NeuroImage: Clinical 15 (2017) 769–779

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

Neurological software tool for reliable atrophy measurement (NeuroSTREAM) of the lateral ventricles on clinical-quality T2-FLAIR MRI scans in multiple sclerosis

Michael G. Dwyer[a,*], Diego Silva[b], Niels Bergsland[a,c], Dana Horakova[d], Deepa Ramasamy[a], Jaqueline Durfee[a], Manuela Vaneckova[e], Eva Havrdova[d], Robert Zivadinov[a,f]

[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] Novartis Pharma AG, Basel, Switzerland
[c] Magnetic Resonance Laboratory, IRCCS Don Gnocchi Foundation, Milan, Italy
[d] Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic
[e] Department of Radiology, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic
[f] MR Imaging Clinical Translational Research Center, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA

ARTICLE INFO

Keywords:
Brain atrophy
Ventricular volume
Automated measurement
Multiple sclerosis

ABSTRACT

Background: There is a need for a brain volume measure applicable to the clinical routine scans. Nearly every multiple sclerosis (MS) protocol includes low-resolution 2D T2-FLAIR imaging.
Objectives: To develop and validate cross-sectional and longitudinal brain atrophy measures on clinical-quality T2-FLAIR images in MS patients.
Methods: A real-world dataset from 109 MS patients from 62 MRI scanners was used to develop a lateral ventricular volume (LVV) algorithm with a longitudinal Jacobian-based extension, called NeuroSTREAM. Gold-standard LVV was calculated on high-resolution T1 1 mm, while NeuroSTREAM LVV was obtained on low-resolution T2-FLAIR 3 mm thick images. Scan-rescan reliability was assessed in 5 subjects. The variability of LVV measurement at different field strengths was tested in 76 healthy controls and 125 MS patients who obtained both 1.5T and 3T scans in 72 hours. Clinical validation of algorithm was performed in 176 MS patients who obtained serial yearly MRI 1.5T scans for 10 years.
Results: Correlation between gold-standard high-resolution T1 LVV and low-resolution T2-FLAIR LVV was $r = 0.99$, $p < 0.001$ and the scan-rescan coefficient of variation was 0.84%. Correlation between low-resolution T2-FLAIR LVV on 1.5T and 3T was $r = 0.99$, $p < 0.001$ and the scan-rescan coefficient of variation was 2.69% cross-sectionally and 2.08% via Jacobian integration. NeuroSTREAM showed comparable effect size ($d = 0.39–0.71$) in separating MS patients with and without confirmed disability progression, compared to SIENA and VIENA.
Conclusions: Brain atrophy measurement on clinical quality T2-FLAIR scans is feasible, accurate, reliable, and relates to clinical outcomes.

1. Introduction

Brain atrophy in multiple sclerosis (MS) was classically thought of as a late-stage phenomenon, secondary to the more salient white matter (WM) lesions (*sclerosi*) lending the disease its name (Murray, 2005). Over the past two decades, though, understanding of brain atrophy in MS has been substantially revised. It is now clear that atrophy begins very early in the disease process (Uher et al., 2014; Kalkers, 2002), can progress relatively independently of overt lesions (Fisniku et al., 2008), affects both gray matter (GM) and WM, and proceeds at up to 5 times the rate associated with normal aging (Miller et al., 2002). Perhaps most important of all, quantitative measurements of atrophy have been shown to be the best correlates and long-term predictors of both cognitive and clinical disability (Benedict et al., 2006; De Stefano et al., 2014; Summers et al., 2008; Zivadinov et al., 2016b).

This revised understanding of the importance and clinical relevance

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

http://dx.doi.org/10.1016/j.nicl.2017.06.022
Received 13 February 2017; Received in revised form 19 May 2017; Accepted 16 June 2017
Available online 19 June 2017

of brain atrophy in MS is largely due to the emergence of quantitative image-based computational techniques for measuring brain atrophy more precisely and accurately than is possible by eye. These techniques include brain parenchymal fraction (BPF), boundary shift integral (BSI), structural image evaluation, using normalization, of atrophy (SIENA), voxel-, tensor-, and deformation-based mapping, and many others (Smith et al., 2002; Avants et al., 2008; Freeborough and Fox, 1997; Ashburner and Friston, 2000). However, despite their success in research, it has proven difficult to translate these approaches to clinical routine data.

Unlike the research-quality MRI sequences underpinning many prior studies, clinical routine images pose many unique challenges. They often have lower signal-to-noise ratio (SNR) and/or spatial resolution due to trade-offs in scanning time. Also due to time constraints, full head coverage is often neglected. Clinical scans also lack standardization, which in turn is compounded by changes in MRI machine and/or scanner upgrades. Finally, they are also more susceptible to artifacts including gradient distortion, wrap-around and patient motion. For all these reasons, retrospective and/or prospective quantitative atrophy analysis in a clinical setting raises numerous theoretical and practical challenges.

Previous efforts at routinely-applicable atrophy measures have included qualitative metrics like ordinal scales (Simon et al., 2006) and semi-quantitative linear metrics like third ventricular width (Benedict et al., 2006). However, they failed to achieve widespread adoption, potentially due to issues of intra−/inter-rater variability and acquisition-related variability. Thus, there is a need for a brain atrophy measurement technique complementary to existing techniques: a fully quantitative method focused on robustness and applicability to clinical routine scans and large pre-existing datasets, while maintaining as much as possible the precision and accuracy of more traditional research methods.

Measurement of ventricular volume (VV) provides a potential solution to many of problems listed above, due to a confluence of physiological and technical factors. The potential for lateral VV (LVV) to be used as an early predictive marker for disability and cognitive impairment, and for use in identifying at-risk patients, was discussed more than a decade ago (Dalton et al., 2004). More recently, it has been shown that VV is one of the best predictors of confirmed MS disability progression and that VV can act as a proxy for more complex brain atrophy measures, like whole brain atrophy, over the long-term. In 176 MS patients followed with serial MRI over 10 years, the effect size between stable and disability-progression groups was similar for whole brain and LVV changes ($d = 0.55$ for whole brain, $d = 0.51$ for LVV), and both measures showed significant separation within the first two years (Zivadinov et al., 2016b). Change in LVV also correlated significantly with changes in GM and cortical atrophy over the 10-year period. In another longitudinal study of 54 patients, VV change over two years was found to be even more predictive of 5-year disability progression than whole brain volume change (Lukas et al., 2010). An even larger retrospective study of 261 subjects also showed VV change as one of the most important predictors of 10-year disability progression (Popescu et al., 2013).

From a technical perspective, LVV measurement may benefit from a number of factors (Table 1). First, the ventricle/tissue border is high contrast on T2-FLAIR (Zivadinov et al., 2016b). This means that contrast-to-noise ratio is generally high, even on rapidly acquired clinical routine scans. Second, the ventricles have a relatively simple shape, resulting in predictable edge positions and considerably less partial-voluming than the highly gyrified cortex. Because of this, LVV measurement may tolerate low resolution scans better than whole brain measurement. Finally, the central position of the lateral ventricles in the head confers two additional advantages: the lateral ventricles are unlikely to be cut off even in partial-head scans, and gradient distortion errors are diminished (Caramanos et al., 2010).

Various methods have previously been devised to segment and/or quantify VV on MRI in both healthy controls and in other diseases. Template-based techniques have been very successful, and include ALVIN (Kempton et al., 2011), ANIMAL (Collins et al., 1995), decision fusion (Heckemann et al., 2006), and multi-atlas selection (Aljabar et al., 2009). These are based on non-linear alignment to a template (or templates), and subsequent transfer of template-space labels to the target image. FIRST (Patenaude et al., 2011) takes an alternative approach, fitting a priori parametric shape models to the target image, which produces excellent results for subcortical nuclei, but faces some difficulties with the ventricles. More-recently, patch-based (Coupé et al., 2011) and machine learning techniques (Julazadeh et al., 2012) have also been proposed. Combination techniques fusing various expert rules and other heuristics with intensity information have seen some success as well (Barra and Boire, 2001; Xia et al., 2004). A direct longitudinal technique - VIENA - modifying the popular SIENA approach to assess edge motion solely along the ventricular border (rather than the whole brain) has also been described and validated (Vrenken et al., 2014).

Table 1
Summary of common clinical-routine imaging challenges and the properties of LVV that allow it to provide mitigation for many of these challenges. (LVV = lateral ventricular volume).

Challenge	LVV mitigation
Low-resolution scans	Low-complexity border and simple topology allow better fitting of edges and use of constraints
Imprecise positioning/ gradient distortions	Nearness to the isocentre alleviates warping issues
Incomplete head coverage	LVV's centrality makes it unlikely to be cut off
Other artifacts	Wrap-around is mitigated by central position, noise by the strength of the border, motion by the use of topological constraints and the simple structure

However, none of these approaches have focused specifically on clinical quality T2-FLAIR scans. Therefore, we set out to develop and validate an LVV quantification tool for this purpose. Our goal was not to produce a method more precise or accurate than those described above. Rather, we sought to produce a highly robust technique capable of operating on clinical-quality T2-FLAIR images and producing results similar to more established techniques run on high-resolution scans. To do this, we extended an earlier idea for whole brain analysis proposed by (Baillard et al., 2001), and adapted it to LVV measurement on T2-FLAIR images. The core of our approach is a combination of multi-atlas segmentation and level-set refinement. We also extended the cross-sectional method for direct longitudinal analysis by including an optional Jacobian integration step. We called the resulting algorithm "Neurological Software Tool for Reliable Atrophy Measurement" - NeuroSTREAM.

2. Methods

2.1. Subjects and datasets

For this study, de-identified, retrospective datasets were used. Demographic and clinical characteristics are reported in Supplement Table 1, while the MRI acquisition characteristics are reported in the Supplement Table 2. The Institutional Review Board of the University of Buffalo approved the use of all datasets.

2.2. Creation of templates and atlases

To provide the basis for template-driven segmentation, we first created three atrophy-level-specific templates. Since the major goal of this work was to produce an algorithm applicable to real-world data, we compiled a non-biased anonymized development dataset of paired T2-FLAIR and high-resolution T1 scans from 109 MS patients across 62

MRI scanners. This dataset was randomly selected from a number of published and unpublished datasets (Santos and Weinstock-Guttman, 2006; Uher et al., 2015; Zivadinov et al., 2017; Zivadinov et al., 2016a, 2016c), using a weighted sampling technique to uniformly cover a wide range of atrophy levels and lesion loads, while simultaneously balancing 1.5T and 3T scans. Scans were converted to NIfTI format, but no other prior pre-processing was performed. This multi-site, multi-scanner dataset was representative of MS in general, with 72% females, and a mean age of 36.4 ± 12.4 years.

On each subject's high-resolution T1 scan, SIENAX (Smith et al., 2001) was applied, with initial BET parameters of f = 0.3 and with the head and neck cleanup option enabled. Manual correction of deskulling was performed as necessary, and the regional option was employed to produce central CSF masks. Using the resulting normalized brain volume measurements, subjects were split into low, medium, and high-atrophy tertiles. From each of the three tertiles, T2-FLAIR images were combined using non-linear symmetric diffeomorphic image registration with trilinear interpolation to generate unbiased, multi-scanner, tertile-specific 1 mm isotropic T2-FLAIR templates (Avants et al., 2008). Reflecting the fact that lesion masks might not be available in future automated analyses, we did not employ cost-function masking. Initial rigid registrations were performed to create a midspace image within each tertile. The final three templates are shown in Fig. 1.

In addition to each template itself, template-specific ventricle atlases were created as follows. Central CSF (cCSF) masks derived from regional SIENAX outputs were manually adjusted by expert operators (DR/JD) to correct errors and to remove non-LVV CSF (e.g. cisterns and third ventricle). The corresponding 3D-T1 scans were then registered to the subjects' T2-FLAIR images using rigid-body registration, and the resulting transforms were composited with the subjects' T2-FLAIR-to-template non-linear transforms to bring each manually corrected LVV mask into template space. Individual subjects' normalized ventricle maps were then combined into a single ventricle map via voxelwise majority vote.

Finally, supplementary binary masks were created directly in each template space for use in later algorithm steps:

- A forced inclusion mask composed of areas within the lateral ventricles that are often automatically misclassified, primarily regions with a high probability of containing choroid plexus.
- A forced exclusion mask composed of areas outside the lateral ventricles but easily automatically misclassified, including the third ventricle and nearby cisterns.
- A template-space sub-AC/PC mask to exclude areas below the anterior-posterior commissure line. Although the inferior horns of the lateral ventricles also lie below this plane, they are small and often lead to misclassification.

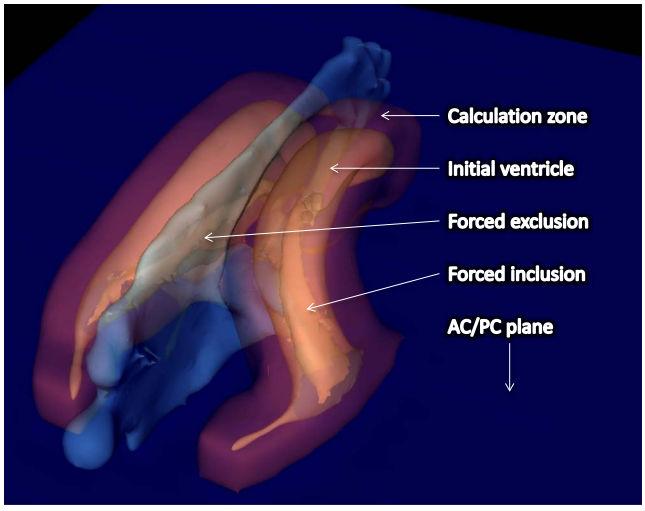


Fig. 2. 3D rendering of key regions and supplementary maps. To incorporate a-priori expert information into the NeuroSTREAM pipelines, binary mask volumes are included with each atrophy-level template. A calculation zone restricts the analysis to a probable-ventricle zone. An initial ventricle map provides a starting point for subsequent level set analysis. Forced inclusion and exclusion areas specifically include and exclude often-misclassified areas like choroid plexus and midline. Finally, an AC/PC plane map is used to reproducibly restrict measurement to the body, anterior, and posterior horns of the lateral ventricles. AC = anterior commissure, PC = posterior commissure.

Examples of these masks for the middle tertile atlas/template pair are shown in Fig. 2.

2.3. Algorithm description

The proposed NeuroSTREAM algorithm can be conceptualized in three phases: pre-processing, template-based segmentation, and level-set refinement. An overview of all processing steps is shown in Fig. 3.

2.3.1. Pre-processing

First, T2-FLAIR images are preprocessed to improve consistency and address specific issues. Initial reorientation and robust field of view selection are performed to ensure that the image is in a roughly standard position and that areas outside the head are cropped out (i.e., extra

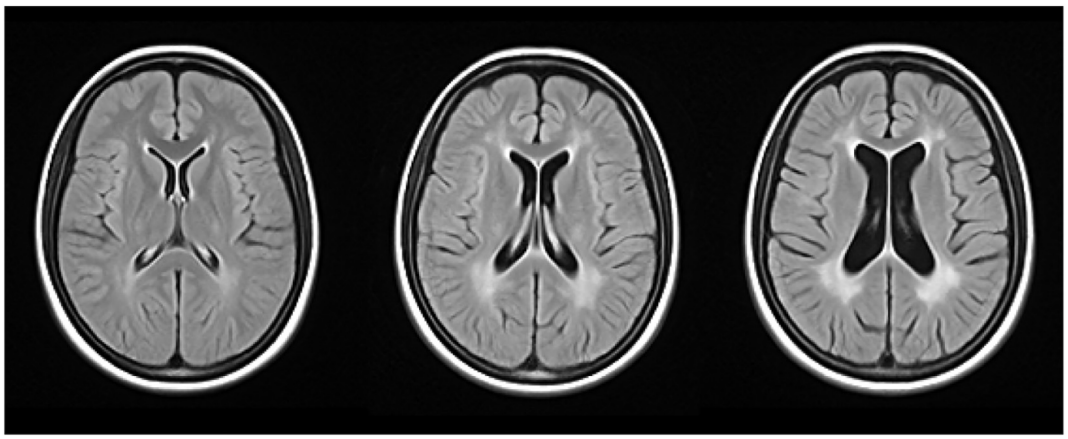


Fig. 1. Individual patients' images encompass a wide range of atrophy levels, making alignment to a single normative template (such as MNI152) problematic. NeuroSTREAM addresses this problem by providing atrophy-specific templates. These MS-specific T2-FLAIR templates were derived from a multi-scanner, multi-field-strength population. The population was split into tertiles based on normalized brain volume before creating a nonbiased, nonlinear template for each tertile. During processing, a subject's scan is aligned to all three templates, and resulting segmentations and maps are derived by a joint-fusion weighting scheme based on how well each alignment performed. Left: low atrophy template; Center: mid-atrophy template; Right: high-atrophy template.

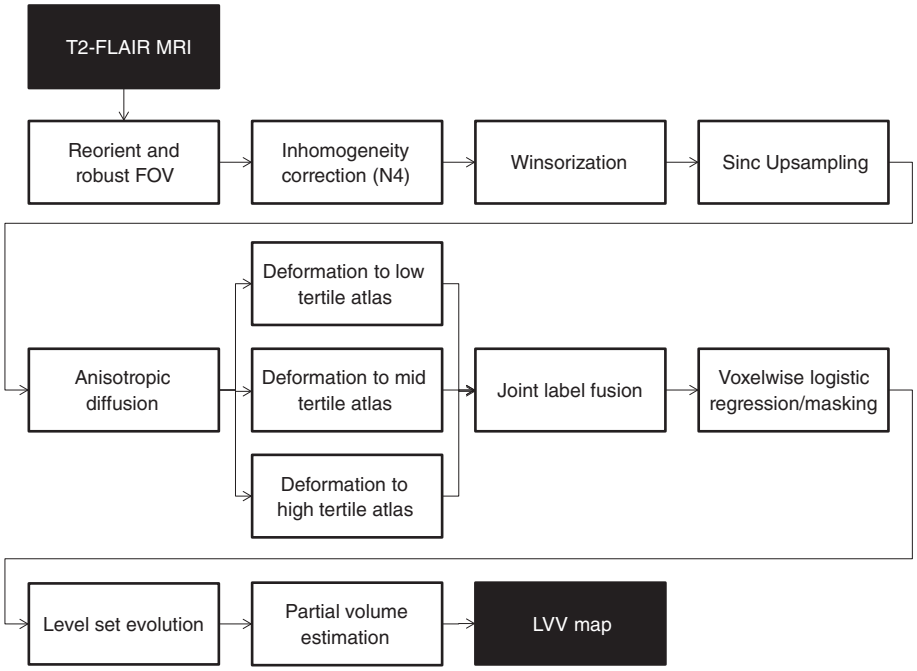


Fig. 3. Schematic overview of the NeuroSTREAM processing pipeline. LVV = lateral ventricular volume, FOV = field of view.

neck tissue). Next, bias field inhomogeneity is corrected using the N4 algorithm (Tustison and Avants, 2010). After these more generally-applicable steps, a T2-FLAIR-specific intensity Winsorization (Wilcox and Wilcox, 2005) routine is employed with an upper cutoff at the 97.5th percentile. This is done to ensure that very bright, subject-specific lesions do not substantially bias subsequent steps, but are also not completely ignored in the fitting. Then, for numerical purposes and to ensure consistency, the image is upsampled to 1 mm cubic voxels using windowed sinc interpolation. Finally, the isotropic, upsampled image is smoothed using a nonlinear anisotropic diffusion algorithm to reduce noise while preserving strong edges (Perona and Malik, 1990).

2.3.2. *Multi-atlas segmentation*

Non-linear alignment to each of the three previously-described templates is then performed via a two-stage process. In stage one, a linear alignment is performed using a 12 degree-of-freedom affine transformation model with a mutual information cost function (Avants et al., 2008). For this alignment, the full head is used. In stage two, a second linear alignment is performed using the same transformation and cost function, but with a cost function weighting map restricted to within 2 cm of the ventricles in template space. Finally, in stage three a non-linear alignment is performed using the SyN transform (Avants et al., 2008), again with the restricted ventricle-specific weighting mask. The final nonlinear transformations from each template are then applied to each of the three ventricular atlases, and the resulting maps are combined into a single map using the joint label fusion technique (Wang and Yushkevich, 2013). This joint fusion technique takes into account the quality of underlying fit of the images, giving more weight to atlases corresponding to templates that have been more accurately warped to the target subject's T2-FLAIR image. The same joint fusion process is applied for each atlas's related supporting maps, using the warp field determined from the atlas itself.

2.3.3. *Level set refinement*

The template-based segmentation described above provides a map of the ventricles that is often of good quality. However, SyN and most similar non-linear warping techniques are toplogy-preserving. While generally a positive feature, this can be problematic with low-resolution scans in which portions of the ventricles may appear effectively disconnected even when upsampled (Coupé et al., 2011). The level-set framework provides an effective complementary approach, since it naturally handles changes in toplogy (Heimann and Meinzer, 2009). Therefore, a numerical level-set based evolution is performed on the initial template-based LVV segmentation. The level-set speed function, which drives the evolution of the segmentation, is produced via a voxel-wise combination of factors. First, two groups of voxels are selected: those within the preliminary LVV and at least 1 mm from the border (inside voxels) and those outside the preliminary LVV between 1 mm and 10 mm from the border (outside voxels). A logistic regression model is then fit between the intensities of these two voxel sets to produce a function predicting probability of LVV membership from image intensity. From this function, a voxel-wise LVV probability map is created, and then refined as follows: all voxels in the forced inclusion mask are set to have a probability of 1, all voxels in the forced exclusion mask are set to have a probability of 0, and all voxels in the sub-AC/PC mask are set to have a probability of 0. Given this speed function and the initial LVV segmentation from the template-based stage, a level-set evolution is then performed with 25 iterations and a curvature value of 0.2 (Yushkevich et al., 2006).

2.3.4. *Partial volume estimation*

Despite the ventricles' comparatively simple topology, most voxels along the edges of the LVV still contain portions of both CSF and surrounding tissue. Common approaches to deal with this issue involve the use of explicit Gaussian mixture models to assign tissue-specific partial volume estimates (PVE) to each voxel (Avants et al., 2011; Zhang et al., 2001). However, this method is not reliable on T2-FLAIR images where unpredictable periventricular lesions substantially skew intensity distributions. To address this, the proposed algorithm instead uses a simpler but more robust super-resolution technique, relying on the level set evolution itself to perform implicit edge-based partial volume estimation. Each voxel within the LVV mask and the previously described speed function is subdivided into 8 subvoxels. An additional level-set refinement is then performed for 5 iterations in this higher-resolution space, and the final map is then subsampled back into the original space. The proportion of subvoxels included within the level set segmentation for each voxel then provides an approximate PVE estimate. This approach is numerically similar to previous level-set based PVE techniques (Rifai et al., 2000).

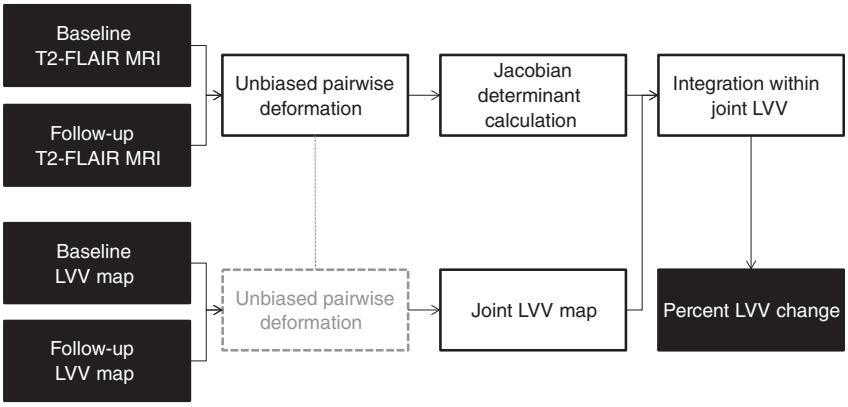


Fig. 4. Schematic overview of the longitudinal Jacobian-based extension to NeuroSTREAM. Shaded area indicates that pairwise deformation results are also used to bring a joint LVV map into a halfway space. LVV = lateral ventricular volume.

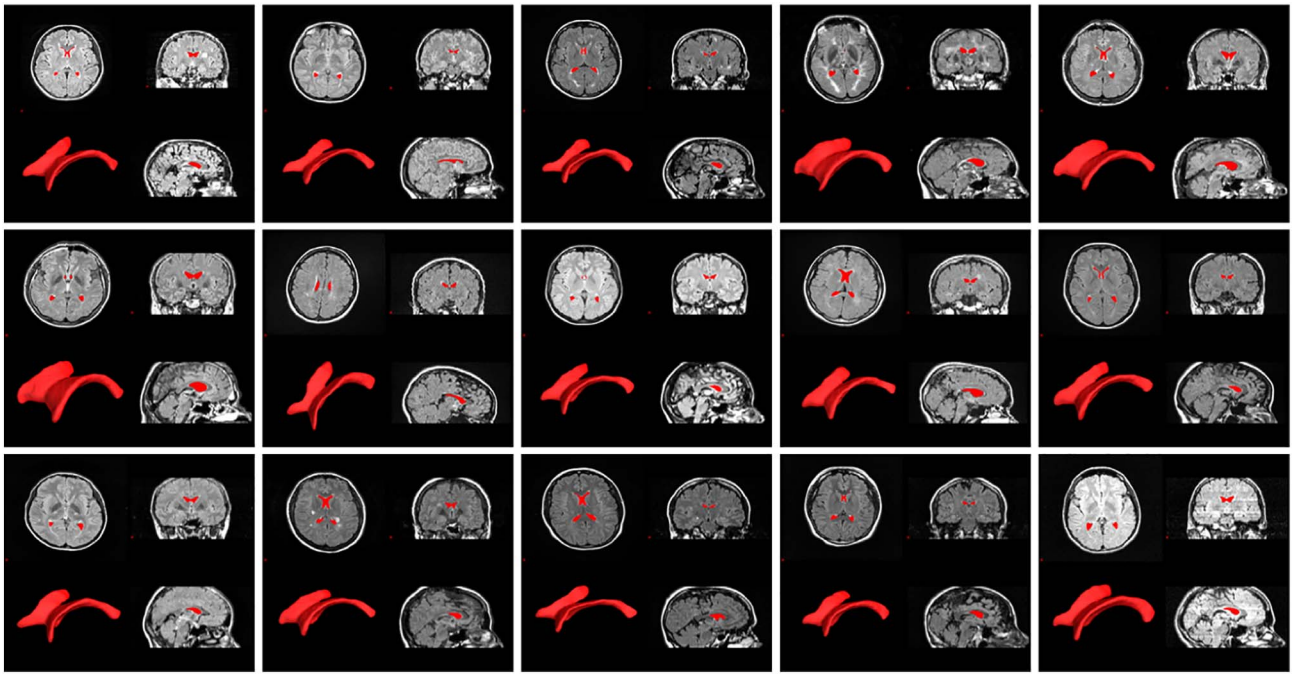


Fig. 5. Sample NeuroSTREAM LVV segmentations demonstrating the range of scan types, intensity profiles, resolutions, ventricular anatomy, and atrophy levels capable of being successfully segmented and quantified. Each sub-image shows an axial (upper left), coronal (upper right), and sagittal (lower-right) view, along with a 3D render of the extracted LVV (lower-left). LVV = lateral ventricular volume.

2.3.5. Direct longitudinal analysis via Jacobian integration

Because direct longitudinal approaches can often provide more accurate and precise outcomes (Zivadinov et al., 2016c) by avoiding compounded measurement errors, we also extended the above algorithm to use a Jacobian integration technique (Nakamura et al., 2014) to evaluate longitudinal T2-FLAIR changes. In this case, the full cross-sectional algorithm is run on both baseline and follow-up images to produce independent LVV maps. Then, the two images are warped together using nonlinear warping into a non-biased halfway space (Avants et al., 2012). Subsequently, the two ventricular masks are also brought into the halfway space and combined to produce a halfway LVV map. At the same time, the warp field is differentiated to produce a Jacobian determinant map. Finally, the Jacobian determinant field is integrated within the joint LVV map to produce a final estimate of percent volumetric change in LVV from baseline to follow-up. An overview of this processing is shown in Fig. 4.

2.3.6. Implementation details

The proposed method was implemented in Python and made use of the NiPype Neuroimaging Python framework (Gorgolewski et al., 2011). Experiments were carried out on a 32-core system with 192GB of RAM. Template alignment and longitudinal warping were performed using the ANTs toolkit (Avants et al., 2008), and logistic speed function prediction was performed with the scikit-learn machine learning toolbox (Pedregosa et al., 2011).

2.4. Validation

To validate the proposed NeuroSTREAM algorithm, we employed a number of complementary approaches in order to assess overall accuracy, precision, inter-scanner stability, and relevance to clinical outcomes. All NeuroSTREAM analyses described below were conducted fully automatically, with manual review of automatically generated segmentation images (and warp images in the case of longitudinal pairs). As the approach is intended to be fully automated, cases were not manually corrected - they were either accepted or rejected outright. Due to the exploratory nature of these validation analyses and the high interdependence between the various observations, correction for multiple comparisons between tests was not employed in this study.

2.4.1. Accuracy and agreement with manual gold standard (cross sectional)

To address accuracy, a manual gold standard dataset was created.

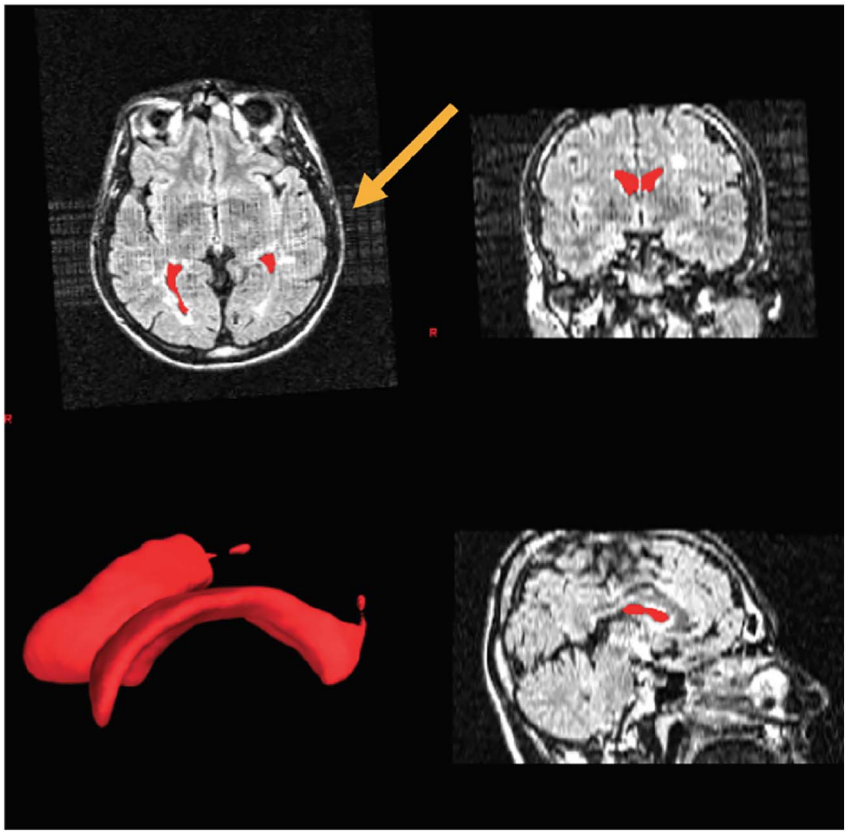


Fig. 6. Representative image of NeuroSTREAM algorithm performance on a scan with substantial RF artifact. The high-contrast of the ventricles, the multi-atlas approach, and the spatial regularization of the level-set refinement allow the algorithm to be very robust to artifacts like this that are common in clinical routine imaging.

An additional 30 independent MRI exams were selected randomly from the same pool as the dataset used for creation of templates and atlases, stratified to include 10 scans each of 3 mm, 5 mm, and 7 mm slice thickness T2-FLAIRs. Each exam also contained a high-resolution T1-weighted image. Lateral ventricles were delineated manually in JIM 6.0 (Xinapse Systems, UK) on the T1-weighted images, beginning at the slice best corresponding to the AC/PC line as determined on orthogonal views. Volumes were computed at a subvoxel level and summed across ROIs.

NeuroSTREAM was then run independently on the corresponding T2-FLAIR images, and the resulting automated volumes were compared to the manual gold standard volumes. Association with NeuroSTREAM was assessed by pairwise correlation, intra-class correlation (ICC), Bland-Altman (BA) plot, and casewise coefficient of variation (CoV). The same analyses were carried out on 3 mm, 5 mm, and 7 mm subsets, as well as acquisition-direction subsets, to determine the effects of these parameters.

2.4.2. Precision via scan-rescan (cross sectional and longitudinal)

To assess precision, we used a previously collected dataset of scan and re-scan sessions on the same 3.0 T GE Signa Excite HD 12.0 Twin Speed 8-channel scanner. A cohort of 6 volunteers (2 HC, 4 MS) with mean age was 28.7 ± 18.7 years, was scanned twice within one week, and protocols included identical 3 mm T2-FLAIR acquisitions.

For this scan-rescan dataset, NeuroSTREAM was run on each scan for each subject, and the Jacobian extension was run on each pair. Cross-sectional association was assessed by pairwise correlation, ICC, Bland-Altman plot, and casewise CoV. Longitudinal error was calculated as the median absolute Jacobian-derived percent change, with the true change assumed to be 0%.

2.4.3. Inter-scanner stability (cross sectional and longitudinal)

To assess stability across scanners, a previously collected dataset consisting of 125 MS patients and 76 healthy controls scanned at both 1.5T and 3T was used. Subjects were 72% female, with a mean age of 42.5 ± 11.1 years. All subjects were examined on both scanners within 72 h, and the order in which subjects were scanned was randomized. The scanners used were a 1.5T GE Signa Excite HD 12.0 8-channel scanner and a 3T GE Signa Excite HD 12.0 Twin Speed 8-channel scanner (General Electric Milwaukee, WI). 2D 3 mm T2-FLAIR acquisitions were included in each scanning session. Sequences were not identical between the two scanners, but rather reflected optimizations for the specific field strengths as would be seen in clinical practice.

As with the scan-rescan dataset, NeuroSTREAM was run on both scans for each subject, and association was again assessed by pairwise correlation, ICC, Bland-Altman plot, and casewise CoV. As above, longitudinal error was calculated as the median absolute Jacobian-derived percent change, with the true change assumed to be 0%.

2.4.4. Clinical relevance (cross sectional and longitudinal)

To assess clinical relevance, we used a 10-year serial validation dataset (Zivadinov et al., 2016b). Serial yearly MRI data were obtained from 176 early RRMS patients who were initially enrolled into the 2-year, double-blind, placebo-controlled Avonex-Steroid-Azathioprine (ASA) study (Havrdova et al., 2009). MRI scans with 1 mm thick 3D-T1s and 3 mm thick 2D T2-FLAIRs were acquired on a 1.5 T scanner at baseline and at yearly intervals, and confirmed disability progression (CDP) status was assessed at the 10-year follow-up. Full details of the design, demographic and clinical characteristics, and results of the study were previously reported (Zivadinov et al., 2016b). 100 patients developed CDP and 76 remained stable.

NeuroSTREAM was run on each scan for each subject. Additionally, the longitudinal Jacobian integration component was used to evaluate pairwise, within-subject changes. SIENAX and SIENA were previously applied to this dataset, and processing included lesion in-painting, manual skull stripping correction, and manual acceptance of results. Prospectively, we also applied the VIENA longitudinal ventricular change measurement tool (Vrenken et al., 2014) on these corrected

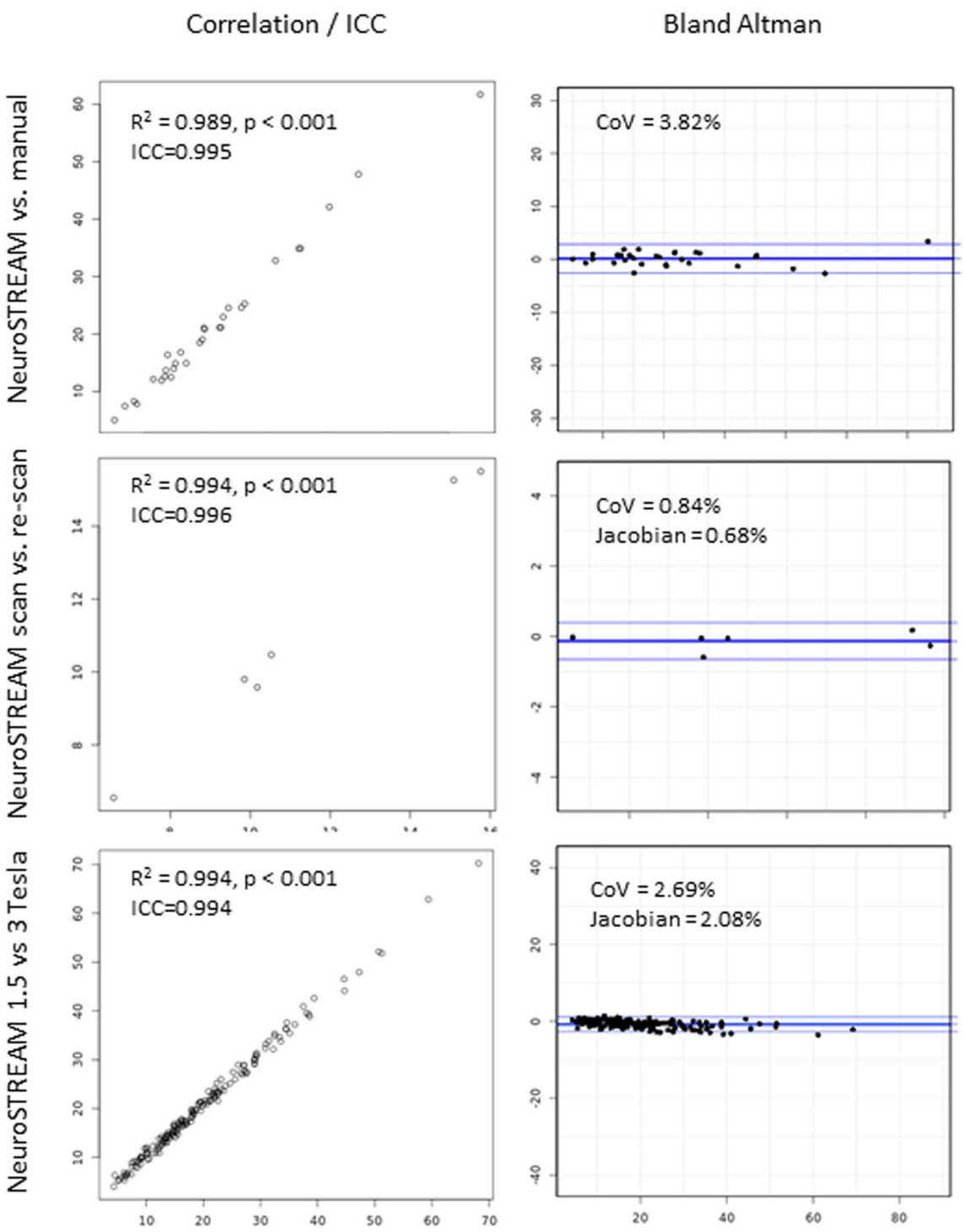


Fig. 7. Correlation scatterplots (left column) and Bland-Altman plots (right column) demonstrating accuracy and precision for the NeuroSTREAM algorithm. These results demonstrate accuracy in relation to manual gold standard measurements (upper row), scan-rescan precision (middle row), and inter-scanner (1.5 vs. 3 T) reliability (bottom row). Bland-Altman plot ranges are set to the scale of the domain, centered on 0. ICC = intra-class correlation (absolute agreement), CoV = coefficient of variation. Jacobian represents the mean absolute Jacobian change between scans, assessed with the longitudinal pipeline.

images. We compared NeuroSTREAM to conventional VV measures - SIENAX ventricular CSF (VCSF) for cross-sectional and VIENA percent ventricular CSF volume change (PVCSFVC) for longitudinal - by investigating the correlations between the various measures. We also evaluated the direct relationship between NeuroSTREAM LVV and whole-brain measures - SIENAX normalized brain volume (NBV) and SIENA percent brain volume change (PBVC). To determine the clinical relevance of NeuroSTREAM, we compared each measure with respect to development of CDP at 10 years, as previously reported for SIENA/SIENAX (Zivadinov et al., 2016b).

3. Results

3.1. Development dataset

The NeuroSTREAM algorithm successfully delineated the LVV on a wide variety of cases (Fig. 5), and was visually assessed as failed in only 1 case (~1%) from the template/atlas creation dataset. The algorithm

Table 2
Correlations between NeuroSTREAM LVV and previously established brain atrophy measures on the clinical 10-year serial dataset (Zivadinov et al., 2016a, 2016b, 2016c). NeuroSTREAM was assessed on conventional T2-FLAIR, and all other target measures were assessed on high-resolution 3D T1 images. Baseline and follow-up values are based on cross-sectional NeuroSTREAM, and change values are based on the Jacobian determinant extension. LVV = lateral ventricular volume, vCSF = ventricular CSF, NBV = normalized brain volume.

	SIENAX vCSF	VIENA	SIENAX NBV	SIENA
Baseline	0.93**	–	− 0.498**	–
Follow-up	0.94**	–	− 0.521**	–
Change	–	0.88**		− 0.581**

** $p < 0.01$.

performed well even in cases with extreme lesion load or corruption by MRI artifacts (Fig. 6).

3.2. Validation

3.2.1. Accuracy and agreement with manual gold standard (cross sectional)

On the independent testing dataset, no cases failed visual quality control. R^2 correlation coefficient between high-resolution, manually-delineated gold standard 3D-T1 volumes and automated volumes (from low-resolution T2-FLAIR) was 0.99 ($p < 0.001$), ICC was 0.99, and CoV was 3.82%, as shown in Fig. 6. The BA plot did not show any evidence of bias as a function of LVV (Fig. 7). When evaluated as a function of slice thickness, R^2 remained 0.99 in all subsets, and ICC was also consistently 0.99 (to 2 significant digits). However, CoV was 3.15% for 3 mm scans, 4.09% for 5 mm scans, and 4.22% for 7 mm scans. Seven scans were sagittally acquired, and the remaining scans were axially acquired. When evaluated as a function of orientation, R^2 and ICC were again 0.99 for both axial and sagittal scans. CoV was 3.84% for axial scans and 3.73% for sagittal scans.

3.2.2. Precision via scan-rescan (cross sectional and longitudinal)

The algorithm did not fail in any cases. R2 correlation coefficient was 0.99, ICC was 0.99, and CoV was 0.84%. The BA plot did not show any evidence of bias as a function of LVV (Fig. 7). Mean absolute Jacobian-derived percent change between scans was 0.68%.

3.2.3. Inter-scanner stability (cross sectional and longitudinal)

Out of 402 total scans, 4 analyses failed (< 1%). R2 correlation coefficient was 0.99, ICC was 0.99, and CoV was 2.69%. The BA plot did not show any evidence of bias as a function of LVV (Fig. 7). Mean absolute Jacobian-derived percent change from 1.5 to 3T was 2.08%.

3.2.4. Clinical relevance (cross sectional and longitudinal)

NeuroSTREAM was visually assessed as failed in 27 out of 1931 individual cross-sectional exams (~1.4%). Longitudinal Jacobian analysis was assessed as failed in 46 out of 1767 longitudinal scan pairs (~2.6%).

During the 10 year time period, PBVC changed by − 6.5%, corresponding to an annualized PBVC of − 0.67%. In contrast, ventricular volume change as measured by VIENA on high-resolution T1 increased by 41.97%, corresponding to an annualized change rate of 3.57% - a > 5-fold greater rate than PBVC. NeuroSTREAM LVV measures on low-resolution T2-FLAIR corresponded with VIENA, with PLVVC increasing by 38.64% over the same period (resulting in an annualized PLVVC of 3.32%).

Correlations between NeuroSTREAM and high-resolution T1-derived SIENAX VCSF and VIENA are reported in Table 2. Cross-sectional correlations were at or above r = 0.9, and 10-year change showed a correlation of 0.88. All p-values were below 0.01. Agreement was very high in all cases. NeuroSTREAM LVV was also significantly correlated with whole brain volumes and volume changes from SIENAX and SIENA, with longitudinal changes showing a correlation of $r = -0.581$ ($p < 0.001$). For comparison, the analogous correlation between high-resolution 3D T1 derived VIENA change with PBVC was $r = -0.635$ ($p < 0.0001$).

Table 3 reproduces the relationship between evolution of brain atrophy and CDP group for whole brain measures previously published (Zivadinov et al., 2016b). Previously, high-resolution-T1-derived PBVC showed significant group separation of confirmed disability progression (48 weeks) vs. stable MS within the first year of follow-up (54.0% difference, $p = 0.053$), and this difference remained significant at all time points of the study (effect sizes 0.27–0.70). VIENA (Table 4) showed similar results, with a significant difference after the first year (46.58% difference, $p < 0.03$) and effect sizes ranging from 0.31 to 0.62. This pattern remained the case with T2-FLAIR-derived NeuroSTREAM LVV, which showed separation after one year (60.0% difference, $p < 0.009$). The differences in NeuroSTREAM LVV also remained significant at all subsequent time points (effect sizes 0.39–0.71, Table 5), with the exception of a trend ($p = 0.052$) at 24 months. Graphical plots of SIENA PBVC, VIENA PVCSFVC, and NeuroSTREAM PLVVC are shown in Fig. 8.

4. Discussion

In this study, we developed and validated a new algorithm for measuring brain atrophy via LVV. LVV reflects both GM and whole brain atrophy (Zivadinov et al., 2016b), has clear borders even on low-quality scans, and is minimally susceptible to common MRI artifacts. Throughout development, we paid specific attention to the issues that arise when measuring brain atrophy on clinical-quality T2-FLAIR images, an MRI sequence performed as part of nearly all diagnostic and monitoring examinations for MS.

Our results show that NeuroSTREAM metrics are comparable to manual gold-standard, and to sophisticated SIENAX and VIENA LVV

Table 3
Reproduced with permission from (Zivadinov et al., 2016b). SIENA percent brain volume change from baseline to follow-up at each time point, derived form high-resolution 3D T1 images. All p-values are age and sex corrected. Between group p-values are derived from ANCOVA analysis. PBVC = percent brain volume change, CDP = confirmed disability progression over 10 years. CDP group shows atrophy differentiation after only 1 year.

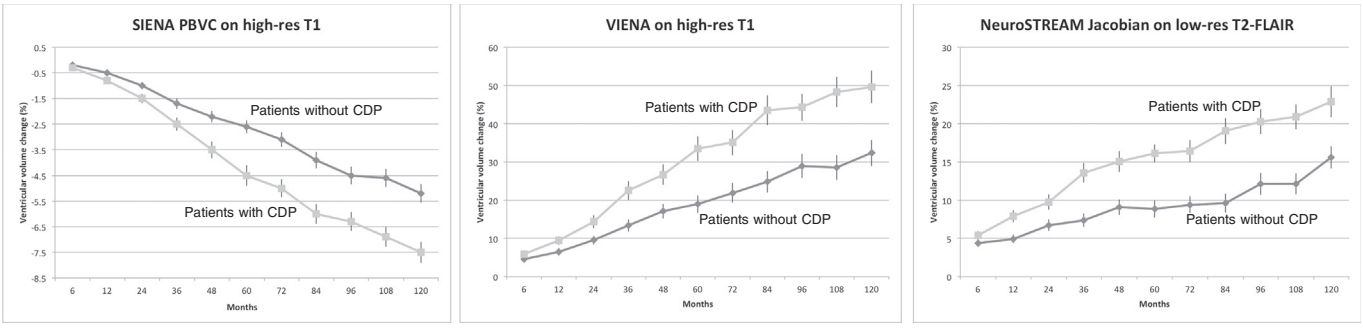
Months from baseline to	Stable group PBVC	n	CDP group PBVC	n	% Difference	Cohen's d	p-value
6	− 0.2 (0.8)	74	− 0.3 (1)	94	47.8	0.11	0.326
12	− 0.5 (0.9)	76	− 0.8 (1.3)	95	54	0.27	0.053
24	− 1 (1.1)	68	− 1.5 (1.6)	85	50.5	0.36	0.01
36	− 1.7 (1.7)	67	− 2.5 (2.4)	89	46.8	0.38	0.003
48	− 2.2 (1.8)	68	− 3.5 (3.1)	87	55.8	0.51	< 0.001
60	− 2.6 (2.1)	67	− 4.5 (3.8)	91	69.6	0.62	< 0.001
72	− 3.1 (2.3)	66	− 5 (3.2)	87	58.2	0.68	< 0.001
84	− 3.9 (2.6)	68	− 6 (3.6)	85	53.3	0.67	< 0.001
96	− 4.5 (2.8)	66	− 6.3 (3.4)	87	40.3	0.58	< 0.001
108	− 4.6 (2.9)	68	− 6.9 (3.6)	84	49.8	0.7	< 0.001
120	− 5.2 (3)	68	− 7.5 (3.8)	85	43.8	0.55	< 0.001

Table 4
VIENA percent ventricular CSF change from baseline to follow-up at each time-point, derived form high-resolution 3D T1 images. All p-values are age and sex corrected. Between group p-values are derived from ANCOVA analysis. CDP = confirmed disability progression, PVCSFVC = percent ventricular CSF volume change.

Months from Baseline to	Stable group PVCSFVC	n	CDP group PVCSFVC	N	% Difference	Cohen's d	p-value
6	4.6 (6.4)	74	5.9 (6.1)	94	29.3	0.21	0.123
12	6.4 (8.5)	76	9.4 (10.9)	95	46.6	0.31	0.032
24	9.5 (9.4)	68	14.4 (15.5)	85	50.9	0.38	0.018
36	13.4 (13.3)	67	22.6 (23.5)	89	68.5	0.41	0.005
48	17.1 (15.6)	68	26.7 (25.1)	87	55.8	0.46	0.005
60	19.0 (18.8)	67	33.5 (30.3)	90	59.4	0.58	0.001
72	21.9 (20.7)	66	35.1 (31.1)	88	60.2	0.5	0.002
84	24.8 (23.5)	68	43.5 (36.0)	85	75.4	0.62	< 0.001
96	29.0 (25.8)	67	44.3 (33.1)	87	53.0	0.52	0.003
108	28.5 (26.8)	68	48.3 (36.2)	84	69.3	0.62	< 0.001
120	32.4 (28.1)	67	49.7 (39.5)	85	53.4	0.5	0.003

Table 5
Percent NeuroSTREAM change from baseline to follow-up at each time-point, derived from T2-FLAIR images. All p-values are age and sex corrected. Between group p-values are derived from ANCOVA analysis. Note that results are comparable to both SIENA PBVC and CDP = confirmed disability progression, PLVVC = percent lateral ventricular volume change.

Months from baseline to	Stable group PLVVC	n	CDP group PLVVC	n	% Difference	Cohen's d	p-value
6	8.8 (8.5)	74	10.8 (10.2)	94	22.8	0.20	0.205
12	9.9 (10.2)	76	15.8 (15.6)	95	60.0	0.45	0.009
24	13.4 (12.4)	68	19.5 (18.6)	85	42.5	0.36	0.052
36	14.8 (14.5)	67	27.2 (23.9)	89	84.3	0.63	0.001
48	18.2 (16.9)	68	30.1 (25.7)	87	65.4	0.55	0.003
60	17.7 (18.8)	67	32.3 (22.2)	90	82.3	0.71	< 0.001
72	18.8 (17.6)	66	32.8 (27.2)	88	74.2	0.61	0.001
84	19.3 (20.9)	68	38.1 (31.2)	85	97.6	0.71	< 0.001
96	24.3 (23.5)	67	40.6 (30.4)	87	67.1	0.60	0.002
108	24.3 (22.4)	68	41.8 (30.1)	84	72.2	0.66	0.001
120	31.2 (23.9)	67	45.7 (37.4)	85	46.6	0.46	0.007



	PBVC: SIENA	PVCSFVC: VIENA	PLVVC: NeuroSTREAM
Mean CDP vs. non-CDP difference (year 1 and later)	51.8%	56.5%	65.3%
Effect size range (year 1 and later)	0.27–0.70	0.31 – 0.62	0.39–0.71

Fig. 8. Evolution of brain atrophy measures in patients with and without confirmed disability progression (CDP) over a 10-year period. Left: whole brain SIENA percent brain volume change (PBVC). Middle: VIENA-derived percent ventricular CSF volume change (PVCSFVC). Right: NeuroSTREAM-derived percent lateral ventricular volume change (PLVVC). SIENA and VIENA are performed on high-resolution 3D-T1 images. Despite being applied to low-resolution T2-FLAIR images, NeuroSTREAM shows comparable separation between CDP and non-CDP groups.

measures. This is despite the fact that NeuroSTREAM is applied to low-resolution images with 300% coarser voxels (3 mm vs. 1 mm). Furthermore, the resulting LVV measures have a similar predictive profile to SIENAX- and VIENA-derived LVV. The algorithm was also extremely robust, failing in only 1 of the development cases (which included very extreme cases), and failing in < 2% of the cross-sectional cases and < 3% of the longitudinal case pairs in the large clinical dataset (despite clear artifacts on many scans and many cases with extensive lesion load).

The inter-scanner reproducibility between 1.5T and 3.0T, at 2.69% cross-sectionally and 2.08% via direct Jacobian, also compares favorably to whole-brain measures performed on high-resolution images. In a recent study, Chu et al. evaluated paired 1.5T and 3.0T scans repeated on the same subjects within 1 month (Chu et al., 2016). They performed SIENAX on each scan, and compared the resulting whole brain volume measurements. Their data showed a 3.37% average absolute difference between scans. The improved results from LVV likely reflect both the distortion protection afforded by closeness to the isocenter and the simpler geometry of the ventricles compared to cortical gyri. When comparing error rates to whole brain measures, it is also important to consider that ventricular volume expands at a higher relative rate of up to 3–5 times that of whole brain volume (Popescu

et al., 2013; Zivadinov et al., 2016b).

Although NeuroSTREAM metrics are in line with other methods, their precision, accuracy, and predictive value generally do not exceed these prior methods. This is as expected, since the goal of NeuroSTREAM is not to improve upon well-validated methods like VIENA in a research setting, but rather to expand the potential reach of automated atrophy quantification to a broader range of clinical exams. The findings of this study indicate that clinical-quality T2-FLAIR can indeed be used for meaningful and reliable brain atrophy measurements.

In addition to prospective clinical-quality scan analysis, there is potentially an important place for a method like that proposed here in the analysis of retrospective datasets. Many previous studies in MS have been performed with lesions as an MRI endpoint, and these studies have not always included high resolution T1-weighted images or used consistent full-head head coverage. Similarly, many latent clinical scans have not been used at all in a research setting. The ability to perform LVV measurement on all these datasets and to pool them together may eventually allow for large-scale data mining and informatics techniques to be brought to bear in providing more individually meaningful predictions in a clinical setting. For example, k nearest neighbor and kernel density techniques are simple and powerful non-parametric techniques that can match individuals for many variables simultaneously. However, these powerful techniques suffer from the "curse of dimensionality" and usually require far more data points than traditional MRI studies can provide.

While designing the NeuroSTREAM algorithm, many trade-offs were made between performance and robustness. For example, intensity Winsorization to minimize the influence of lesions is a relatively simplistic approach, whereas many automated lesion detection/classification algorithms have been devised and even used routinely in practice (Lladó et al., 2011). However, these approaches generally require multiple input images (e.g. T2-FLAIR and T1) and complex statistical models that must be tuned to specific scanners and imaging protocols. Therefore, incorporating these approaches would have limited the generalizability of the LVV algorithm and potentially increased the number of failed cases. Similarly, logistic regression is a basic technique for classification, potentially improved upon by more modern machine learning methods like SVD, decision trees, and others. However, logistic regression is a very robust technique, and the proposed method of applying it on a case-by-case basis allows for a fast and widely applicable means of assigning probabilities to individual image voxels.

An advantage of the proposed method is that it is fully automated. In theory, other methods for atrophy measurement like SIENAX and SPM are also fully automated. However, on scans with imperfect quality, manual corrections to de-skulling procedures are often required to avoid failed cases or erroneous data. By limiting itself to a more easily segmentable region, the proposed NeuroSTREAM technique can be more fully automated.

5. Limitations and future work

We have shown both here and previously that LVV is significantly correlated with both whole brain atrophy and clinical outcomes. However, it remains a proxy, with all the attendant caveats of any proxy measure. In particular, mild enlargement of the ventricles with unknown etiology is a relatively common incidental finding that may not reflect overall brain atrophy. This may therefore limit the applicability of single, cross-sectional LVV measurements to individual subjects in general, even outside the scope of the currently proposed algorithm.

Another potential issue is that of normalization. Normalizing brain volume measures to intracranial volume has improved inter-subject data comparison with other techniques, by correcting for non-pathological natural variations in brain size (Sanfilipo et al., 2004). For the current algorithm, we found more variability and introduction of analysis failures in estimating intracranial volume on low-quality T2-FLAIR scans than was gained by including it (once sex is taken into account as a factor). Therefore, we did not include it in the present work. However, future approaches may be able to provide a more accurate estimate of intracranial volume without sacrificing accuracy or robustness.

The method at present is also only applicable to T2-FLAIR scans. We chose this because it is the "lowest common denominator" of clinical MS MRI imaging protocols. Although this makes it generally and widely applicable, many individual researchers may have other scans available. With high-resolution research quality images, alternate approaches like SIENA/X, VIENA, and SPM are likely the best option. However, there may be a middle ground where low-resolution T1-weighted images, T2 images, or PD images may be present, and might provide valuable additional information for LVV segmentation. We are therefore working to extend the method to these additional image weightings. Indeed, there is also a potential value for even high-resolution T1 data, in providing a more accurate control for NeuroSTREAM measures taken in practice.

6. Conclusions

NeuroSTREAM is able to measure LVV accurately, precisely, and reliably on clinical-quality T2-FLAIR images, and changes in NeuroSTREAM-derived LVV relate to clinical outcomes. This provides an important complementary approach for performing meaningful atrophy analysis of real-world imaging datasets and clinical routine scans where no high-resolution T1-weighted images are available.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.nicl.2017.06.022.

References

Aljabar, P., et al., 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46 (3), 726–738.
Ashburner, J., Friston, K., 2000. Voxel-based morphometry—the methods. NeuroImage 11 (6), 805–821.
Avants, B.B., et al., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41.
Avants, B., et al., 2011. An open source multivariate framework for N-tissue segmentation with evaluation on public data. Neuroinformatics 9 (4), 381–400.
Avants, B., et al., 2012. Eigenanatomy improves detection power for longitudinal cortical change. In: Medical Image Computing and Computer-assisted Intervention: MICCAI … International Conference on Medical Image Computing and Computer-Assisted Intervention. 15(Pt 3). pp. 206–213.
Baillard, C., Hellier, P., Barillot, C., 2001. Segmentation of brain 3D MR images using level sets and dense registration. Med. Image Anal. 5 (3), 185–194.
Barra, V., Boire, J., 2001. Automatic segmentation of subcortical brain structures in MR images using information fusion. IEEE Trans. Med. Imaging 20 (7), 549–558.
Benedict, R.H.B.R., et al., 2006. Neocortical atrophy, third ventricular width, and cognitive dysfunction in multiple sclerosis. Arch. Neurol. 63 (9), 1301–1306.
Caramanos, Z., et al., 2010. Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. NeuroImage 49 (2), 1601–1611.
Chu, R., et al., 2016. Whole brain volume measured from 1.5 T versus 3T MRI in healthy subjects and patients with multiple sclerosis. (Journal of).
Collins, D.L., et al., 1995. Automatic 3-D model-based neuroanatomical segmentation. Hum. Brain Mapp. 208 (3), 190–208.
Coupé, P., et al., 2011. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage 54 (2), 940–954.
Dalton, C., Chard, D., Davies, G., 2004. Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. Brain.
De Stefano, N., et al., 2014. Clinical relevance of brain volume measures in multiple sclerosis. CNS Drugs 28 (2), 147–156.
Fisniku, L.K., et al., 2008. Gray matter atrophy is related to long term disability in multiple sclerosis. Ann. Neurol. 64 (3), 247–254.
Freeborough, P.A., Fox, N.C., 1997. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Trans. Med. Imaging 16 (5), 623–629.
Gorgolewski, K., et al., 2011. Nipype: a flexible, lightweight and extensible neuroimaging

data processing framework in python. Front. Neuroinform. 5, 13.
Havrdova, E., et al., 2009. Randomized study of interferon beta-1a, low-dose azathioprine, and low-dose corticosteroids in multiple sclerosis. Mult. Scler. 15 (8), 965–976 (Houndmills, Basingstoke, England).
Heckemann, R.A., et al., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33 (1), 115–126.
Heimann, T., Meinzer, H., 2009. Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal.
Julazadeh, A., et al., 2012. A novel automated approach for segmenting lateral ventricle in MR images of the brain using sparse representation classification and dictionary learning. In: International Conference on Information Science. IEEE, pp. 888–893.
Kalkers, N.F., 2002. Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype. Arch. Neurol. 59 (10), 1572–1576.
Kempton, M.J., et al., 2011. A comprehensive testing protocol for MRI neuroanatomical segmentation techniques: evaluation of a novel lateral ventricle segmentation method. NeuroImage 58 (4), 1051–1059.
Lladó, X., et al., 2011. Automated detection of multiple sclerosis lesions in serial brain MRI. Neuroradiology.
Lukas, C., Minneboo, A., Groot, V. De, 2010. Early central atrophy rate predicts 5 year clinical outcome in multiple sclerosis. J. Neurol.
Miller, D.H., et al., 2002. Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. Brain J. Neurol. 125 (Pt 8), 1676–1695.
Murray, T.J., 2005. Multiple Sclerosis: The History of a Disease, 1st ed. Demos Medical Publishing, New York.
Nakamura, K., et al., 2014. Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis. NeuroImage: Clinical 4, 10–17.
Patenaude, B., et al., 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage 56 (3), 907–922.
Pedregosa, F., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Perona, P., Malik, J., 1990. Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. 12 (7), 629–639.
Popescu, V., et al., 2013. Brain atrophy and lesion load predict long term disability in multiple sclerosis. J. Neurol. Neurosurg. Psychiatry 84 (10), 1082–1091.
Rifai, H., et al., 2000. Segmentation of the skull in MRI volumes using deformable model and taking the partial volume effect into account. Med. Image Anal. 4 (3), 219–233.
Sanfilipo, M., et al., 2004. Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. Neuroimage.
Santos, R., Weinstock-Guttman, B., 2006. Dynamics of interferon-β modulated mRNA biomarkers in multiple sclerosis patients with anti-interferon-β neutralizing antibodies. (Journal of).
Simon, J.H., et al., 2006. Standardized MR imaging protocol for multiple sclerosis: consortium of MS centers consensus guidelines. AJNR Am. J. Neuroradiol. 27 (2), 455–461.
Smith, S., et al., 2001. Normalized accurate measurement of longitudinal brain change. J. Comput. 25 (3), 466–475.
Smith, S.M., et al., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17 (1), 479–489.
Summers, M., et al., 2008. Cognitive impairment in multiple sclerosis can be predicted by imaging early in the disease. J. Neurol. Neurosurg. Psychiatry 79 (8), 955–958.
Tustison, N., Avants, B., 2010. N4ITK: improved N3 bias correction. Med. Imaging 29 (6), 1310–1320.
Uher, T., et al., 2014. Longitudinal MRI and neuropsychological assessment of patients with clinically isolated syndrome. J. Neurol. 261 (9), 1735–1744.
Uher, T., Horakova, D., Kalincik, T., 2015. Early magnetic resonance imaging predictors of clinical progression after 48 months in clinically isolated syndrome patients treated with intramuscular interferon β-1a. (European journal of).
Vrenken, H., et al., 2014. Validation of the automated method VIENA: an accurate, precise, and robust measure of ventricular enlargement. Hum. Brain Mapp. 35 (4), 1101–1110.
Wang, H., Yushkevich, P.A., 2013. Multi-atlas segmentation with joint label fusion and corrective learning-an open source implementation. Front. Neuroinform. 7, 27.
Wilcox, R., Wilcox, R., 2005. Trimming and Winsorization. In: Encyclopedia of Biostatistics. John Wiley & Sons, Ltd., Chichester, UK.
Xia, Y., et al., 2004. A knowledge-driven algorithm for a rapid and automatic extraction of the human cerebral ventricular system from MR neuroimages. NeuroImage 21 (1), 269–282.
Yushkevich, P.A., et al., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 31 (3), 1116–1128.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20 (1), 45–57.
Zivadinov, R., Ramasamy, D., et al., 2016a. Cerebral microbleeds in multiple sclerosis evaluated on susceptibility-weighted images and quantitative susceptibility maps: a case-control study. Radiology.
Zivadinov, R., Uher, T., Hagemeier, J., 2016b. A serial 10-year follow-up study of brain atrophy and disability progression in RRMS patients. Mult. Scler.
Zivadinov, R., et al., 2016c. 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., Khan, N., Medin, J., 2017. An observational study to assess brain MRI change and disease progression in multiple sclerosis clinical practice—the MS-MRIUS study. J. Neuroimaging. 27 (3), 339–347.