Psychiatry Research: Neuroimaging 193 (2011) 182–190

Contents lists available at ScienceDirect

Psychiatry Research: Neuroimaging

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

Assessing the reproducibility of the SienaX and Siena brain atrophy measures using the ADNI back-to-back MP-RAGE MRI scans

Keith S. Cover [a,*], Ronald A. van Schijndel [b], Bob W. van Dijk [a], Alberto Redolfi [c], Dirk L. Knol [d], Giovanni B. Frisoni [c], Frederik Barkhof [b], Hugo Vrenken [a,b,e] and neuGRID [1] and the Alzheimer's Disease Neuroimaging Initiative [2]

[a] Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, The Netherlands
[b] Department of Radiology, VU University Medical Center, Amsterdam, The Netherlands
[c] Laboratory of Epidemiology & Neuroimaging, IRCCS San Giovanni di Dio Fatebenefratelli, Via Pilastroni 4-25125 Brescia, Italy
[d] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands
[e] MS Center Amsterdam and Alzheimer Center, VU University Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 19 August 2010
Received in revised form 26 February 2011
Accepted 27 February 2011

Keywords:
Alzheimer's disease
Power calculations
Non Gaussian distributions

ABSTRACT

SienaX and Siena are widely used and fully automated algorithms for measuring whole brain volume and volume change in cross-sectional and longitudinal MRI studies and are particularly useful in studies of brain atrophy. The reproducibility of the algorithms was assessed using the 3D T1 weighted MP-RAGE scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. The back-to-back (BTB) MP-RAGE scans in the ADNI data set makes it a valuable benchmark against which to assess the performance of algorithms of measuring atrophy in the human brain with MRI scans. A total of 671 subjects were included for SienaX and 385 subjects for Siena. The annual percentage brain volume change (PBVC) rates were -0.65 ± 0.82%/year for the healthy controls, -1.15 ± 1.21%/year for mild cognitively impairment (MCI) and -1.84 ± 1.33%/year for AD, in line with previous findings. The median of the absolute value of the reproducibility of SienaX's normalized brain volume (NBV) was 0.96% while the 90th percentile was 5.11%. The reproducibility of Siena's PBVC had a median of 0.35% and a 90th percentile of 1.37%. While the median reproducibility for SienaX's NBV was in line with the values previously reported in the literature, the median reproducibility of Siena's PBVC was about twice that reported. Also, the 90th percentiles for both SienaX and Siena were about twice the size that would be expected for a Gaussian distribution. Because of the natural variation of the disease among patients over a year, a perfectly reproducible whole brain atrophy algorithm would reduce the estimated group size needed to detect a specified treatment effect by only 30% to 40% as compared to Siena's.

1. Introduction

Measures of volume and volume changes of the brain using magnetic resonance images are becoming widely used to monitor the state and progression of diseases such as Alzheimer's disease (AD) and multiple sclerosis (MS). Several early magnetic resonance imaging (MRI) studies demonstrated the potential of brain atrophy as a measure of Alzheimer's disease (Fox et al., 1996a, 1996b; Jack et al., 1998, 1999; Killiany et al., 2000) and MS (Losseff et al., 1996; Rudick et al., 1999; Molyneux et al., 2000; Miller et al., 2002). Recent results have shown the continued and growing interest in atrophy algorithms in both AD (Jack et al., 2008; Sluimer et al., 2009, 2010) and MS (Bermel and Bakshi, 2006; de Stefano et al., 2007; Altmann et al., 2009; Barkhof et al., 2009; Korteweg et al., 2009). Indeed, in their recent review Barkhof et al. (2009) recommended whole brain atrophy as one of the preferred MRI outcomes for phase II neuroprotection and repair trials in MS. In addition Frisoni et al. (2010) stated that rates of whole brain and hippocampal atrophy are sensitive markers of neurodegeneration and can be used as secondary outcomes in phase III trials of potentially disease-modifying therapies in AD.

In addition to anatomically specific algorithms such as FIRST [FMRIB Integrated Registration and Segmentation Tool, Oxford University, Oxford UK], voxel-based morphometry (VBM) (Ashburner and Friston, 2000) and Freesurfer (Fischl et al., 2002), there has been a

* Corresponding author at: VU University Medical Center, Department of Physics and Medical Technology, Postbus 7057, 1007MB Amsterdam, The Netherlands. Tel.: +31 20 444 0677; fax: +31 20 444 4816.
E-mail address: Keith@kscover.ca (K.S. Cover).
[1] *see* http://neuGRID.eu.
[2] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf).

0925-4927/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.pscychresns.2011.02.012

growing use of whole brain atrophy algorithms. These include Siena and SienaX (Smith et al., 2001; Smith, 2002; Smith et al., 2002, 2007; Battaglini et al., 2008), brain boundary shift interval (BBSI) (Freeborough and Fox, 1997; Fox et al., 2000) and the brain parenchymal fraction (BPF) (Rudick et al., 1999). The growth in the use of atrophy algorithms has driven an interest in evaluating atrophy algorithms performance including their accuracy and reproducibility (Horsfield et al., 2003; Jasperse et al., 2007; Klauschen et al., 2009; Frisoni et al., 2010). The various algorithms have been compared against each other (Sormani et al., 2004; Zivadinov et al., 2005; Smith et al., 2007) and their performance on various types of MRI sequences has been assessed (Neacsu et al., 2008). In addition, Sormani et al. (2004) reported Siena to have only half the error of automatic seed growing that used a semiautomated technique for brain parenchyma segmentation. In the technique a seed was positioned in any part of the cerebral parenchyma and a region of interest (ROI) was grown from the seed using upper and lower intensity thresholds, which were set interactively by a user.

Many studies have used the whole brain atrophy algorithms SienaX and Siena (Smith et al., 2002, 2007), which are part of the FSL package. SienaX measures the volume of the brain from a single MRI scan and then normalizes it to a standard skull to yield a normalized brain volume (NBV). The NBV can be thought of as the fraction of the skull that is filled with brain. In contrast, Siena measures the percentage brain volume change (PBVC) between two scans of the same subject. Thus SienaX is useful in cross-sectional studies when the longitudinal scans required by Siena are not available. Siena is preferred for longitudinal studies because it has better reproducibility than SienaX as Siena finds the volume changes between two scans of the same subject. That both algorithms are fully automated and widely available makes them particularly appealing.

The reproducibility of brain atrophy algorithms has been calculated across the literature in a consistent manner. The reproducibility has been calculated from the BTB difference over a group of subjects. The BTB difference for a subject is the difference between the same algorithm for brain atrophy applied to two MP-RAGEs of the same subject when the MP-RAGEs are acquired within a short period of time of each other. As SienaX measures the NBV of each subject, SienaX's BTB difference is commonly expressed as the percentage change between the two volumes. Since Siena measures the percentage brain volume change (PBVC) between two MP-RAGEs of the same subject at different points in time, the difference between the two PBVC values also has units of percentage. Precise definitions of the BTB differences are provided in the methods section.

Various statistics can be used to summarize the BTB differences as a reproducibility over particular groups. The most common one in the literature is the median of the absolute value of the BTB difference (Smith et al., 2007). This statistic will be used in the current paper unless otherwise stated. The definitions of other statistics of the BTB differences are described in the Methods.

For SienaX Smith et al. (2002) reported a brain volume reproducibility of 0.5% to 1.0%, based on axial 2D T1 weighted fast field echo scans from 16 healthy controls (HC). All subjects were each scanned at the same center and twice within 1 week. Using the same data set as the SienaX results, Smith et al. (2002) also reported a reproducibility for Siena of 0.15%.

In a later article, and using 3D acquired T1 weighted scans acquired at a single center, Smith et al. (2007) reported the reproducibility of Siena for 185 back-to-back (BTB) acquisitions acquired from 68 subjects, comprising 45 patients with AD and 23 age-matched controls, as 0.16%. The addition statistic of 0.27% for the mean of the absolute value of the BTB difference was also provided for the same data set. Using incremental atrophy summation (IAS), Smith et al. (2007) also found the median and mean absolute difference reproducibility of Siena to be 0.16% and 0.20%, respectively.

Other studies have estimated the between- and within-individual variability from longitudinal scans without BTB MP-RAGEs (e.g., Schott et al., 2006). While requiring less scan time, since only one 3D T1 weighted acquisition is required rather than two, the within-individual variability will be sensitive to a broader range of variabilites than the BTB difference. The additional variabilities would include patient positioning and anything, other than AD, that would cause the patients' brain volumes to change between visits. Also, determining the distribution of the reproducibility of the MP-RAGEs requires several assumptions regarding disease modeling. The full impact of these assumptions on the variability of the result may be difficult to predict.

The MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack et al., 2008) offer several advantages when estimating the reproducibility of brain volume algorithms such as SienaX and Siena. The ADNI study acquired scans at more than 50 sites across North America. It has more than 800 subjects in a cohort composed of HC, mild cognitive impairment (MCI) and AD. MRI scans of each subject are acquired yearly, providing a large cross-sectional and longitudinal data set.

A unique characteristic of the ADNI MRI protocol making it particularly well suited to the study of the reproducibility of brain volume algorithms is that each subject's MR visit includes BTB acquisitions of the 3D T1 weighted magnetization prepared rapid gradient echo (MP-RAGE) sequence (Mugler and Brookeman, 1990) with identical parameters. These BTB MP-RAGEs, which were all acquired at 1.5 T, were included in the ADNI protocol to ensure that at least one was of satisfactory quality (Jack et al., 2008). However, for most patients, both scans are of high quality. This high quality is partly because in the ADNI protocol the acquisition of the second BTB MP-RAGE starts within seconds of the completion of the first, thus the subjects MRI scans are likely very similar. In addition, any variability introduced by the acquisition of the MRI scans over more than 50 acquisition sites is in line with that of current clinical trials.

BTB MP-RAGE scans are rarely included in MRI studies because of the additional acquisition time required. Reproducibility studies often remove the patient from a MRI before repeating a scan later the same day or within the next days or weeks thus introducing the variable of patient repositioning. Therefore, ADNI BTB MP-RAGEs are particularly well suited to isolating the variability introduced by the reproducibility of a particular algorithm of brain volume change from other sources of variability in a study. Thus, given the wide variety of MRI scanners and sites in the ADNI study, the ADNI BTB MP-RAGEs are a particularly good benchmark against which to assess the performance of brain volume change algorithms.

It should be kept in mind that the pair of ADNI BTB MP-RAGEs is of degraded value when it comes to applying atrophy algorithms. Most MRI scanners introduce distortions into their MRI images because of nonlinear gradients. If these distortions, usually called gradient distortion (GD), are not corrected for by post acquisition processing they can lead to systematic errors in brain volume change algorithms. While both of the BTB MP-RAGEs are available without ADNI's post acquisition processing, only one of each BTB scans has received the post acquisition processing. However, as long as the patient is in the same position in the MRI scanner for both of the BTB scans, the GD will be the same, and thus will not affect the brain volume algorithm BTB difference. Most recently, Caramanos et al. (2010) detailed the potential detrimental effects of GD on Siena reproducibility. They showed the primary source of systematic errors in the atrophy rates was displacement along the z direction of the MRI scanners.

Several recent publications have used the ADNI data set to assess the performance of MRI scanners (Clarkson et al., 2009; Gunter et al., 2009; Kruggel et al., 2010) and algorithms to measure brain volume change (Morra et al., 2008; Chupin et al., 2009; Morra et al., 2010). However, to date no one has used the BTB MP-RAGE as a benchmark for evaluating the reproducibility of the volume change algorithms.

The current study assessed the reproducibility of both the SienaX and Siena atrophy algorithms for measuring brain volume change and

demonstrates how the BTB MP-RAGEs in the ADNI protocol provide an excellent benchmark on which to test the reproducibility of brain atrophy algorithms.

2. Methods

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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 primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

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. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research — approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org.

To insure GD was not a factor in the reproducibility the displacement of subject heads was calculated for each BTB scan. To measure the displacement between BTB MP-RAGEs, the Flirt algorithm in FSL was used to linearly co-register the BTB MP-RAGEs using rigid body (6 d.f.). The resulting rigid body transform parameters were then used to calculate the displacement of each voxel from the first to second BTB MP-RAGE. The maximum displacement observed inside the BET brain mask of the first BTB MP-RAGE was then determined and used as an indication of the amount of displacement between MP-RAGEs. As a further check of the influence of the quality of the BTB MP-RAGEs, a scatter plot of the SienaX BTB differences versus the Siena BTB differences was plotted.

For this reproducibility study, FSL version 4.1.4 was used. For both SienaX and Siena, the brain extraction tool (BET) was run with only the −B option, which implements a correction for spatial signal inhomogeneity and robust brain center estimation for optimized brain extraction.

All BTB MP-RAGEs that were acquired before the summer of 2009 were downloaded from the ADNI website. None of the scans included in this study had received any post acquisition processing such as N3 correction, GD or B1 non uniformity correction. To reduce the likelihood of poor quality MRI scans, any subjects that had more than 2 MP-RAGEs during a patient visit were excluded from the reproducibility analysis.

For the SienaX reproducibility study, only the month 0 scan for each patient, referred to as the screening scan in the ADNI data set, was used. In a total of 13 subjects SienaX failed to yield values for one or both of the MP-RAGEs leaving a total of 671 subjects in the study. For Siena, the month 0 and month 12 scans were used. For Siena, one subject failed to yield a value giving a total of 385 subjects for further analysis. The subjects that failed in SienaX and Siena were different. More information on group sizes is given in Table 1. The first PBVC value ($PBVC_1$) was calculated from the first acquired MP-RAGE of both patient visits, and the second PBVC value ($PBVC_2$) was calculated from the second acquired MP-RAGE of both patient visits.

For this article, the reproducibility of a volume algorithm was calculated from the BTB difference for each subject. The BTB difference was calculated by the second minus the first of the BTB MP-RAGEs. As SienaX measures the NBV in litres, its BTB difference ($BTBD_{SX}$) was calculated as the volume of the second scan minus the first with the difference divided by their average value and then expressed as a percentage. The SienaX BTB difference expressed as an equation is

$$BTBD_{SX} = 200(NBV_2 - NBV_1) / (NBV_2 + NBV_1) \quad (1)$$

A symmetrized version of the percentage change was used because there was no reason to treat either of the BTB MP-RAGEs preferentially.

As Siena already yields a percentage change between two patient visits, its BTB difference ($BTBD_S$) was defined to be the PBVC value of the second acquired of the BTB MP-RAGE minus PBVC value of the first acquired. The Siena BTB difference expressed as an equation is

$$BTBD_S = PBVC_2 - PBVC_1 \quad (2)$$

Several statistics of the BTB differences were calculated to give a clearer picture of the distributions of the BTB differences. In addition to the median of the absolute value of the BTB differences (Smith et al., 2007), three other statistics were included in this article. The 90th percentile of the absolute value of the BTB difference, when compared with the median, gives a measure of how close the distributions are to Gaussian. The standard deviation of the BTB difference is also included because it was used in the group size calculations. As Smith et al. (2007) also used the mean of the absolute value of the BTB differences, this statistic is included for completeness.

In addition to the reproducibility, to characterize the three diagnostic groups and to compare our results with previous findings, we also calculated for each of the groups the mean NBV and mean annual atrophy rate. The latter was calculated as PBVC/year by dividing PBVC by the time between the month 0 and month 12 scans to correct for minor variations in follow-up duration between subjects.

An important issue in the design of clinical trials is the smallest group size that still has sufficient statistical power to detect an expected treatment effect. A typical question would be what group size would be required to detect a reduction in the whole brain atrophy by 0.5 percentage points — for example from a mean value of −1.8% to −1.3%. There is a common statistical standard that a group size must satisfy for it to be considered to be sufficiently large to detect a specified treatment effect. We assume that both the treated and untreated groups are of the

Table 1
Reproducibility of SienaX and Siena for the BTB difference for the combined and diagnostic groups. In addition to the number of subjects, four difference statistics for the various groups are presented for the BTB reproducibility.

	SienaX NBV				Siena PBVC			
	Combined	HC	MCI	AD	Combined	HC	MCI	AD
Number of subjects	671	183	330	158	385	105	195	85
Median abs value (%)	0.96	0.92	0.91	1.11	0.35	0.34	0.32	0.41
90th percentile abs (%)	5.11	4.86	4.95	5.29	1.37	1.12	1.46	1.62
Standard deviation (%)	3.37	2.95	3.37	3.77	0.95	0.72	0.91	1.01
Mean abs value (%)	1.97	1.80	1.96	2.22	0.59	0.49	0.55	0.64

same size and that the group size must be large enough such that the significance of the statistical test falls below $p = 0.05$ at least 80% of the time (Altmann et al., 2009; Holland et al., 2009).

The testing of the statistical power of various group sizes (Holland et al., 2009) was implemented in this study using bootstrapping (Mooney and Duval, 1985; Altmann et al., 2009). Bootstrapping is a particularly robust method for group size estimation as it can handle normal and non normal distributions equally well.

For bootstrapping to find the smallest group size that has sufficient statistical power, three distributions have to be determined. The first, referred to as the disease distribution, is the distribution of the volume measured over a representative set of subjects in the normal progression of the disease. The second, referred to as the reproducibility distribution, is the distribution of the reproducibility for individual measurements of the atrophy. The third, referred to as the treatment distribution, simulates the atrophy after treatment and is calculated from the disease distribution and depends on the specified treatment effect. For bootstrapping, all three distributions were derived from the ADNI data set including the treatment distribution which was derived via the disease distribution.

The disease distribution was calculated by averaging the two volume measures from each pair of BTB MP-RAGEs. For SienaX, each element of the disease distribution set corresponded to the average NBV of a subject

$$(NBV_2 + NBV_1)/2 \tag{3}$$

Similarly, for Siena, each element of the disease distribution set corresponded to the average PBVC of a subject

$$(PBVC_2 + PBVC_1)/2 \tag{4}$$

The reproducibility distribution was calculated by taking half the difference of the volume measures from the BTB MP-RAGEs. For SienaX, each element of the reproducibility distribution set is

$$(NBV_2 - NBV_1)/2 \tag{5}$$

The corresponding equation for Siena is

$$(PBVC_2 - PBVC_1)/2 \tag{6}$$

The mathematics of the disease distribution is different from the reproducibility distribution because the disease distribution is signal while the treatment distribution is noise. With ideal noise the standard deviations add as the squares. Thus, Eqs. (5) and (6) require the division by a first square root of two because the BTB difference is the difference between two measurements that are assumed to have the same standard deviation. Division by a second square root of two accounts for the disease distribution being the average of the pair of BTB volume algorithms, thus reducing the variance of the disease distribution. An underlying assumption of these calculations is that the disease and reproducibility distributions are uncorrelated.

The treatment distribution was derived from the disease distribution by multiplication by a scaling factor. For example, to simulate a treatment distribution that has a mean that is 0.5 percentage points larger than SienaX's disease distribution, the disease distribution was multiplied by 1.005. For Siena, to simulate a treatment distribution with a mean of −1.3% from a disease distribution with a mean of −1.8%, the disease distribution was multiplied by 0.7222.

During the bootstrap calculation, to determine the power of a particular group size to detect a particular treatment effect, 100,000 realizations of a test version of the disease and treatment distributions were generated. For each realization, each disease and treatment test distribution was generated by selecting randomly, with replacement, from the corresponding measured distributions. Each subject's value in the disease and treatment test distributions then had a BTB difference added to its value that was selected randomly, with replacement, from the reproducibility distribution. The disease and treatment distributions were then compared using the two-sided Mann–Whitney–Wilcoxon nonparametric test (Altman, 1991). The fraction of the p-values less than the statistical significance threshold of 0.05 was then calculated to see whether it exceeded the required power of 0.8.

Three different versions of the reproducibility distribution were used during the bootstrap. The first simulated a single MP-RAGE. The second simulated the average of two BTB MP-RAGEs. The third assumed the brain volume measurement was perfectly reproducible and thus no reproducibility value was added to either the disease or treatment test distributions.

All SienaX and Siena calculations were run on the DAS3 cluster using the MirageGRID software (Sluimer et al., 2009). The total time to complete all of the Siena subjects was under 10 h, about 200 times faster than on a single computer. The calculation of the SienaX reproducibility required similar resources.

3. Results

Fig. 1 shows the distribution of SienaX's NBV and Siena's PBVC for the diagnostic groups presented as box–whisker plots. Both distributions for both of the BTB MP-RAGE are shown. From the first of the BTB MP-RAGE of each subject, the mean NBV for the HC group was 1.48 L with a standard deviation of 0.10 L. The MCI group had a mean volume of 1.45 ± 0.09 L and the AD group had a mean volume of 1.45 ± 0.10 L.

The mean annual PBVC, as obtained by taking the first MP-RAGE of the BTB MP-RAGE, was −0.65%/year with a standard deviation of 0.82%/year for the HC group. The MCI group had a mean PBVC of −1.15 ± 1.21%/year and the AD group had a mean of -1.84 ± 1.33%/y. As expected, the corresponding values for the second scans were nearly identical to the first. They were −0.66 ± 0.81%/year, −1.14 ± 1.30%/year and −1.88 ± 1.47%/year. For the AD group there was a slightly larger range of atrophy rates on the second MP-RAGE than the first, although the statistical significance is unclear. As with the comparable values for NBV, even though there was no correction for GD, the trend to more rapid disease progression in the PBVC from HC to MCI to AD was clear.

The BTB difference of the NBV was calculated for the pair of BTB MP-RAGE for each subject. The histogram is shown in Fig. 2 including

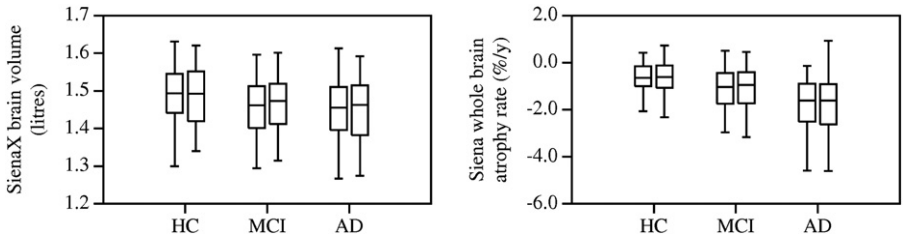

Fig. 1. Box-and-whisker plot of (left) baseline brain volume and (right) whole brain atrophy rate, by diagnostic group. Horizontal line inside the box is median value, the box boundaries are the 25 and 75 percentiles, and the whiskers are the 5 and 95 percentiles. For each diagnostic group, the box-and-whisker plot of the left corresponds to the first of the BTB MP-RAGE and on the right to the second.

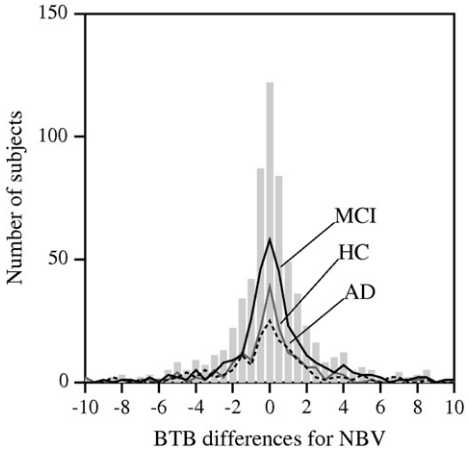

Fig. 2. Histogram of the reproducibility, as quantified by the BTB difference, of SienaX's NBV over 672 subjects, including the breakdown by diagnostic groups.

the breakdown by diagnostic group. Table 1 provides some statistics of the NBV reproducibility, as calculated from the BTB difference, for the combined and diagnostics groups. Fig. 3 shows the scatter plot of the BTB difference of the NBV versus the displacement of the subject's head between the BTB MP-RAGE. The displacement of most heads is less than 5 mm. The few heads that were displaced by more than 5 mm did not have BTB differences that were substantially worse than those with less displacement.

Fig. 4 shows the Bland–Altman scatter plot of the average NBV of the BTB MP-RAGE versus the difference of the NBV. From visual inspection of Fig. 4, there was no evidence of dependence of the difference in the NBV on the average of the NBV.

Fig. 5 shows the histogram of the BTB differences of Siena's PBVC. A slight asymmetry in the histogram of about 0.1% is apparent indicating that the negative BTB differences are slightly more common among the various groups than the positive ones. The asymmetry is discussed further in the Discussion. Table 1 provides some statistics of the PBVC BTB differences for the combined and diagnostics groups.

Fig. 6 shows the scatter plot of the difference of the PBVCs versus the maximum displacement of the subject's head between BTB MP-RAGE. In this case, since there were two patient visits for each subject, and thus two BTB MP-RAGEs, the maximum displacement of the two patient visits for each subject was used to generate Fig. 6. As with SienaX, the displacement of most heads is less than 5 mm and the few heads that were displaced by more than 5 mm did not have differences in the BTB PBVCs that were substantially worse than those with less displacement.

Fig. 7 shows the Bland–Altman scatter plot of the average PBVCs of the BTB MP-RAGE versus the difference of the PBVCs. As can be seen from Fig. 7, similar to what was observed for SienaX, there was little if any dependence of the difference in the PBVC to the average of the PBVC.

Fig. 8 shows the scatter graph of the SienaX versus the Siena BTB difference for each subject. Even though about half a dozen subjects fell outside the main pattern, the shape of the main pattern is clear. The subjects with the poorest NBV reproducibility have good PBVC reproducibility while the subjects with poor PBVC reproducibility have good NBV reproducibility. This point is discussed further in the Discussion.

Table 2 shows the group sizes, estimated by bootstrap simulation, SienaX would require to detect a particular difference in NBV. Table 3 shows the corresponding group sizes for Siena. The reason the three different types of reproducibility are not presented for SienaX is covered in the Discussion. For Siena, the group sizes for the average of 2 MP-RAGE scans are only 10% to 13% smaller than for a single MP-RAGE scan. Also for Siena, the group sizes for a perfectly reproducible whole brain atrophy algorithm, using the disease variation over 1 year, are 30% to 40% smaller than for a single MP-RAGE scan.

4. Discussion

The annual whole brain atrophy rates measured in the current study were in good agreement with those presented in the literature. For example Sluimer et al. (2008), on a different cohort than ADNI but using Siena to measure the brain atrophy, reported −1.2% for MCI and −1.9% for AD. These values compares favorably with the −1.15% and −1.84% measured in the current study. Fox et al. (1999) reported an annual atrophy rate of −2.0% for AD patients. Evans et al. (2010), for the ADNI cohort but using the semi-automated boundary shift integral (BSI), obtained values of −0.49%, −1.05% and −1.50% (HC, MCI, and AD). While the HC and MCI values are close to the current study, the AD value is slightly different. It is interesting to note there is good agreement between the current study and previous studies even though the MP-RAGE in the current study received no correction for GD.

Several measures of reproducibility were used in this manuscript. These included the median of the absolute value of the BTB differences (Eqs. (1) and (2)), which is commonly used in the literature. A novel measure of the reproducibility used was the 90 percentile of the absolute value of the BTB differences. The 90 percentile reproducibility measure was included so the shape of the BTB distributions could be compared to Gaussian. The standard deviation of the BTB distributions was also calculated.

For SienaX, the median of the absolute value of the BTB difference found, 0.96% (Table 1), was in line with the 0.5% to 1.0% reported by Smith et al. (2002). However, although no values for the 90th percentile have been previously published, at 5.11%, it was unexpectedly high. For a Gaussian distribution with a median of the absolute value at 0.96%, the 90th percentile would be expected to be at 2.34%. Thus, the measured 90th percentile of the reproducibility for SienaX is more than twice the value that would be expected for a Gaussian distribution. Thus, at least 10%, and likely more, of the BTB differences for each of the groups must be outliers to a Gaussian distribution.

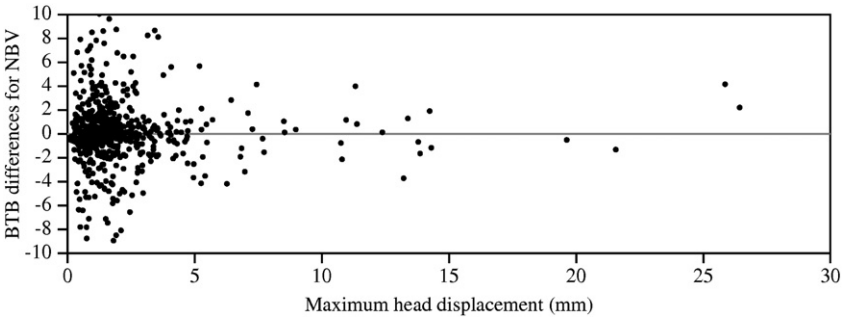

Fig. 3. Scatter plot of the reproducibility of SienaX's NBV versus the displacements of the heads between BTB MP-RAGE.

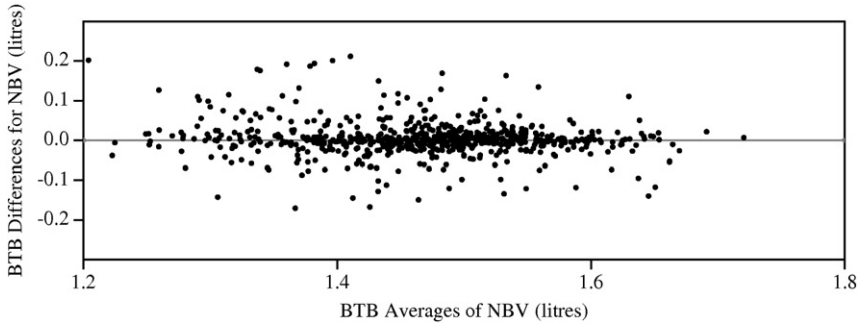

Fig. 4. Bland–Altman scatter plot of the difference between SienaX's NBV for BTB MP-RAGE versus the average of the two NBVs.

For Siena, the absolute median of the PBVC's BTB difference was about twice as big as the published values. Smith et al. (2007) reported an absolute mean reproducibility of 0.16% for BTB difference. The current study found the comparable value of reproducibility to be 0.35%. One possible cause of the doubling in the reproducibility may be due to the wide variety of scans from the more than 50 acquisition sites included in the ADNI study. While the Smith et al. (2007) study had a mix of HC and AD subjects and used a 3D T1 weighted sequence, it only acquired the MRI scans at a single center. As with SienaX, the 90th percentile of reproducibility for Siena at 1.33% was much larger than expected for a Gaussian distribution.

The reproducibility of the SienaX and Siena is reasonably close to the values published in the literature. This result is impressive considering the wide variety of MRI scanners that were employed in the more than 50 acquisition sites included in the ADNI study. This result bodes well for the use of SienaX and Siena in clinical trials. However, the reproducibility of SienaX and Siena has distributions that have much larger shoulders than a Gaussian and any analysis that assumes a Gaussian distribution may yield inaccurate results. The use of bootstrapping and the Mann–Whitney–Wilcoxon statistical test in the current study handled Gaussian and non Gausssian distributions equally well and avoided this potential error.

The primary contributions to the variation in the brain volume change from MP-RAGE to MP-RAGE and patient to patient are likely measurement reproducibility and disease variation. The ADNI BTB study allows these two sources of variation to be separated from each other. While other factors may introduce some variation, such as intervals between scans or disease progression, because of the study design these variations will likely be incorporated into the disease variability rather than the reproducibility.

The standard deviation of the measured atrophy rates of the subjects, σ_{measured}, is related to the standard deviation of the disease atrophy rates, σ_{disease}, and the standard deviation of the reproducibility, σ_{repro}, by the equation

$$\sigma^2_{\text{measured}} = \sigma^2_{\text{disease}} + \sigma^2_{\text{repro}} \tag{7}$$

assuming the distributions are uncorrelated. The σ_{measured} and σ_{repro} values can be determined from the ADNI data set for all combinations of the SienaX and Siena volume measures and the MCI and AD diagnostic groups. For example, the measured AD group of Siena had a $\sigma_{\text{measured}} = 1.33\%$. The reproducibility of single MP-RAGE measurement is the standard deviation of the BTB difference, which is 1.01%, divided by the square root of 2, yielding $\sigma_{\text{repro}} = 0.71\%$. Thus, from Eq (7), $\sigma_{\text{disease}} = 1.12\%$ over 1 year. Also, according to Eq. (7) the squares of the standard deviations of the disease and reproducibilities are summed to get the standard deviation of the measured distribution. Therefore, most of the variation in the measured distribution for the AD diagnostic group measured with Siena over 1 year is due to the disease variation rather than the reproducibility. This is also true for the MCI diagnostic group.

If the MP-RAGE scans were acquired for 1 to 2 more years on each subject, the disease standard deviation would be expected to grow while the reproducibility would be expected to stay the same. Thus for Siena's measurements of brain atrophy in AD patients in multiyear studies the main source of variation will not be measurement error, but disease variation. As averaging of more MRI scans and longer duration studies allows the averaging down of the reproducibility but not the disease variation, Siena measurements in AD reaches the point of diminishing returns within 1 or 2 years.

This characteristic of Siena is also indicated by the required group sizes for specified treatment effects. Increasing from 1 to 2 MP-RAGE scans only reduces group sizes by about 10%. Indeed, if a version of Siena could be implemented with perfect reproducibility, it would only reduce group sizes by about 35% in a one year study. In a multiple year study, because of the relatively large variation in the atrophy rates over MCI and AD patients, this reduction in group size can be expected to be substantially less. Consequently major improvements in Siena's reproducibility should only have minor effects on the required group sizes on multiyear studies.

The standard deviation of SienaX's BTB difference of 3.8% is nearly 4 times larger than that of Siena. SienaX, because of its poor reproducibility when compared to Siena, should benefit much more from frequent scanning and large duration studies to average down the reproducibility. Altmann et al. (2009) noted this behavior after applying SienaX and Siena to a multiyear study of whole brain atrophy in multiple sclerosis. Analysis of multiyear scans for subjects in the ADNI study should also confirm this behavior.

The foregoing calculation is based on the assumption that the disease variation and BTB differences are uncorrelated. However, we cannot be sure this assumption is completely true. Examination of Table 1 suggests that the BTB differences may be slightly worse for more advanced disease. However, the BTB differences between the groups are small and it is not clear how this slight difference would

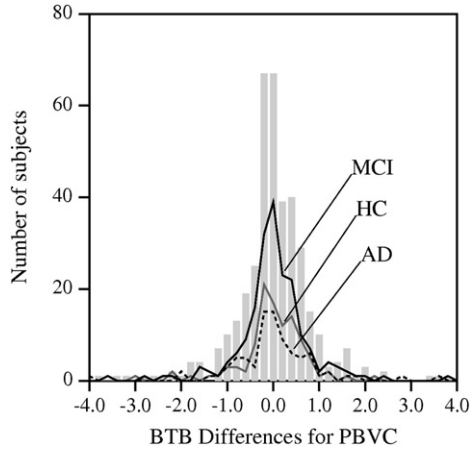

Fig. 5. Histogram of the reproducibility, as quantified by the BTB difference, of Siena's PCBV over 385 subjects including the breakdown by diagnostic groups.

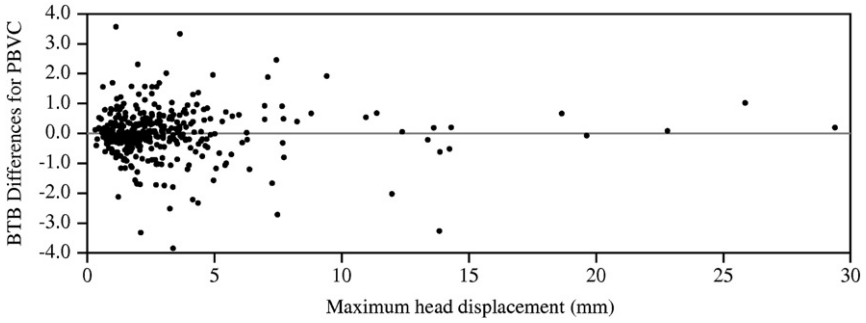

Fig. 6. Scatter plot of the reproducibility of Siena's PCBV versus the displacements of the heads between BTB MP-RAGEs.

affect the separation between the disease variety and the atrophy measurement reproducibility BTB differences.

From the results of the current study it is possible to check whether the poor reproducibility found in some subjects is due to poor quality MP-RAGEs. Applying SienaX to a poor quality MPRAGE can be expected to yield a NBV with an outlying value. Also, applying Siena to a pair of MP-RAGES, either of which has a poor quality, can be expected to yield an outlying PBVC. So if any of the 4 MP-RAGEs per subject are of poor quality then in at least half the subjects both the NBV and PBVC will yield outlying volume measures.

The scatter graph in Fig. 8 shows that for most subjects both SienaX and Siena have small BTB differences. For those subjects that do have poor reproducilibilty in SienaX or Siena the other measure appears to have good reproducibility, yielding a distinctive cross pattern to the scatter graph. Only about half a dozen subjects of the 385 subjects had outlying values for both SienaX and Siena. Therefore, it can be stated with confidence that the vast majority of outlying NBV and PBVC measures in this study are not related to poor MP-RAGE quality since poor image quality should yield poor values for both BTB differences for the subject. Although not shown, Fig. 8 was also plotted for the HC, MCI and AD groups separately. All groups showed similar patterns to the combined group with none showing substantially more outliers. Thus, the underlying cause of the outliers remains unclear.

Figs. 3 and 6 demonstrated that the displacement of subjects' heads between BTB MP-RAGEs had little effect on the reproducibility of both SienaX and Siena. This minimal effect may be due to the limited motion of the patients head between BTB MP-RAGEs. As the second scan begins within seconds of the completion of the first, the likely motions of the head are left–right rotation and anterior-superior tilting. Motion in the *z* direction is very restricted because the subject remains on the table during the MRI scan. Caramanos et al. (2010) demonstrated that whole brain volume algorithms were particularly sensitive to displacement in the *z* direction in the MRI. Thus the displacement between the BTB MP-RAGEs in the ADNI study is probably not particularly sensitive to GD distortion.

There is a slight asymmetry in the Siena histogram (Fig. 5) of the order of about 0.1%. To check the possibility that the asymmetry might be due to post acquisition analysis, the full analysis for each of the 385 subjects, including all the volume algorithms was repeated twice. The first repeat had the order of the BTB MP-RAGE switched for both patient visits for all subjects. The second repeat had the order of the 0 month and 12 month scans interchanged. Consistent with the proper performance of the post acquisition software, for both repeats, the PBVC histogram was mirrored left-to-right but otherwise identical. The results of the additional post processing strongly suggest that the asymmetry in the Siena histogram is somehow related to the acquisition of the MRI scans rather than its post acquisition processing.

One possible explanation of the asymmetry is that a subset of subjects was moving more in the second of the BTB MP-RAGE with the movement being more common in the AD group. This increased movement may explain the slight increase in the range of PBVC values for the second AD MP-RAGE shown in Fig. 1. It may also explain why for PBVC the MCI group sizes required to detect a treatment effect, shown in Table 3, are slightly smaller than for AD. Further investigation is needed to resolve this issue. Of course in practice, the disease progression in MCI is slower than AD so detection of any slowing of the disease will require larger treatment groups for MCI than AD.

While the ADNI BTB MP-RAGEs provide a valuable way to assess the reproducibility of algorithms to measure brain volume changes it should be kept in mind that other sources of variation can affect the outcome of a clinical trial besides measurement reproducibility and disease variation. Therefore, any additional sources of variation would likely be incorporated into the disease variation part of Eq. (7).

In conclusion, for a fully automatic algorithm for measuring whole brain atrophy, Siena performed well, especially considering that ADNI is such a large and varied data set. SienaX, as is well documented in the literature, performed less well. While the unexpectedly large 90th percentiles of SienaX's and Siena's reproducibilities indicated non Gaussian distributions, none of the processing used in this study assumed a Gaussian distribution. It was demonstrated that Siena's reproducibility was sufficiently small that, in multiyear studies, it is likely that any further improvement of Siena's reproducibility will yield only diminishing returns. This study also demonstrated that the ADNI data set provides a valuable benchmark by which to assess the reproducibilities of current and future volume algorithms.

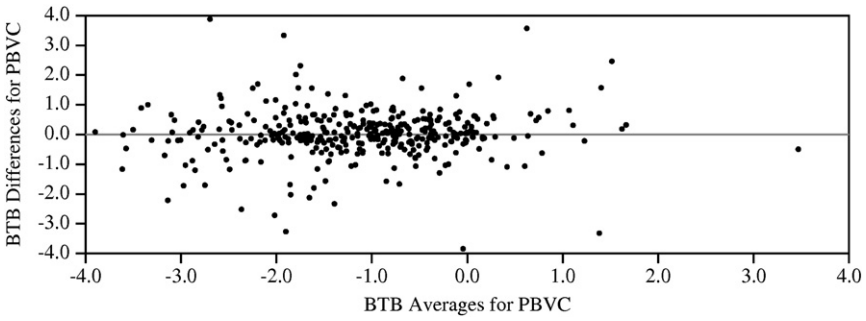

Fig. 7. Bland–Altman scatter plot of the difference between Siena's PCBV for BTB MP-RAGE versus the average of the two PCBVs.

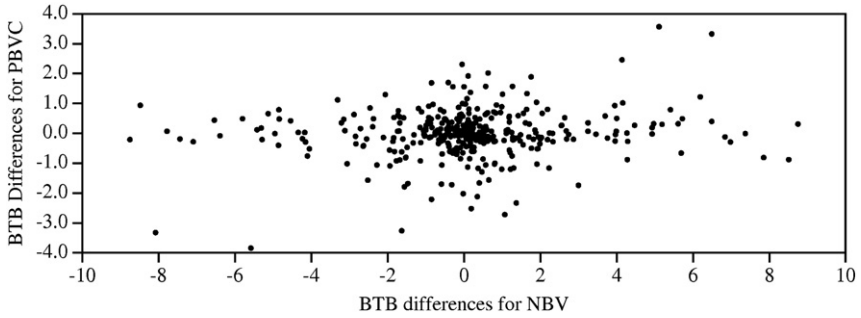

Fig. 8. Scatter plot of the reproducibility of SienaX versus that of Siena for each subject.

Table 2
SienaX sample size estimates using bootstrapping for the MCI and AD. While the treatment effect varies with the cohort, a 50% treatment effect in AD is usually about 0.6 percentage points per year for whole brain atrophy measures.

Percentage point reduction	MCI	AD
1.0	510	608
0.9	617	736
0.8	777	944
0.7	1034	1221
0.6	1387	1621
0.5	1951	2165

Table 3
Siena sample size estimates using bootstrapping for the MCI and AD for a single MP-RAGE, the average of the two BTB MP-RAGEs and perfect reproducibility. While the treatment effect varies with the cohort, a 50% treatment effect in AD is usually about 0.6 percentage points per year for whole brain atrophy measures.

Percentage point reduction	MCI			AD		
	1 MP-RAGE	2 MP-RAGE	Perfect	1 MP-RAGE	2 MP-RAGE	Perfect
1.0	15	13	9	24	21	16
0.9	19	17	12	31	27	20
0.8	24	21	16	40	35	28
0.7	33	29	22	54	48	37
0.6	46	41	32	76	68	54
0.5	70	63	50	113	101	80
0.4	116	104	84	182	163	123
0.3	220	198	158	335	300	210

Acknowledgements

The research for this paper was funded by neuGRID (http://neuGRID.eu) (European Union FP7 contract number RI-211714). MirageGRID was developed under funding from Netherlands' Virtual Laboratory for e-Science VL-e Project. The DAS3 computer cluster at the Computer Science department of the VU University was used for the high performance computing. Additional support for this study was received from the Department of Physics and Technology, VU University Medical Center in Amsterdam. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI 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 California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

References

Altman, D.G., 1991. Practical Statistics for Medical Research. Chapman and Hall, London.
Altmann, D.R., Jasperse, B., Barkhof, F., Beckmann, K., Filippi, M., Kappos, L.D., Molyneux, P., Polman, C.H., Pozzilli, C., Thompson, A.J., Wagner, K., Yousry, T.A., Miller, D.H., 2009. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. Neurology 72, 595–601.
Ashburner, J., Friston, K.L., 2000. Voxel-based morphometry—the methods. Neuroimage 11 (6 Pt 1), 805–821.
Barkhof, F., Calabresi, P.A., Miller, D.H., Reingold, S.C., 2009. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. Nature Reviews Neurology 5, 256–266.
Battaglini, M., Smith, S.M., Brogi, S., De Stefano, N., 2008. Enhanced brain extraction improves the accuracy of brain atrophy estimation. Neuroimage 40, 583–589.
Bermel, R.A., Bakshi, R., 2006. The measurement and clinical relevance of brain atrophy in multiple sclerosis. Lancet Neurology 5, 158–170.
Caramanos, Z., Fonov, V.S., Francis, S.J., Narayanan, S., Pike, G.B., Collins, D.L., Arnold, D.L., 2010. Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. Neuroimage 49, 3498-3498.
Chupin, M., Gerardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehericy, S., Benali, H., Garnero, L., Colliot, O., Alzheimer's Dis Neuroimaging Initi, 2009. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19, 579–587.
Clarkson, M.J., Ourselin, S., Nielsen, C., Leung, K.K., Barnes, J., Whitwell, J.L., Gunter, J.L., Hill, D.L., Weiner, M.W., Jack, C.R., Fox, N.C., 2009. The Alzheimer's disease neuroimaging initiative. Comparison of phantom and registration scaling corrections using the ADNI cohort. Neuroimage 47, 1506–1513.
De Stefano, N., Battaglini, M., Smith, S.M., 2007. Measuring brain atrophy in multiple sclerosis. Journal of Neuroimaging 17 (Suppl. 1), 10S–15S.
Evans, M.C., Barnes, J., Nielsen, C., Kim, L.G., Clegg, S.L., Blair, M., Leung, K.K., Douiri, A., Boyes, R.G., Ourselin, S., Fox, N.C., 2010. Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. European Radiology 20, 674–682.
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
Fox, N.C., Freeborough, P.A., Rossor, M.N., 1996a. Visualisation and quantification of rates of atrophy in Alzheimer's disease. Lancet 348, 94–97.
Fox, N.C., Warrington, E.K., Freeborough, P.A., Hartikainen, P., Kennedy, A.M., Stevens, J.M., Rossor, M.N., 1996b. Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. Brain 119 (Pt 6), 2001–2007.
Fox, N.C., Scahill, R.I., Crum, W.R., Rossor, M.N., 1999. Correlation between rates of brain atrophy and cognitive decline in AD. Neurology 52, 1687–1689.
Fox, N.C., Jenkins, R., Leary, S.M., Stevenson, V.L., Losseff, N.A., Crum, W.R., Harvey, R.J., Rossor, M.N., Miller, D.H., Thompson, A.J., 2000. Progressive cerebral atrophy in MS: a serial study using registered, volumetric MRI. Neurology 54, 807–812.
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 Transactions on Medical Imaging 16, 623–629.
Frisoni, G.B., Fox, N.C., Jack Jr., C.R., Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology 6 (2), 67–77.
Gunter, J.L., Bernstein, M.A., Borowski, B.J., Ward, C.P., Britson, P.J., Felmlee, J.P., Schuff, N., Weiner, M., Jack, C.R., 2009. Measurement of MRI scanner performance with the ADNI phantom. Medical Physics 36, 2193–2205.

Holland, D., Brewer, J.B., Hagler, D.J., Fenema-Notestine, C., Dale, A.M., Alzheimer's Dis Neuroimaging Initi, 2009. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America 106, 20954–20959.

Horsfield, M.A., Rovaris, M., Rocca, M.A., Rossi, P., Benedict, R.H., Filippi, M., Bakshi, R., 2003. Whole-brain atrophy in multiple sclerosis measured by two segmentation processes from various MRI sequences. Journal of the Neurological Sciences 216, 169–177.

Jack Jr., C.R., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1998. Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. Neurology 51, 993–999.

Jack, C.R.J., Petersen, R.C., Xu, Y.C., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Waring, S.C., Tangalos, E.G., Kokmen, E., 1999. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. Neurology 52 (7), 1397–1407.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging 27 (4), 685–691.

Jasperse, B., Valsasina, P., Neacsu, V., Knol, D.L., De Stefano, N., Enzinger, C., Smith, S.M., Ropele, S., Korteweg, T., Giorgio, A., Anderson, V., Polman, C.H., Filippi, M., Miller, D.H., Rovaris, M., Barkhof, F., Vrenken, H., 2007. Intercenter agreement of brain atrophy measurement in multiple sclerosis patients using manually-edited SIENA and SIENAX. Journal of Magnetic Resonance Imaging 26, 881–885.

Killiany, R.J., Gomez-Isla, T., Moss, M., Kikinis, R., Sandor, T., Jolesz, F., Tanzi, R., Jones, K., Hyman, B.T., Albert, M.S., 2000. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease [see comments]. Annals of Neurology 47, 430–439.

Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., 2009. Evaluation of automated brain MR image segmentation and volumetry methods. Human Brain Mapping 30, 1310–1327.

Korteweg, T., Rovaris, M., Neacsu, V., Filippi, M., Comi, G., Uitdehaag, B.M., Knol, D., Polman, C.H., Barkhof, F., Vrenken, H., MAGNIMS Collaboration, 2009. Can rate of brain atrophy in multiple sclerosis be explained by clinical and MRI characteristics? Multiple Sclerosis 15, 465–471.

Kruggel, F., Turner, J., Muftuler, L.T., Alzheimers Dis Neuroimaging Initia, 2010. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49, 2123–2133.

Losseff, N.A., Wang, L., Lai, H.M., Yoo, D.S., GawneCain, M.L., McDonald, W.I., Miller, D.H., Thompson, A.J., 1996. Progressive cerebral atrophy in multiple sclerosis. A serial MRI study. Brain 119, 2009–2019.

Miller, D.H., Barkhof, F., Frank, J.A., Parker, G.J., Thompson, A.J., 2002. Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. Brain 125, 1676–1695.

Molyneux, P.D., Kappos, L., Polman, C., Pozzilli, C., Barkhof, F., Filippi, M., Yousry, T., Hahn, D., Wagner, K., Ghazi, M., Beckmann, K., Dahlke, F., Losseff, N., Barker, G.J., Thompson, A.J., Miller, D.H., 2000. The effect of interferon beta-1b treatment on MRI measures of cerebral atrophy in secondary progressive multiple sclerosis. Brain 123, 2256–2263.

Mooney, C.Z., Duval, R.D., 1985. Bootstrapping: a Nonparametric Approach to Statistical Inference. Sage Publications, Newbury Park, Ca.

Morra, J.H., Tu, Z., Apostolova, L., Green, A., Avedissian, C., Madsen, S., Parikshak, N., Hua, X., Toga, A., Jack, C., Weiner, M., Thompson, P., Alzhei, The, 2008. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. Neuroimage 43, 59–68.

Morra, J.H., Tu, Z.W., Apostolova, L.G., Green, A.E., Toga, A.W., Thompson, P.M., 2010. Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. IEEE Transactions on Medical Imaging 29, 30–43.

Mugler, J.P., Brookeman, J.R., 1990. 3-Dimensional magnetization-prepared rapid gradient-echo imaging (3D MP-RAGE). Magnetic Resonance in Medicine 15, 152–157.

Neacsu, V., Jasperse, B., Korteweg, T., Knol, D.L., Valsasina, P., Filippi, M., Barkhof, F., Rovaris, M., Vrenken, H., MAGNIMS Study Grp, 2008. Agreement between different input image types in brain atrophy measurement in multiple sclerosis using SIENAX and SIENA. Journal of Magnetic Resonance Imaging 28, 559–565.

Rudick, R.A., Fisher, E., Lee, J.C., Simon, J., Jacobs, L., 1999. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing–remitting MS. Neurology 53, 1698–1704.

Schott, J.M., Frost, C., Whitwell, J.L., Macmanus, D.G., Boyes, R.G., Rossor, M.N., Fox, N.C., 2006. Combining short interval MRI in Alzheimer's disease: implications for therapeutic trials. Journal of Neurology 253, 1147–1153.

Sluimer, J.D., Vrenken, H., Blankenstein, M.A., Fox, N.C., Scheltens, P., Barkhof, F., van der Flier, W.M., 2008. Whole-brain atrophy rate in Alzheimer disease: identifying fast progressors. Neurology 70, 1836–1841.

Sluimer, J.D., van der Flier, W.M., Karas, G.B., van Schijndel, R., Barnes, J., Boyes, R.G., Cover, K.S., Olabarriaga, S.D., Fox, N.C., Scheltens, P., Vrenken, H., Barkhof, F., 2009. Accelerating regional atrophy rates in the progression from normal aging to Alzheimer's disease. European Radiology 19, 2826–2833.

Sluimer, J.D., Bouwman, F.H., Vrenken, H., Blankenstein, M.A., Barkhof, F., van der Flier, W.M., Scheltens, P., 2010. Whole-brain atrophy rate and CSF biomarker levels in MCI and AD: a longitudinal study. Neurobiology of Aging 31, 758–764.

Smith, S.M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17, 143–155.

Smith, S.M., De Stefano, N., Jenkinson, M., Matthews, P.M., 2001. Normalized accurate measurement of longitudinal brain change. Journal of Computer Assisted Tomography 25, 466–475.

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

Smith, S.M., Rao, A., De Stefano, N., Jenkinson, M., Schott, J.M., Matthews, P.M., Fox, N.C., 2007. Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: cross-validation of BSI, SIENA and SIENAX. Neuroimage 36, 1200–1206.

Sormani, M.P., Rovaris, M., Valsasina, P., Wolinsky, J.S., Comi, G., Filippi, M., 2004. Measurement error of two different techniques for brain atrophy assessment in multiple sclerosis. Neurology 62, 1432–1434.

Zivadinov, R., Grop, A., Sharma, J., Bratina, A., Tjoa, C.W., Dwyer, M., Zorzon, M., 2005. Reproducibility and accuracy of quantitative magnetic resonance imaging techniques of whole-brain atrophy measurement in multiple sclerosis. Journal of Neuroimaging 15, 27–36.