Magnetic Resonance Imaging 34 (2016) 455–461

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.mrijournal.com

Atlas based brain volumetry: How to distinguish regional volume changes due to biological or physiological effects from inherent noise of the methodology

Roland Opfer [a,b,*], Per Suppa [b], Timo Kepp [b], Lothar Spies [b], Sven Schippling [a], Hans-Jürgen Huppertz [c]; the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Neuroimmunology and Multiple Sclerosis Research, Department of Neurology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
[b] jung diagnostics GmbH, Hamburg, Germany
[c] Swiss Epilepsy Centre, Zurich, Switzerland

ARTICLE INFO

Article history:
Received 23 November 2015
Accepted 18 December 2015
Available online xxxx

Keywords:
Magnetic resonance imaging
Atlas based brain volumetry
Dartel
Intrascanner variability
Repeat scanning
Elastic image registration

ABSTRACT

Fully-automated regional brain volumetry based on structural magnetic resonance imaging (MRI) plays an important role in quantitative neuroimaging. In clinical trials as well as in clinical routine multiple MRIs of individual patients at different time points need to be assessed longitudinally. Measures of inter- and intrascanner variability are crucial to understand the intrinsic variability of the method and to distinguish volume changes due to biological or physiological effects from inherent noise of the methodology.

To measure regional brain volumes an atlas based volumetry (ABV) approach was deployed using a highly elastic registration framework and an anatomical atlas in a well-defined template space. We assessed inter- and intrascanner variability of the method in 51 cognitively normal subjects and 27 Alzheimer dementia (AD) patients from the Alzheimer's Disease Neuroimaging Initiative by studying volumetric results of repeated scans for 17 compartments and brain regions.

Median percentage volume differences of scan–rescans from the same scanner ranged from 0.24% (whole brain parenchyma in healthy subjects) to 1.73% (occipital lobe white matter in AD), with generally higher differences in AD patients as compared to normal subjects (e.g., 1.01% vs. 0.78% for the hippocampus). Minimum percentage volume differences detectable with an error probability of 5% were in the one-digit percentage range for almost all structures investigated, with most of them being below 5%. Intrascanner variability was independent of magnetic field strength. The median interscanner variability was up to ten times higher than the intrascanner variability.

1. Introduction

A very common approach in quantitative MRI to measure regional brain volumes is atlas based volumetry (ABV), which classifies the image on a voxel level into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) compartments and warps the resulting tissue probability maps into a well-defined template space using elastic image registration (normalization). In a second step it uses an atlas of predefined regions of interest in that template space to extract regional brain volumes [15]. The method has already been applied to neurodegenerative diseases in both cross-sectional and longitudinal studies, demonstrating its sensitivity to detect volume changes within 6 month intervals [11,12,14,17].

Obviously, volumetric accuracy depends critically on registration performance: the more precisely the registration in the template space matches individual anatomy the more accurate volumetric results are obtained [18]. To enhance registration accuracy high-dimensional diffeomorphic image registration approaches, such as diffeomorphic anatomical registration through exponentiated Lie algebra (DARTEL) [2], have been proposed. DARTEL is part of the Statistical Parametric Mapping (SPM) software (www.fil.ion.ucl.ac.uk/spm) and is a highly elastic registration method resulting in a more precise registration which has been confirmed in numerous studies [7,22,25,30].

* Corresponding author at: Martinistrasse 52, D-20251 Hamburg, Germany. Tel.: +49 40/81 97 6967 12; fax: +49 40/81 97 6967 17.
E-mail address: roland.opfer@jung-diagnostics.de (R. Opfer).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

http://dx.doi.org/10.1016/j.mri.2015.12.031
0730-725X/

In the past, inter- and intrascanner variability of ABV was studied by Huppertz and co-workers [15] who deployed the unified segmentation algorithm [3] of SPM5 for normalization. MRI data from a single healthy volunteer scanned three times on six different scanners were used to calculate the variability of volumetric results for a number of GM and WM regions. Furthermore, Eggert and coworkers [8] investigated the reliability of whole brain GM segmentation deploying five current automated segmentation pipelines using a larger public database of real images from a single scanner. The present study extends this previous work and provides a comprehensive assessment of intrascanner variability of ABV for whole brain and tissue compartments (GM, WM) as well as cortical and subcortical brain structures using DARTEL. In contrast to the study of Huppertz et al. [15] the assessment is based not only on one single healthy volunteer but on 78 subjects comprising cognitively normal subjects as well as patients with Alzheimer's dementia (AD). As opposed to Eggert et al. we assessed the variability of volumetric results not only for GM but also for WM and various subregions. The results of this study help to distinguish volume changes due to biological or physiological effects from inherent noise of the methodology and are an important basis for future clinical studies using SPM and DARTEL algorithms for atlas-based MRI volumetry.

2. Material and methods

MRI data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public private partnership. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California–San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada.

2.1. Image acquisition

Three-dimensional (3D) T1-weighted magnetization prepared rapid gradient echo (MPRAGE) scans of 51 cognitively normal subjects (Normal) and 27 Alzheimer dementia patients (AD) were included in this study and unprocessed MRI scans were downloaded from the ADNI repository (www.adni-info.org). The data have been acquired at different imaging centers (ADNI study centers) using diverse MR scanners. Prior to study initiation acquisition protocols were harmonized to achieve a similar image quality (contrast-to-noise, spatial resolution, resistance to artifact, speed, etc.) across scanner platforms [16]. While acquisition parameters, such as echo time, repetition time, inversion time etc. varied depending on the scanner type, all images have slice thickness of 1.2 mm in common. The in-plane resolution was slightly higher (average 1.0 mm, range 0.94–1.30 mm). The population of normal subjects (female: 64.7%; mean age: 75.7 $\pm$ 4.9 years) and AD patients (female: 60.7%; mean age: 74.6 $\pm$ 8.7 years) in this study is identical to the corresponding subgroups studied in Wolz et al. [31] for assessing the inter- and intrascanner variability of a hippocampal volume quantification algorithm. Only one of the AD patients (ADNI subject 133_S_1170) had to be excluded from our study (in comparison to the cohort investigated in Wolz et al.) because the field of view did not cover the whole skull in one of the 4 scans provided by ADNI. Each subject was scanned two times on two different scanner platforms at 1.5 T and 3 T field strengths within a period of a few weeks yielding 312 MRIs in total. The two scans per scanner were acquired back-to-back during a single imaging session.

2.2. Image processing

All images were processed on MATLAB Version 14 (R2014b, The Mathworks, Natick, USA) using SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK) (www.fil.ion.ucl.ac.uk/spm). For segmentation of the brain into GM, WM and CSF components we used the unified segmentation engine of SPM12 with default parameters, except that the image data were sampled every 2 mm instead of the default 3 mm as proposed by Herron and coworkers [13]. Normalization of brains to Montreal Neurological Institute (MNI) template space was performed using diffeomorphic anatomical registration through exponentiated Lie algebra (DARTEL), a high-dimensional elastic image registration technique [2]. GM and WM component images in native space resulting from unified segmentation were used as an input to the DARTEL process. We ran DARTEL with default parameters and 'with existing templates'. Targets for normalization were the IXI555 templates provided in the VBM12 toolbox by C. Gaser (http://dbm.neuro.uni-jena.de/vbm) which are already in MNI template space. The normalized component images were 'modulated' in order to preserve the overall volume. In previous SPM releases this meant that the determinant of the Jacobian of the transformation field was locally applied in order to preserve the volume. In the new SPM12 release, however, the warped data are not scaled by the Jacobian determinants when generating the 'modulated' data. Instead, the original voxels are projected into their new location in the warped images. This exactly preserves the tissue count, but has the effect of introducing aliasing artifacts (such as ripples and even holes where voxel values in the original native image are missing), especially if the original data are at a lower resolution than the warped images (cf. SPM12 manual, page 46). To circumvent this problem of aliasing artifacts the component images in native space were rescaled to 0.5 mm voxel resolution before normalization to MNI space.

2.3. Atlas based volumetry

Volumetric measures of intracranial compartments and brain structures were derived from the normalized and modulated component images processed with DARTEL. Volumetric measures of brain structures were calculated by a voxel-by-voxel multiplication and subsequent integration of normalized and modulated component images (GM, WM or CSF) with predefined binary masks from different brain atlases. Binary masks for the frontal, parietal, occipital and temporal lobe were derived from a cerebral lobe atlas defined in the Montreal Neurological Institute (MNI) template space and published by Fonov et al. [10]. For the hippocampus, which comprises cornus ammonis, fascia dentata and subiculum substructures as defined by Amunts et al. [1], a binary mask was taken from a freely available toolbox [9]. Binary masks for the caudate nucleus and putamen were derived from the LPBA40 atlas [28]. The thalamus mask and the mask for lateral ventricles were taken from wfu pickatlas [20]. A 30% margin was added to the lateral ventricle masks to accommodate the wide variety of ventricle sizes. The corpus callosum mask was composed of binary masks from subregions, i.e. genu, body and splenium, which were taken from the ICBM-DTI-81 white-matter labels atlas [24,27]. All masks were interpolated to a 1.5 mm isotropic resolution, which matches the resolution of the modulated and normalized component images. Lobar volumes were calculated for both GM and WM. For the caudate nucleus and hippocampus only GM partial volumes were estimated. Similarly for corpus callosum only the WM volume was assessed. Thalamic and putamen volumes were estimated using the sum of

GM and WM components. In total, 17 different brain structures and compartments were evaluated.

2.4. Volumetric assessment

The ABV method was applied to each scan (4 scans for each of the 78 subjects) to determine the volumes of all compartments and substructures investigated. The results were split into 4 groups: 1) Normal 1.5 T (n = 51), 2) Normal 3 T (n = 51), 3) AD 1.5 T (n = 27), 4) AD 3 T (n = 27)). For each group the mean and standard deviation of the volumetric measurements were determined.

2.5. Interscanner variability of volumetric results

With two 1.5 and 3 T scans for each patient we can build 4 pairs of 1.5 T versus 3 T scans. The volumetric results for each pair, V_1 and V_2, of the two measurements for a specific structure were quantitatively compared by computing the *absolute percentage volume difference* as defined by

$$\Delta_{1.5Tvs3T} := 200 * |V_1 - V_2| / (V1 + V2).$$

For each patient and each structure we obtained four $\Delta_{1.5Tvs3T}$ measurements. Subsequent calculations of mean volumes and standard deviations of investigated brain structures/compartments and of absolute percentage volume differences between different scanners were based on all possible pairings of 1.5 and 3 T scans.

2.6. Intrascanner variability of volumetric results

The same metric was used to compare volumetric results for each pair of back-to-back scans (scan–rescan) on the same scanner and in the same subject. For each patient and each structure we obtained two difference measurements ($\Delta_{1.5Tvs1.5T}$ and Δ_{3Tvs3T}).

Throughout this paper the absolute percentage volume difference is used as a metric for inter- and intrascanner variability. The same metric was used in similar studies [6,23,31]. Obviously, in an ideal setting Δ should be zero. In reality, however, due to small variations in the acquisition process and sensitivity of the applied algorithms to these variations, Δ may deviate from zero. The subject's disease characteristics as well as the scanner field strength may have an additional influence.

2.7. Statistical analysis

Since the collections of absolute percentage volume differences are not normally distributed (Δ is always positive) non-parametric methods were used to describe the distribution (as recommended in [5,6]). For each group and structure the 25th, 50th, 75th, and 95th percentiles are reported. We statistically compared the intrascanner variability between 1.5 and 3 T scans deploying the Wilcoxon signed rank test, the nonparametric equivalent to the paired t-test. The null hypothesis assumes that $\Delta_{1.5Tvs1.5T} - \Delta_{3Tvs3T}$ comes from a distribution with zero median. The same approach was used to compare the intrascanner variability (combining 1.5 and 3 T scans) between healthy subjects and AD patients. For the comparison the Wilcoxon rank-sum test was deployed because the scan–rescan results of these subgroups (AD and healthy subject) are statistically independent. The statistical analysis was performed using the MATLAB 2014a Statistics and Machine Learning Toolbox.

3. Results

3.1. Interscanner variability of volumetric results based on 1.5 T and 3 T scans

For each of the four groups (Normal vs. AD; 1.5 vs. 3 T), Table 1 shows mean volumes and standard deviations of all investigated brain structures and compartments as determined by ABV. As a measure of interscanner variability, the median, the 75th, and the 95th percentile of the absolute percentage volume differences ($\Delta_{1.5Tvs3T}$) between the 1.5 and 3 T scan are listed. The median of $\Delta_{1.5Tvs3T}$ ranged from 1.8% for the brain parenchyma (BP) in normal subjects to 11.5% for occipital lobe white matter in the AD group.

3.2. Intrascanner variability of volumetric results

Fig. 1 shows the intrascanner variability of volumetric results for 1.5 T ($\Delta_{1.5Tvs1.5T}$) and 3 T scans (Δ_{3Tvs3T}) by depicting the median and interquartile ranges of absolute percentage volume differences for each of the four groups separately. In Fig. 2 the same data are shown but grouped differently (normals versus AD). Overall the median of the absolute percentage volume differences between back-to-back scans ranged from 0.24% (BP) to 1.06% (occipital WM) for normal subjects, and from 0.29% (BP) to 1.73% (again for occipital WM) in AD patients. There were no statistically significant differences in intrascanner variability between the 1.5 and 3 T scans. However, intrascanner variability was generally higher in AD patients compared to normal subjects. The difference reached statistical significance for the hippocampus, white matter occipital lobe and the corpus callosum. Table 2 summarizes intrascanner variabilities of ABV in detail, i.e. the median, the 75th, and the 95th percentile of the absolute percentage volume differences for both normal subjects and AD patients (since intrascanner variabilities did not depend on field strength results for 1.5 T and 3 T were combined).

4. Discussion

In this study we investigated MRI inter- and intrascanner variability of an ABV method with the following main findings:

- The majority of investigated structures had median intrascanner variabilities of less than 1% between scans, being lowest in case of the brain parenchymal (BP) volume (0.24% and 0.29% in normal subjects and AD patients, respectively). The only outlier, but still with a median percentage volume difference below 2% was the occipital WM (cf. Table 2) which is the smallest of the investigated lobar structures and thus perhaps more prone to be affected by the intrinsic measurement variability. Overall, these results indicate a favorable reliability of ABV for repeated MRI measurements on the same scanner.
- The median interscanner variability was up to 10 times higher than the intrascanner variability. This confirms results from previous studies that automatic volumetry highly depends on the scanner platform [19].

Intrascanner variability was independent of magnetic field strength. However, in AD patients, intrascanner variability was generally higher than in normal subjects. While this difference was relatively small for median values, it became most obvious for the 95th percentile, which more than doubled in AD patients for about half of the investigated structures (Table 2). This indicates that outliers were responsible for the increased intrascanner variability. These outliers are most probably the result of motion artifacts in MR images, a hypothesis which is supported when visually inspecting MR images of AD patients with especially high intrascanner variability. Fig. 3 shows example images of two AD patients in

Table 1
Comparison of volumetric results: Normal 1.5 T (n = 51), Normal 3 T (n = 51), AD 1.5 T (n = 27), and AD 3 T (n = 27). Table shows mean and standard deviation (in brackets) of the volumes (in ml) of the compartments and subregions. Columns next to volume values show median, the 75th and the 95th percentile of $\Delta_{1.5Tvs3T}$.

	Normal (n = 51)					AD (n = 27)				
	1.5 T	3 T	$\Delta_{1.5Tvs3T}$			1.5 T	3 T	$\Delta_{1.5Tvs3T}$		
			median	75th	95th			median	75th	95th
GM	573.9 [61.9]	602.3 [64.6]	4.5	6.9	10.5	525.0 [87.4]	542.9 [89.2]	5.9	7.3	12.6
WM	429.1 [50.2]	416.0 [48.1]	3.4	5.3	9	414.7 [67.8]	400.7 [67.2]	4	5.5	7.9
BP	1003.0 [94.3]	1018.3 [97.5]	1.8	3	5.4	939.7 [147.0]	943.6 [141.4]	2.2	3.2	5.8
Frontal GM	181.4 [21.4]	191.7 [21.9]	5	8.4	12.1	169.9 [27.8]	176.6 [28.2]	6.4	9	11.8
Temporal GM	133.0 [15.8]	138.7 [16.9]	4.1	6.2	9.6	114.7 [21.3]	118.3 [20.7]	4.9	7.6	14.7
Parietal GM	85.0 [9.8]	89.4 [10.2]	5.9	9.2	13.5	76.9 [13.1]	79.6 [13.4]	5.9	8.3	14.4
Occipital GM	50.7 [7.1]	55.4 [7.4]	9.6	11.7	18	46.3 [11.3]	50.2 [11.3]	8.4	11.8	19.7
Lateral Ventricles	46.7 [15.7]	45.0 [15.0]	3.4	5.3	9.4	69.7 [36.7]	69.2 [38.6]	3.5	7.7	13
Hipp	6.4 [0.8]	6.6 [0.8]	2.6	5.1	11.9	5.2 [1.0]	5.4 [1.1]	5.1	7.6	15.4
Frontal WM	154.1 [19.6]	148.9 [18.3]	4.3	6.2	9.7	152.2 [27.4]	146.7 [26.3]	3.8	6.3	8.7
Temporal WMT	74.6 [9.4]	73.8 [9.5]	1.9	3.6	6.1	69.2 [13.2]	67.7 [12.7]	2.4	4.7	8.5
Parietal WM	84.5 [10.6]	80.6 [10.4]	5.3	7.8	11.8	81.4 [12.9]	77.6 [13.6]	5.8	8.1	10.2
Occipital WM	38.1 [5.4]	34.2 [5.1]	10.8	15.4	19.1	37.2 [7.3]	33.1 [6.6]	11.5	15.7	20.7
CC	20.2 [2.5]	20.5 [2.5]	2.1	4.4	6.9	19.6 [3.2]	20.1 [3.5]	2.7	5	8.6
NC	8.4 [1.0]	9.0 [1.2]	6.7	9.2	16.2	8.4 [1.6]	8.8 [1.4]	8.2	12.7	21.6
Thalamus	10.5 [0.9]	11.1 [1.0]	5.7	7.5	10.5	9.7 [1.2]	10.3 [1.3]	6.2	8.3	10.9
Putamen	18.8 [1.6]	19.1 [1.6]	1.8	3.3	5.5	18.2 [2.8]	18.4 [2.7]	1.6	3.4	5.5

Legend: GM = gray matter, WM = white matter, BP = brain parenchyma, CC = corpus callosum, NC = nucleus caudate, Hipp = hippocampus.

whom movement artifacts in one of the scans were accompanied by increased variability of volumetric results between scans. AD patients but also patients with other neurological diseases may be less capable than normal subjects of lying still during the acquisition process. It does not come as a surprise that this can negatively affect intrascanner variability of volumetric measurements. It is important

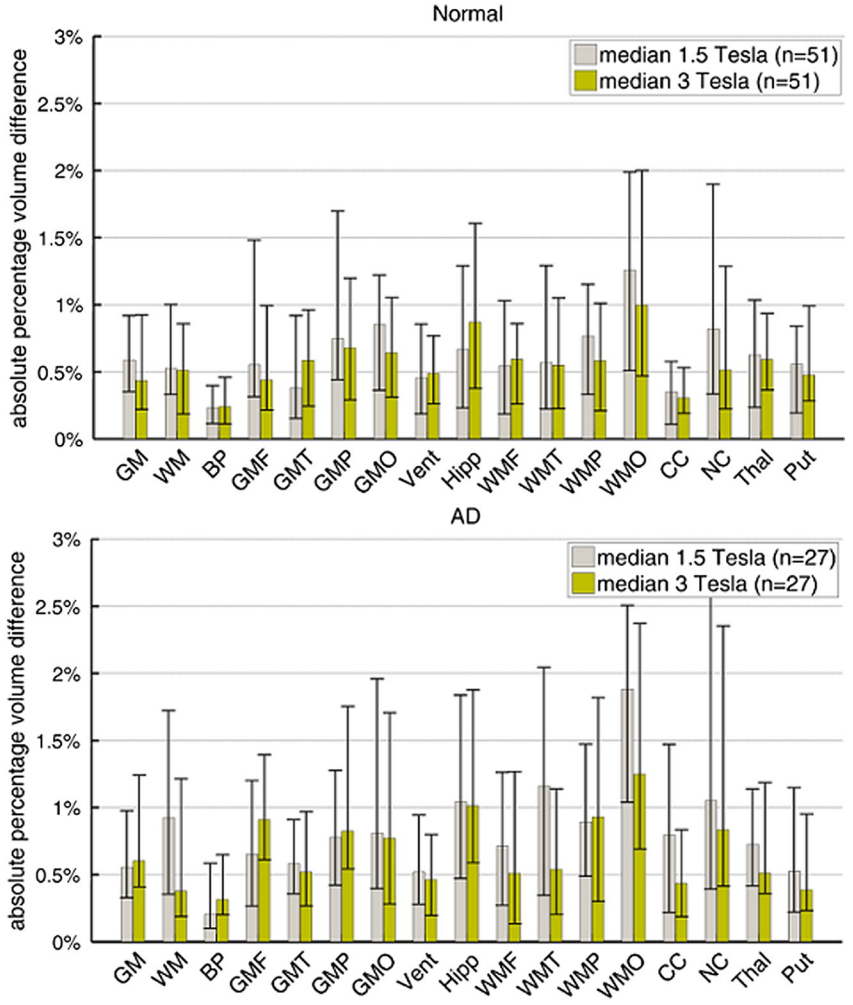

Fig. 1. The bar plots illustrate the difference in scan–rescan variability between 1.5 T ($\Delta_{1.5Tvs1.5T}$) and 3 T (Δ_{3Tvs3T}) using 51 cognitively normal subjects and 27 AD patients. The interquartile range (25th percentile, 75th percentile) is illustrated by the lower and the upper whisker. Statistically significant ($p < 0.05$) differences are marked with a blue asterisk.

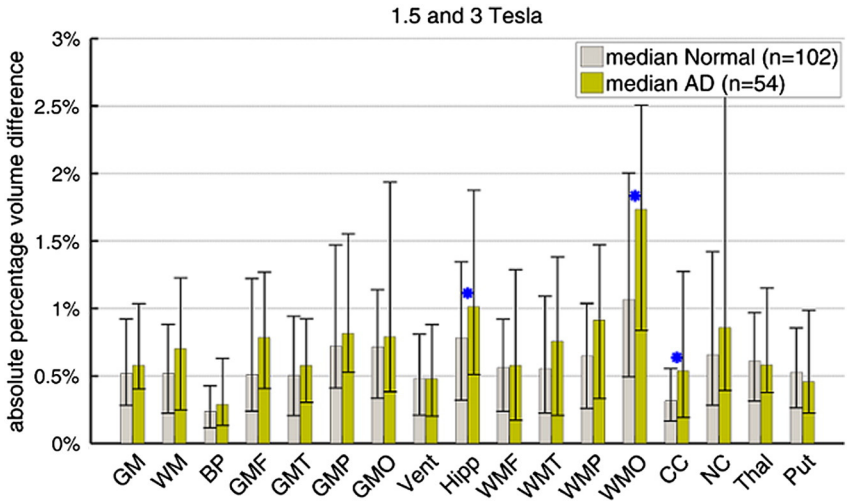

Fig. 2. The bar plots illustrate the difference in scan–rescan variability between the 27 Alzheimer (AD) patients and the 51 cognitively normal subjects (combining $\Delta_{1.5Tvs1.5T}$ and Δ_{3Tvs3T}). The interquartile range (25th percentile, 75th percentile) is illustrated by the lower and the upper whisker. Statistically significant ($p < 0.05$) differences are marked with a blue asterisk.

to be aware of this since many studies are not conducted in healthy subjects but in cognitively and/or physically impaired patients.

Table 2 summarizes the intrascanner variability for ABV methods. The table may help users to estimate the minimum absolute percentage volume difference between two MRI scans of the same subject necessary to detect a significant volume change beyond the level of intrinsic measurement variability. For example, a hippocampal volume change of more than 5.41% between two MRI scans using the same scanner can be regarded as significant with an error probability of 5% (cf. last column of Table 2, i.e. the 95th percentiles of AD patients).

4.1. Comparison with other methods and studies

Smith et al. [29] assessed intrascanner variability of BP volume for a fully automated method (called SIENA) of longitudinal change analysis. SIENA is frequently being used in multiple sclerosis drug trials and is considered as a gold standard in assessing longitudinal BP volume changes [26]. A median absolute error of 0.2% has been reported in this paper which compares well to the median absolute percentage volume difference in our analysis (0.24%) (Table 2).

In the study by Cover and coworkers [6] ADNI data were used to assess intrascanner variability of SIENAX which is the cross-sectional version of the SIENA. The authors reported a median error of 0.92% (for healthy subjects), which is almost four times as high as suggested by the results of our study (0.24%).

Eggert and coworkers [8] systematically explored intrascanner variability of GM for 5 different segmentation engines. They evaluated a total of 20 scans from 10 healthy subjects as provided by the publicly available OASIS database [21]. Each subject was scanned twice within 12 days. They calculated 0.6% mean volume difference for GM for "new segment" which is part of SPM8 and which became the standard segmentation procedure under SPM12 (deployed in our paper). This value is in quantitative agreement with our results (0.52%) (Table 2).

In the study of Wolz et al. [31] the intrascanner variability of an automated hippocampus volumetry method (LEAP algorithm) was investigated in the same ADNI data that were used in the present study. The median percentage volume differences of the scan pairs compare to our results (for the hippocampus) as follows: AD patients at 1.5 and 3 T: 1.56% and 1.06%, respectively (LEAP) vs. 1.04% and 1.01%, respectively (ABV); 75th percentile 2.75% and 2.38% (LEAP) vs. 1.83% and 1.88%, respectively (ABV). Cognitively normal subjects at 1.5 and 3 T: 0.93% and 0.89%, respectively (LEAP) vs. 0.66% and 0.86%, respectively (ABV); 75th percentile 1.91% and 1.97% (LEAP) vs. 1.28% and 1.60%, respectively (ABV). Overall, hippocampus volumetry using ABV seems to have slightly lower intrascanner variability than LEAP.

If a patient receives a follow-up scan and a certain brain volume loss is measured it is necessary to distinguish between true biological/pathological volume changes and changes due to the intrinsic variability of the methodology applied. The intrascanner variability can be influenced by several factors, such as effects of scanner and image quality, individual brain anatomy, or the ability of a patient to lie still during image acquisition. For a specific patient and a specific scanner the intrascanner variability is unknown. However, since the ADNI data were acquired in many different centers with different scanners and comprises data of many healthy subjects as well as patients, the range of intrascanner variabilities given in Table 2 can be regarded as an estimate of intrascanner variabilities occurring in clinical routine. If the ABV method proposed in this study is deployed, the results of Table 2, as indicated above,

Table 2
Intrascanner variability for ABV: the median, the 75th, and the 95th percentile of the absolute percentage volume differences of the scan–rescans ($\Delta_{1.5Tvs1.5T}$ and Δ_{3Tvs3T}) for cognitively normal subjects and patients with Alzheimer's disease (AD).

	Normal $\Delta_{1.5Tvs1.5T}$ and Δ_{3Tvs3T}			AD $\Delta_{1.5Tvs1.5T}$ and Δ_{3Tvs3T}		
	median	75th	95th	median	75th	95th
GM	0.52	0.92	2.67	0.58	1.04	2.40
WM	0.52	0.88	1.96	0.70	1.23	4.59
BP	0.24	0.43	1.28	0.29	0.63	2.59
Frontal GM	0.51	1.22	3.24	0.78	1.27	4.32
Temporal GM	0.50	0.94	2.21	0.58	0.92	2.91
Parietal GM	0.72	1.47	5.21	0.81	1.55	5.92
Occipital GM	0.71	1.14	2.69	0.79	1.94	4.29
Lateral Ventricles	0.48	0.81	2.12	0.48	0.88	2.54
Hipp	0.78	1.34	2.36	1.01	1.88	5.41
Frontal WM	0.56	0.92	2.02	0.58	1.29	4.54
Temporal WMT	0.55	1.09	2.32	0.76	1.38	7.98
Parietal WM	0.65	1.04	2.79	0.91	1.47	4.42
Occipital WM	1.06	2.00	4.15	1.73	2.51	5.66
CC	0.31	0.55	1.97	0.54	1.27	4.09
NC	0.66	1.42	4.33	0.86	2.68	10.61
Thalamus	0.61	0.97	1.96	0.58	1.15	3.96
Putamen	0.53	0.85	1.92	0.46	0.98	2.13

Legend: GM = gray matter, WM = white matter, BP = brain parenchyma, CC = corpus callosum, NC = nucleus caudate, Hipp = hippocampus.

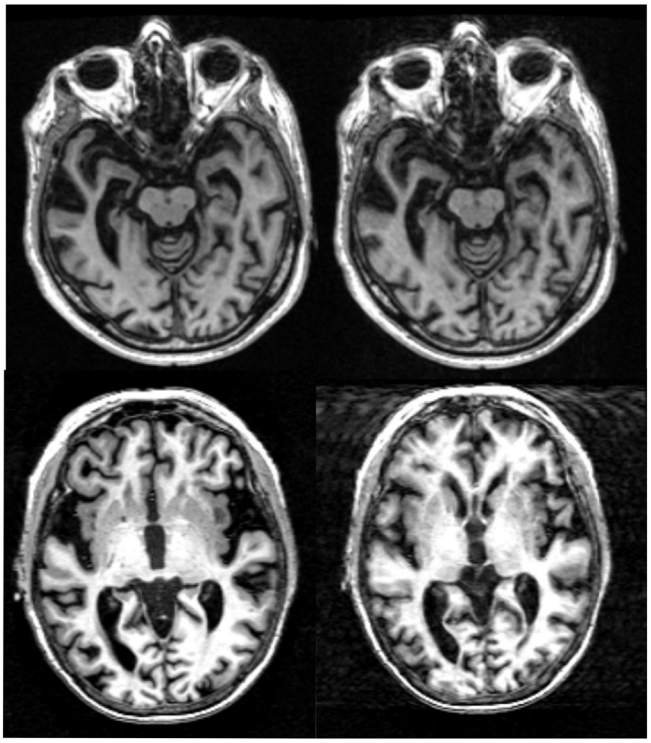

Fig. 3. Axial slices of two Alzheimer patients scanned twice with 3 T machines. The second scan (right column) shows strong motion artifacts. For these two patients the percentage volume differences of the white matter between scan and rescan were 5.2% (first row) and 12.58% (second row), respectively.

can help to assess volume changes beyond noise levels at single subject level. The 95th percentiles in this table can be regarded as the minimum absolute percentage volume difference between two MRI scans necessary to assume a significant volume change with an error probability of 5%. Based on the 95th percentiles of AD patients, volume changes from 2.4% (for GM) up to 10.6% (for caudate nucleus) would be required, with a median of 4.3% for all brain structures. Naturally, these thresholds would decrease for group studies, depending on the sample size. However, using the 95th percentiles determined in the AD patients of this study might also be too conservative considering that apparently MRIs with movement artifacts have not been excluded from the ADNI database (Fig. 3). We deliberately chose to use all available data of AD patients but with a stricter quality control the thresholds for significance could be even lower. Furthermore, for patients with expectably lesser degree of brain atrophy (e.g. patients with multiple sclerosis) it might be appropriate to refer to the 95th percentiles of cognitively normal subjects in this study. The thresholds for significance would then range between 1.3% and 4.3% (again for BP and caudate nucleus, respectively), with a median of 2.3%. It should be noted that (except for the caudate nucleus in AD patients) all thresholds for significance are in the one-digit percentage range. For most structures the thresholds are even below 5% which is in the range of the typical annual volume loss expected in AD patients [4]. These appear to be very favorable results, especially when considering that these thresholds are valid at a single patient level which renders them useful in clinical routine. In addition, the underlying volume changes are hardly discernable by pure visual inspection of MR images.

4.2. Methodological considerations and limitations

The calculation of intrascanner variabilities in this study was based on scans acquired back-to-back during a *single* imaging session. However, variabilities may increase with repeated sessions due to changes in subject positioning, differences in pre-scan and shim settings, and magnetic field drift. Morey et al., e.g., showed that the reliability of volume measures for an interscan interval of 1 h was higher than for an interval of 1 week [23]. The potential influence of these additional sources of variability could not be assessed in this study. Thus, our results may underrate the scan–rescan variability of scans acquired in *different* sessions.

It should be noted that without “ground truth” for volumetric measures, this study was not intended to investigate the anatomical accuracy of ABV. However, the accuracy of DARTEL has already been shown in numerous studies [7,22,25,30].

5. Conclusion

The results of this study indicate that fully-automated ABV deploying high-dimensional image registration techniques with a large number of degrees of freedom, such as DARTEL in SPM12, promising improved anatomical accuracy, is robust with regard to intrascanner variability and hence suitable to be included into future clinical applications. Median absolute percentage volume differences of less than 1% between scans and rescans using the same scanner were determined for the majority of investigated structures, with higher values in AD patients as compared to normal subjects. Intrascanner variability did not depend on magnetic field strength. As a consequence the minimum absolute percentage volume changes detectable for an error probability of 5% by ABV in a single subject measured repeatedly using the same MR scanner are in the one-digit percentage range for almost all structures, for most of them even below 5%. The interscanner variability was up to ten times higher than the intrascanner variability and therefore, longitudinal volume changes can only be assessed when baseline and follow-up are acquired on the same scanner.

Acknowledgments

Sven Schippling is supported by the Clinical Research Priority Program (CRPP) of the University of Zurich, Zurich, Switzerland as well as the Betty and David Koetser Foundation for Brain Research, Zurich, Switzerland. Hans-Jürgen Huppertz is supported by the Swiss Epilepsy Foundation, Zurich, Switzerland.

Data collection and sharing for this project were funded by the Alzheimer's disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; ; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1] Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, et al. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat Embryol* (*Berl*) 2005;210(5–6):343–52. http://dx.doi.org/10.1007/s00429-005-0025-5.

[2] Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage 2007;38(1):95–113. http://dx.doi.org/10.1016/j.neuroimage.2007.07.007.

[3] Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005;26(3):839–51. http://dx.doi.org/10.1016/j.neuroimage.2005.02.018.

[4] Barnes J, Bartlett JW, van de Pol LA, Loy CT, Scahill RI, Frost C, et al. A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. Neurobiol Aging 2009;30(11):1711–23. http://dx.doi.org/10.1016/j.neurobiolaging.2008.01.010.

[5] Cover KS, van Schijndel RA, Popescu V, van Dijk BW, Redolfi A, Knol DL, et al. The SIENA/FSL whole brain atrophy algorithm is no more reproducible at 3 T than 1.5 T for Alzheimer's disease. Psychiatry Res 2014;224(1):14–21. http://dx.doi.org/10.1016/j.pscychresns.2014.07.002.

[6] Cover KS, van Schijndel RA, van Dijk BW, Redolfi A, Knol DL, Frisoni GB, et al. Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans. Psychiatry Res 2011;193(3):182–90. http://dx.doi.org/10.1016/j.pscychresns.2011.02.012.

[7] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage 2011;56(2):766–81. http://dx.doi.org/10.1016/j.neuroimage.2010.06.013.

[8] Eggert LD, Sommer J, Jansen A, Kircher T, Konrad C. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. PLoS One 2012;7(9), e45081. http://dx.doi.org/10.1371/journal.pone.0045081.

[9] Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 2005;25(4):1325–35. http://dx.doi.org/10.1016/j.neuroimage.2004.12.034.

[10] Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins DL. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage 2009;47(*Supplement 1*(0)):S102. http://dx.doi.org/10.1016/S1053-8119(09)70884-5.

[11] Frings L, Mader I, Landwehrmeyer BG, Weiller C, Hull M, Huppertz HJ. Quantifying change in individual subjects affected by frontotemporal lobar degeneration using automated longitudinal MRI volumetry. Hum Brain Mapp 2012;33(7):1526–35. http://dx.doi.org/10.1002/hbm.21304.

[12] Frings L, Yew B, Flanagan E, Lam BY, Hull M, Huppertz HJ, et al. Longitudinal grey and white matter changes in frontotemporal dementia and Alzheimer's disease. PLoS One 2014;9(3), e90814. http://dx.doi.org/10.1371/journal.pone.0090814.

[13] Herron TJ, Kang X, Woods DL. Automated measurement of the human corpus callosum using MRI. Front Neuroinform 2012;6:25. http://dx.doi.org/10.3389/fninf.2012.00025.

[14] Hoglinger GU, Huppertz HJ, Wagenpfeil S, Andres MV, Belloch V, Leon T, et al. Tideglusib reduces progression of brain atrophy in progressive supranuclear palsy in a randomized trial. Mov Disord 2014;29(4):479–87. http://dx.doi.org/10.1002/mds.25815.

[15] Huppertz HJ, Kröll-Seger J, Kloppel S, Ganz RE, Kassubek J. Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. Neuroimage 2010;49(3):2216–24. http://dx.doi.org/10.1016/j.neuroimage.2009.10.066.

[16] Jack Jr CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27(4):685–91. http://dx.doi.org/10.1002/jmri.21049.

[17] Kassubek J, Pinkhardt EH, Dietmaier A, Ludolph AC, Landwehrmeyer GB, Huppertz HJ. Fully automated atlas-based MR imaging volumetry in Huntington disease, compared with manual volumetry. AJNR Am J Neuroradiol 2011;32(7):1328–32. http://dx.doi.org/10.3174/ajnr.A2514.

[18] Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 2009;46(3):786–802. http://dx.doi.org/10.1016/j.neuroimage.2008.12.037.

[19] Kruggel F, Turner J, Muftuler LT, Alzheimer's Disease Neuroimaging, I.. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 2010;49(3):2123–33. http://dx.doi.org/10.1016/j.neuroimage.2009.11.006.

[20] Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 2003;19(3):1233–9. http://dx.doi.org/10.1016/s1053-8119(03)00169-1.

[21] Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 2007;19(9):1498–507. http://dx.doi.org/10.1162/jocn.2007.19.9.1498.

[22] Matsuda H, Mizumura S, Nemoto K, Yamashita F, Imabayashi E, Sato N, et al. Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer disease. AJNR Am J Neuroradiol 2012;33(6):1109–14. http://dx.doi.org/10.3174/ajnr.A2935.

[23] Morey RA, Selgrade ES, Wagner II HR, Huettel SA, Wang L, McCarthy G. Scan–rescan reliability of subcortical brain volumes derived from automated segmentation. Hum Brain Mapp 2010;31(11):1751–62. http://dx.doi.org/10.1002/hbm.20973.

[24] Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, et al. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage 2008;40(2):570–82. http://dx.doi.org/10.1016/j.neuroimage.2007.12.035.

[25] Peelle JE, Cusack R, Henson RN. Adjusting for global effects in voxel-based morphometry: gray matter decline in normal aging. Neuroimage 2012;60(2):1503–16. http://dx.doi.org/10.1016/j.neuroimage.2011.12.086.

[26] Radue EW, Barkhof F, Kappos L, Sprenger T, Haring DA, de Vera A, et al. Correlation between brain volume loss and clinical and MRI outcomes in multiple sclerosis. Neurology 2015. http://dx.doi.org/10.1212/wnl.0000000000001281.

[27] Rohlfing T. Incorrect ICBM-DTI-81 atlas orientation and white matter labels. Front Neurosci 2013;7:4. http://dx.doi.org/10.3389/fnins.2013.00004.

[28] Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, et al. Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 2008;39(3):1064–80. http://dx.doi.org/10.1016/j.neuroimage.2007.09.031.

[29] Smith SM, De Stefano N, Jenkinson M, Matthews PM. Normalized accurate measurement of longitudinal brain change. J Comput Assist Tomogr 2001;25(3):466–75.

[30] Takahashi R, Ishii K, Miyamoto N, Yoshikawa T, Shimada K, Ohkawa S, et al. Measurement of gray and white matter atrophy in dementia with Lewy bodies using diffeomorphic anatomic registration through exponentiated lie algebra: a comparison with conventional voxel-based morphometry. AJNR Am J Neuroradiol 2010;31(10):1873–8. http://dx.doi.org/10.3174/ajnr.A2200.

[31] Wolz R, Schwarz AJ, Yu P, Cole PE, Rueckert D, Jack Jr CR, et al. Robustness of automated hippocampal volumetry across magnetic resonance field strengths and repeat images. Alzheimers Dement 2014;10(4):430–438.e432. http://dx.doi.org/10.1016/j.jalz.2013.09.014.