site 1)[a]		
Site 2	0.16 (0.04), <0.001	0.15 (0.08), 0.07
Site 3	0.01 (0.04), 0.84	−0.31 (0.20), 0.11
Site 4	0.06 (0.04), 0.13	0.06 (0.09), 0.49
MRI		
Scan interval, y	−0.18 (0.03), <0.001	0.05 (0.03), 0.10
Switch of field strength (reference 1.5T⟷3T)		
1.5T⟷1.5T	0.75 (0.21), <0.001	0.94 (0.23), <0.001
3T⟷3T	0.79 (0.21), <0.001	0.62 (0.26), 0.02
Switch of vendor (reference: GE⟷Philips)		
Siemens⟷Philips	1.2 (0.82), 0.15	2.1 (0.85), 0.01
GE⟷GE	1.1 (0.58), 0.06	1.1 (0.58), 0.07
Philips⟷Philips	1.1 (0.58), 0.05	0.77 (0.59), 0.20
Siemens⟷Siemens	1.1 (0.58), 0.05	1.1 (0.58), 0.07
Icobrain software version (reference: 3.1.1)		
3.1.2	−0.42 (0.59), 0.47	
3.1.4	−0.36 (0.58), 0.53	
Intrascanner vs interscanner: Intrascanner	−0.10 (0.04), 0.03	−0.03 (0.05), 0.59
RADIOLOGICAL (MS)		
BV of MRI-1[b], ml	5.2×10^{-4} (1.8×10^{-4}), 0.004	7.1×10^{-4} (3.8×10^{-4}), 0.06
GMV of MRI-1[2], ml	5.3×10^{-4} (2.6×10^{-4}), 0.04	-1.2×10^{-3} (5.6×10^{-4}), 0.03
FLAIR lesion volume (MRI-1)[b], ml	-9.5×10^{-3} (2.2×10^{-3}), <0.001	−0.01 (9.3×10^{-3}), 0.20
T1 hypointense lesion volume (MRI-1)[b], ml	−0.01 (2.2×10^{-3}), <0.001	4.7×10^{-3} (0.01), 0.71
FLAIR lesion volume change,%	0.01 (0.01), 0.45	
CLINICAL (MS)		
Relapse number between MRI-1 & 2[c]	−0.01 (0.03), 0.71	
Relapse in month pre MRI (reference: no relapses or relapse pre MRI-1 & 2)		
Relapse pre MRI-1 only	−0.15 (0.07), 0.04	−0.15 (0.07), 0.03
Relapse pre MRI-2 only	−0.03 (0.07), 0.71	−0.05 (0.08), 0.54
Corticosteroid use in month pre MRI (reference: no corticosteroid or corticosteroid pre MRI-1 & 2)		
Corticosteroid pre MRI-1 only	−0.06 (0.08), 0.48	
Corticosteroid pre MRI-2 only	−0.08 (0.10), 0.41	
DMT use during MRI pair (yes)	−0.07 (0.05), 0.16	
Duration of time on DMT used during MRI pair (time pre MRI-1), y	-9.0×10^{-4} (6.3×10^{-3}), 0.89	
Proportion of time on DMT during MRI pair,%	-3.9×10^{-4} (4.6×10^{-4}), 0.40	
Proportion of time on low-efficacy DMT[d] during MRI pair,%	2.3×10^{-4} (4.0×10^{-4}), 0.57	
Proportion of time on high-efficacy DMT[e] during MRI pair,%	-3.4×10^{-4} (3.2×10^{-4}), 0.29	
EDSS at visit closest to MRI-1[f]	−0.02 (0.01), 0.02	-1.1×10^{-3} (9.5×10^{-3}), 0.91
MS course at visit closest to MRI-1 (reference: CIS)		
RR	−0.06 (0.12), 0.61	
SP	−0.11 (0.13), 0.42	

Abbreviations: BV, brain volume; CIS, clinically isolated syndrome; DMT, disease modifying therapy; EDSS, expanded disability status scale; FLAIR, fluid-attenuated inversion recovery; GE, General Electric; GMV, grey matter volume; MRI, magnetic resonance imaging; MRI-1, first MRI in pair; MRI-2, second MRI in pair; MS, multiple sclerosis; PBVC, percentage brain volume change; pts, patients; RR, relapsing-remitting; SE, standard error; SP, secondary-progressive; T, tesla; y, years.

[a] Site 1=Brain and Mind Centre (Sydney, AU), Site 2=Royal Melbourne Hospital (Melbourne, AU), Site 3=Box Hill Hospital (Melbourne, AU), Site 4=General University Hospital (Prague, CZ).

[b] $N = 1880$.

[c] Excluding 1 month pre MRI-2.

[d] Low-efficacy DMT include: interferon beta, glatiramer acetate, teriflunomide.

[e] High-efficacy DMT include: alemtuzumab, autologous stem-cell transplant, cladribine (Cladribine), dimethyl-fumarate, fingolimod, mitoxantrone, natalizumab, ocrelizumab, rituximab.

[f] If visit within 12 months of MRI-1. $N = 1782$.

[g] Age at MS was excluded from the multivariable model due to a collinear relationship with age at MRI.

(cardiovascular risk, trauma) or treatment effects (pseudoatrophy (Sastre-Garriga et al., 2020; Rocca et al., 2017)). These factors are likely significant, with one study on healthy subjects demonstrating that scanner upgrade had effects comparable to that of using different scanners (Takao et al., 2013). However, it does highlight that further work is required before brain volumetry can be incorporated for individualised patient monitoring in MS.

Another limitation is that the scan interval in our cohort was relatively short. We demonstrated that longer scan intervals reduce the variability of PBVC, which is not unexpected as the contribution of variability in each individual is divided by the follow-up interval (Battaglini et al., 2019; Opfer et al., 2018; Narayanan et al., 2020). Battaglini et al. found that the 80th normative percentile of annualised PBVC was 50 % smaller for participants with 2 year than with 1 year follow-up (Battaglini et al., 2019). In our study, using the sample size calculation for different scan intervals, we found that lengthening the interval to ≥6 months reduced the sample size to 243 per arm, and this figure was similar for the 8 and 12 month cutoffs. Therefore, it is worthwhile considering an increase to the scan interval from 4 months to a minimum 6 months. This could be done by changing the MRI pair inclusion criteria or combining ≥3 MRIs from a patient to construct pairs with longer intervals, resulting is less attrition of scans. Additionally, it may be necessary to study multiple timepoints for each patient to establish a trajectory and further minimise variability.

5. Conclusion

This real-world study using icobrain highlights the challenges and limitations of brain volumetry measurements on group and individual data in clinical settings and the need for standardisation to perform it

adequately. Using this commercially available automated software, we were able to define a real-world volumetric dataset to detect group-level PBVC changes with low variance. Future development in this area will need to improve the signal to noise ratio to enable use of volumetric data in multi-site settings to support clinical decisions in individual patients.

Based on the observations made in this cohort, we recommend applying strict selection criteria for volumetric analysis such as icobrain in real-world settings, in line with the MAGNIMS guidelines (Sastre-Garriga et al., 2020; Wattjes et al., 2015). We suggest that longitudinal MRI analysis be performed on the same scanner, and recommend implementation of a thorough QC process, including scan similarity measures. Other suggestions include increasing the scan interval of analysed MRI pairs and optimising the MRI protocol for volumetric software, if possible, by the MS centre.

CRediT authorship contribution statement

Ai-Lan Nguyen: Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Maria Pia Sormani:** Writing – review & editing, Supervision, Methodology. **Dana Horakova:** Writing – review & editing, Resources. **Eva H Havrdova:** Writing – review & editing, Resources. **Michael H Barnett:** Writing – review & editing, Resources. **Nicola De Stefano:** Writing – review & editing. **Marco Battaglini:** Writing – review & editing. **Manuela Vaneckova:** Writing – review & editing, Resources. **Elaine Lui:** Writing – review & editing, Resources. **Frank Gaillard:** Writing – review & editing, Resources. **Patricia M Desmond:** Writing – review & editing, Resources. **Hayden Prime:** Writing – review & editing, Resources. **Mineesh Datta:** Writing – review & editing, Resources. **Anneke Van der Walt:** Writing – review & editing, Supervision. **Vilija G Jokubaitis:** Writing – review & editing, Supervision. **Femke Podevyn:** Writing – review & editing, Software, Data curation. **Robert Zivadinov:** Writing – review & editing. **Bianca Weinstock-Guttman:** Writing – review & editing. **Marie B D'hooghe:** Writing – review & editing. **Guy Nagels:** Writing – review & editing. **Vincent Van Pesch:** Writing – review & editing. **Guy Laureys:** Writing – review & editing. **Liesbeth Van Hijfte:** Writing – review & editing. **Jeannette Lechner-Scott:** Writing – review & editing. **Francesco Patti:** Writing – review & editing. **Edgardo Cristiano:** Writing – review & editing. **Juan I Rojas:** Writing – review & editing. **Diana M Sima:** Writing – review & editing, Software. **Wim Van Hecke:** Writing – review & editing, Software. **Tomas Kalincik:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Helmut Butzkueven:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Wim Van Hecke reports a relationship with Icometrix NV that includes: board membership and equity or stocks. Diana M. Sima reports a relationship with Icometrix NV that includes: employment. Guy Nagels reports a relationship with Icometrix NV that includes: consulting or advisory and equity or stocks. Femke Podevyn reports a relationship with Icometrix NV that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This study was supported by a Postgraduate Scholarship and Ian Ballard Travel Award from Multiple Sclerosis Research Australia (Dr Nguyen), and an Australian Government Research Training Program Scholarship (Dr Nguyen). This research received financial support from Biogen, Novartis and Roche. The MSBase Foundation is a not-for-profit organization that receives support from Roche, Merck, Biogen, Novartis, Sanofi Genzyme, and Alexion.

Acknowledgements

We thank Angela Agostinelli (Royal Melbourne Hospital), Michael Fitzgerald (Box Hill Hospital), Kerry Sims (I-MED Radiology), Erin Branagan (I-MED Radiology) and Tim Wang (Brain and Mind Centre) for their MRI technical assistance. We thank Tomas Uher (Charles University & General University Hospital, Prague), members of the Clinical Outcomes Research (CORe) Unit at the University of Melbourne (Izanne Roos, Sifat Sharmin, Ibrahima Diouf, Charles Malpas) and the Biostatistics Unit at the University of Genoa (Luca Carmisciano, Francesca Bovis, Alessio Signori) for their support and assistance with the statistical design and coding. We would also like to thank the MSBase Registry Team (Charlotte Sartori and Eloise Hinson) for their administrative support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.msard.2024.106148.

References

Popescu, V., Agosta, F., Hulst, H.E., et al., 2013. Brain atrophy and lesion load predict long term disability in multiple sclerosis. J. Neurol. Neurosurg. Psychiatr. 84, 1082–1091. https://doi.org/10.1136/jnnp-2012-304094, 2013/03/26.

Uher, T., Horakova, D., Kalincik, T., et al., 2015. Early magnetic resonance imaging predictors of clinical progression after 48 months in clinically isolated syndrome patients treated with intramuscular interferon β-1a. Eur. J. Neurol. 22, 1113–1123. https://doi.org/10.1111/ene.12716, 2015/04/22.

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

Sastre-Garriga, J., Pareto, D., Battaglini, M., et al., 2020. MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. Nat. Rev. Neurol. 16, 171–182. https://doi.org/10.1038/s41582-020-0314-x.

Smith, S.M., De Stefano, N., Jenkinson, M., et al., 2001. Normalized accurate measurement of longitudinal brain change. J. Comput. Assist. Tomogr. 25, 466–475, 2001/05/15.

Smith, S.M., Zhang, Y., Jenkinson, M., et al., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage 17, 479–489, 2002/12/17.

Giorgio, A., De Stefano, N, 2013. Clinical use of brain volumetry. J. Magn. Reson. Imaging 37, 1–14. https://doi.org/10.1002/jmri.23671.

Jain, S., Sima, D.M., Ribbens, A., et al., 2015. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. NeuroImage Clin. 8, 367–375. https://doi.org/10.1016/j.nicl.2015.05.003, 2015/06/25.

Smeets, D., Ribbens, A., Sima, D.M., et al., 2016. Reliable measurements of brain atrophy in individual patients with multiple sclerosis. Brain Behav. 6, e00518. https://doi.org/10.1002/brb3.518, 2016/10/01.

Steenwijk, M., Amiri, H., Schoonheim, M., et al., 2017. Agreement of MSmetrix with established methods for measuring cross-sectional and longitudinal brain atrophy. NeuroImage Clin. 15, 843–853. https://doi.org/10.1016/j.nicl.2017.06.034.

Beadnall, H.N., Wang, C., Van Hecke, W., et al., 2019. Comparing longitudinal brain atrophy measurement techniques in a real-world multiple sclerosis clinical practice cohort: towards clinical integration? Ther. Adv. Neurol. Disord. 12. https://doi.org/10.1177/1756286418823462, 17562864188234622019/02/06.

D'hooghe, M.B., Gielen, J., Van Remoortel, A., et al., 2019. Single MRI-based volumetric assessment in clinical practice is associated with MS-related disability. J. Magn. Reson. Imaging 49, 1312–1321. https://doi.org/10.1002/jmri.26303, 2018/12/31.

Fragoso, Y.D., Wille, P.R., Abreu, M., et al., 2017. Correlation of clinical findings and brain volume data in multiple sclerosis. J. Clin. Neurosci. 44, 155–157. https://doi.org/10.1016/j.jocn.2017.06.006, 2017/07/01.

Durand-Dubief, F., Belaroussi, B., Armspach, J., et al., 2012. Reliability of longitudinal brain volume loss measurements between 2 sites in patients with multiple sclerosis: comparison of 7 quantification techniques. AJNR Am. J. Neuroradiol. 33, 1918–1924. https://doi.org/10.3174/ajnr.A3107.

Biberacher, V., Schmidt, P., Keshavan, A., et al., 2016. Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. Neuroimage 142, 188–197. https://doi.org/10.1016/j.neuroimage.2016.07.035.

Zivadinov, R., Khan, N., Medin, J., et al., 2017. An observational study to assess brain MRI change and disease progression in multiple sclerosis clinical practice-the MS-

MRIUS study. J Neuroimaging 27, 339–347. https://doi.org/10.1111/jon.12411, 2016/12/05.
Zivadinov, R., Bergsland, N., Korn, J.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. https://doi.org/10.3174/ajnr.A5442, 2017/11/23.
Storelli, L., Rocca, M.A., Pagani, E., et al., 2018. Measurement of whole-brain and gray matter atrophy in multiple sclerosis: assessment with MR imaging. Radiology 288, 554–564. https://doi.org/10.1148/radiol.2018172468, 2018/05/01.
Butzkueven, H., Chapman, J., Cristiano, E., et al., 2006. MSBase: an international, online registry and platform for collaborative outcomes research in multiple sclerosis. Mult. Scler. 12, 769–774. https://doi.org/10.1177/1352458506070775, 2007/02/01.
Studholme, C., Hill, D.L.G., Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. Patt. Recogn. 32, 71–86.
Sima, D.M., Horáková, D., Nguyen, A.-L., et al., 2019. Assessing the reliability of longitudinal MRI examinations in multiple sclerosis follow-up. Mult. Scler. 25 (S2), 253.
Huppertz, H.J., Kröll-Seger, J., Klöppel, S., et al., 2010. Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. Neuroimage 49, 2216–2224. https://doi.org/10.1016/j.neuroimage.2009.10.066, 2009/10/28.
Takao, H., Hayashi, N., Ohtomo, K., 2011. Effect of scanner in longitudinal studies of brain volume changes. J. Magnet. Resonan. Imaging: JMRI 34, 438–444. https://doi.org/10.1002/jmri.22636.
Lee, H., Nakamura, K., Narayanan, S., et al., 2019. Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements. Neuroimage 184, 555–565. https://doi.org/10.1016/j.neuroimage.2018.09.062, 2018/09/22.
Rocca, M., Battaglini, M., Benedict, R., et al., 2017. Brain MRI atrophy quantification in MS: from methods to clinical application. Neurology 88, 403–413. https://doi.org/10.1212/WNL.0000000000003542.
Takao, H., Hayashi, N., Ohtomo, K., 2013. Effects of the use of multiple scanners and of scanner upgrade in longitudinal voxel-based morphometry studies. J. Magnet. Reson. Imaging: JMRI 38, 1283–1291. https://doi.org/10.1002/jmri.24038.
Battaglini, M., Gentile, G., Luchetti, L., et al., 2019. Lifespan normative data on rates of brain volume changes. Neurobiol. Aging 81, 30–37. https://doi.org/10.1016/j.neurobiolaging.2019.05.010.
Opfer, R., Ostwaldt, A.C., Walker-Egger, C., et al., 2018. Within-patient fluctuation of brain volume estimates from short-term repeated MRI measurements using SIENA/FSL. J. Neurol. 265, 1158–1165. https://doi.org/10.1007/s00415-018-8825-8, 2018/03/16.
Narayanan, S., Nakamura, K., Fonov, V.S., et al., 2020. Brain volume loss in individuals over time: source of variance and limits of detectability. Neuroimage 214, 116737. https://doi.org/10.1016/j.neuroimage.2020.116737, 2020/03/21.
Wattjes, M.P., Rovira, À., Miller, D., et al., 2015. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-establishing disease prognosis and monitoring patients. Nat. Rev. Neurol. 11, 597–606. https://doi.org/10.1038/nrneurol.2015.157, 2015/09/15.