Journal of Alzheimer's Disease xx (2021) x–xx
DOI 10.3233/JAD-210596
IOS Press

Reproducibility of Brain Volume Changes in Longitudinal Voxel-Based Morphometry Between Non-Accelerated and Accelerated Magnetic Resonance Imaging

Hidemasa Takao*, Shiori Amemiya and Osamu Abe for the Alzheimer's Disease Neuroimaging Initiative[1]
Department of Radiology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan

Accepted 11 June 2021
Pre-press 21 July 2021

Abstract.
Background: Scan acceleration techniques, such as parallel imaging, can reduce scan times, but reliability is essential to implement these techniques in neuroimaging.
Objective: To evaluate the reproducibility of the longitudinal changes in brain morphology determined by longitudinal voxel-based morphometry (VBM) between non-accelerated and accelerated magnetic resonance images (MRI) in normal aging, mild cognitive impairment (MCI), and Alzheimer's disease (AD).
Methods: Using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) 2 database, comprising subjects who underwent non-accelerated and accelerated structural T1-weighted MRI at screening and at a 2-year follow-up on 3.0 T Philips scanners, we examined the reproducibility of longitudinal gray matter volume changes determined by longitudinal VBM processing between non-accelerated and accelerated imaging in 50 healthy elderly subjects, 54 MCI patients, and eight AD patients.
Results: The intraclass correlation coefficient (ICC) maps differed among the three groups. The mean ICC was 0.72 overall (healthy elderly, 0.63; MCI, 0.75; AD, 0.63), and the ICC was good to excellent (0.6–1.0) for 81.4% of voxels (healthy elderly, 64.8%; MCI, 85.0%; AD, 65.0%). The differences in image quality (head motion) were not significant (Kruskal–Wallis test, $p = 0.18$) and the within-subject standard deviations of longitudinal gray matter volume changes were similar among the groups.
Conclusion: The results indicate that the reproducibility of longitudinal gray matter volume changes determined by VBM between non-accelerated and accelerated MRI is good to excellent for many regions but may vary between diseases and regions.

Keywords: Acceleration, aging, Alzheimer's disease, gray matter, intraclass correlation coefficient, mild cognitive impairment, morphology, parallel imaging, reliability, stability

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://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.

*Correspondence to: Hidemasa Takao, MD, PhD, Department of Radiology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. Tel.: +81 3 5800 8666; Fax: +81 3 5800 8935; E-mail: takaoh-tky@umin.ac.jp.

INTRODUCTION

In recent years, longitudinal structural magnetic resonance imaging (MRI) has become widely used to estimate the rate of brain atrophy during normal aging and in a variety of neurodegenerative disorders. Between-subject morphological differences are usually significantly greater than the within-subject morphological changes. Extensive between-subject variability in brain morphology reduces the sensitivity for detecting changes in brain morphology. Longitudinal structural MRI reduces the variability associated with the between-subject differences in brain morphology by using the individual subjects as their own controls. This may avoid some of the problems caused by secular trends and between-subject variation. However, the statistical power to detect changes in brain morphology can be limited by measurement errors. Nevertheless, to quantify the changes in brain morphology from serial MRI scans in a precise manner, it is important that the acquisition conditions at baseline and at subsequent scans are as similar as possible.

Sufficient reliability is essential when using neuroimaging as a potential biomarker of neurodegenerative disorders, especially when monitoring longitudinal changes and treatment effects. Many previous studies have evaluated the reliability of structural T1-weighted imaging [1–15] and diffusion imaging [16–23]. Scan acceleration techniques, such as parallel imaging, can reduce scan times and are especially useful in subjects who cannot tolerate longer scans, and are therefore widely used in neuroimaging. Parallel imaging shortens scan times (typically by a factor of 2 to 3) by a reduction in the number of phase-encoding steps during image acquisition using the spatial information inherent in receiver coils. On the other hand, shorter scan times may cause a reduced signal-to-noise ratio and parallel imaging relies on the accuracy of the coil calibration data. However, few studies have investigated the effects of scan acceleration on the estimated longitudinal changes in brain morphology [24–28]. In addition, we are unaware of any studies that have fully investigated the reproducibility of longitudinal changes in brain morphology between non-accelerated and accelerated imaging on a voxel-wise basis. It is also unclear whether the type of disease affects the reproducibility.

We obtained 3.0 T structural T1-weighted MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database to determine the reproducibility (i.e., variation due to different scan sequences) of the longitudinal (2-year) changes in brain morphology, measured by longitudinal voxel-based morphometry (VBM), between non-accelerated and accelerated scans in healthy elderly subjects, patients with mild cognitive impairment (MCI), and patients with Alzheimer's disease (AD).

MATERIALS AND METHODS

Subjects

This study used data from the ADNI 2 database (available at http://adni.loni.usc.edu) comprising subjects who underwent non-accelerated and accelerated structural T1-weighted MRI at screening and at a 2-year follow-up (i.e., 2 [1 non-accelerated and 1 accelerated] scans $\times$ 2 time-points per subject) on 3.0 T Philips scanners. The study included 112 subjects: 50 healthy control subjects, 54 patients with MCI, and eight patients with AD. The mean age (range) at screening was 72.3 ± 6.3 years (healthy control subjects, 72.5 ± 5.4 years [64.1–83.7 years]; patients with MCI, 71.2 ± 6.8 years [56.7–88.7 years]; patients with AD, 78.1 ± 5.5 years [70.3–86.6 years]). The mean scan interval (range) was 2.1 ± 0.1 years (healthy control subjects, 2.1 ± 0.1 years [1.9–2.4 years]; patients with MCI, 2.0 ± 0.1 years [1.8–2.2 years]; patients with AD, 2.0 ± 0.04 years [2.0–2.1 years]).

The ADNI was launched in 2003 as a public–private partnership, led by the Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI was to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The ADNI was approved by the institutional review boards of all participating sites. Written informed consent was obtained from all participants.

Imaging data acquisition

MRI scans were performed using 3.0 T Philips scanners at multiple sites and the same ADNI 3.0 T imaging protocol (http://adni.loni.usc.edu). Various models of scanners were used, but each subject was scanned at screening and follow-up using the same scanner. Non-accelerated structural T1-weighted images were acquired using a three-dimensional (3D) magnetization-prepared rapid gradient-echo (MP-RAGE) sequence in 170 sagittal slices

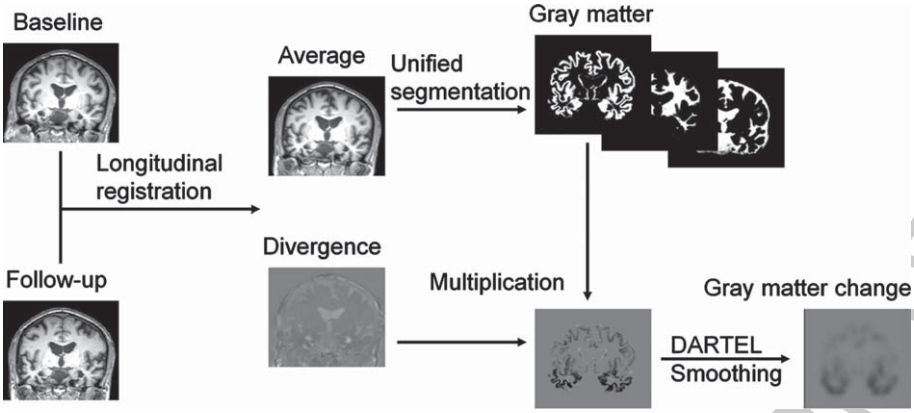

Fig. 1. Overview of the longitudinal voxel-based morphometry (VBM) conducted using statistical parametric mapping (SPM) 12 software. DARTEL, Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra.

(repetition time = 6.8 ms; echo time = 3.1 ms; inversion time = 900 ms; flip angle = 9°; field of view = 256 × 240 mm; slice thickness = 1.2 mm with no gap; acquisition matrix = 256 × 240; image matrix = 256 × 256; reconstructed voxel size = 1.0 × 1.0 × 1.2 mm; scan time = 9 : 06). Accelerated structural T1-weighted images were acquired using the 3D MP-RAGE sequence with sensitivity encoding (SENSE) acceleration (phase reduction = 1, phase oversampling factor = 1.5, slice reduction = 1.8) in 170 sagittal slices (repetition time = 6.8 ms; echo time = 3.1 ms; inversion time = 900 ms; flip angle = 9°; field of view = 270 × 253 mm; slice thickness = 1.2 mm with no gap; acquisition matrix = 244 × 227; image matrix = 256 × 256; reconstructed voxel size = 1.05 × 1.05 × 1.2 mm; scan time = 5 : 34). B1 non-uniformity correction was integrated into the sequences and correction for gradient non-linearity distortion was not applied because of the linearity of Phillips gradient systems. The non-parametric non-uniform intensity normalization algorithm N3 was used to correct the MP-RAGE images for non-uniform intensity [29–31].

The quality of the MP-RAGE images was subjectively graded as good, adequate, or poor by three radiologists with 21 (H.T.), 10, and 2 years of experience in neuroradiology independently and in a blinded manner. In case of disagreements, final evaluations were made by consensus.

Image processing

Image processing was primarily performed using statistical parametric mapping (SPM) 12 software developed in the Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London and MATLAB 9.1 (Mathworks, Sherborn, MA). The image processing steps described below are summarized in Fig. 1.

Longitudinal registration of pairs (obtained at screening and 2 years later) of MP-RAGE images was performed by pairwise inverse-consistent alignment between the first and second scans for each subject, while incorporating bias field correction [32] to calculate the mid-point average images and to map the divergences in velocity fields (representing the rates of volumetric expansion/contraction). The mid-point average images were segmented into gray matter, white matter, and cerebrospinal fluid using the unified segmentation algorithm [33], and using the International Consortium for Brain Mapping gray matter, white matter, cerebrospinal fluid, bone, soft tissue, and air/background templates as priors. The segmented gray matter and white matter images, and the maps of longitudinal gray matter volume changes, which were calculated by multiplying the gray matter images by the divergence maps, were spatially normalized using the Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) algorithm [34]. The normalized images were modulated to correct voxel intensity for volume displacement during normalization to reflect brain volume, and were smoothed using an 8 mm kernel.

Statistical analysis

To examine the reproducibility of the longitudinal changes in gray matter volume between non-accelerated and accelerated structural T1-weighted

imaging, the intraclass correlation coefficient (ICC) was calculated for each voxel using a single-measurement, absolute-agreement, two-way mixed-effects model [35, 36] in MATLAB 9.1, as follows:

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)\,MS_E + \frac{k}{n}\,(MS_C - MS_E)}$$

where

$$MS_R \text{ (mean square for rows)} = \frac{SS_R}{n-1}$$

$$MS_C \text{ (mean square for columns)} = \frac{SS_C}{k-1}$$

$$MS_E \text{ (mean square for error)} = \frac{SS_E}{(n-1)\,(k-1)}$$

$$SS_T \text{ (total sum of squares)} = \sum x_T^2 - \frac{\left(\sum x_T\right)^2}{N}$$

$$SS_R \text{ (sum of squares for rows)} = \sum_i^n \frac{\left(\sum x_i\right)^2}{k} - \frac{\left(\sum x_T\right)^2}{N}$$

$$SS_C \text{ (sum of squares for columns)} = \sum_j^k \frac{\left(\sum x_j\right)^2}{n} - \frac{\left(\sum x_T\right)^2}{N}$$

$$SS_E \text{ (sum of squares for error)} = SS_T - SS_R - SS_C$$

$$\sum x_T = \sum_i^n \sum_j^k x_{ij}, \quad \sum x_T^2 = \sum_i^n \sum_j^k x_{ij}^2$$

$$N = n \times k, \quad n = \text{number of subjects (rows)}$$

$$k = \text{number of measurements (columns)}$$

(here 2 = 1 non-accelerated + 1 acceraleted)

Histogram analysis was performed for each ICC map with a histogram bin width of 0.002 and a range of –1.0 to 1.0. Only voxels with a volume of > 0.05 on all gray matter images were included in the ICC calculation and histogram analysis. The ICC was interpreted using Cicchetti's criteria, which classify an ICC of <0.40 as poor, 0.40–0.59 as fair, 0.60–0.74 as good, and 0.75–1.00 as excellent [37].

The mean and within-subject standard deviation images of longitudinal gray matter volume changes were calculated from non-accelerated and accelerated images. The standard deviation images of longitudinal gray matter volume changes were calculated from non-accelerated images.

To evaluate the effect of image quality on the reproducibility of longitudinal changes in gray matter volume between non-accelerated and accelerated imaging, we used the Kruskal–Wallis test to compare the image quality among healthy control subjects, patients with MCI, and patients with AD using SPSS Statistics 22 (IBM, Armonk, NY). The significance level was set at $p < 0.05$.

RESULTS

ICC maps and histogram analysis

The voxel-wise ICC maps of the longitudinal changes in gray matter volume over 2 years for reproducibility between non-accelerated and accelerated imaging in healthy control subjects, patients with MCI, and patients with AD are shown in Fig. 2. The results of the histogram analysis (frequency polygons) of the ICC maps are shown in Fig. 3. The ICC maps and their frequency polygons differed among the three groups of subjects. The mean ICC was 0.72 overall (0.63 for healthy control subjects, 0.75 for patients with MCI, and 0.63 for patients with AD). The median ICC was 0.75 overall (0.66 for healthy control subjects, 0.79 for patients with MCI, and 0.71 for patients with AD). The histogram peak was 0.81 overall (0.70 for healthy control subjects, 0.84 for patients with MCI, and 0.85 for patients with AD). The distribution of the voxel-wise ICC estimates is summarized in Table 1. Overall, the reproducibility was excellent (ICC = 0.75–1.00) for 49.3% of voxels (23.6% for healthy control subjects, 60.2% for patients with MCI, and 43.3% for patients with AD). The reproducibility was good to excellent (ICC = 0.60–1.00) for 81.4% of voxels (64.8% for healthy control subjects, 85.0% for patients with MCI, and 65.0% for patients with AD).

Longitudinal changes in gray matter volume at 2 years

The mean longitudinal changes in gray matter volume at 2 years in the three groups of subjects are shown in Fig. 4. The results of the histogram analysis (frequency polygons; histogram bin width = 0.0002,

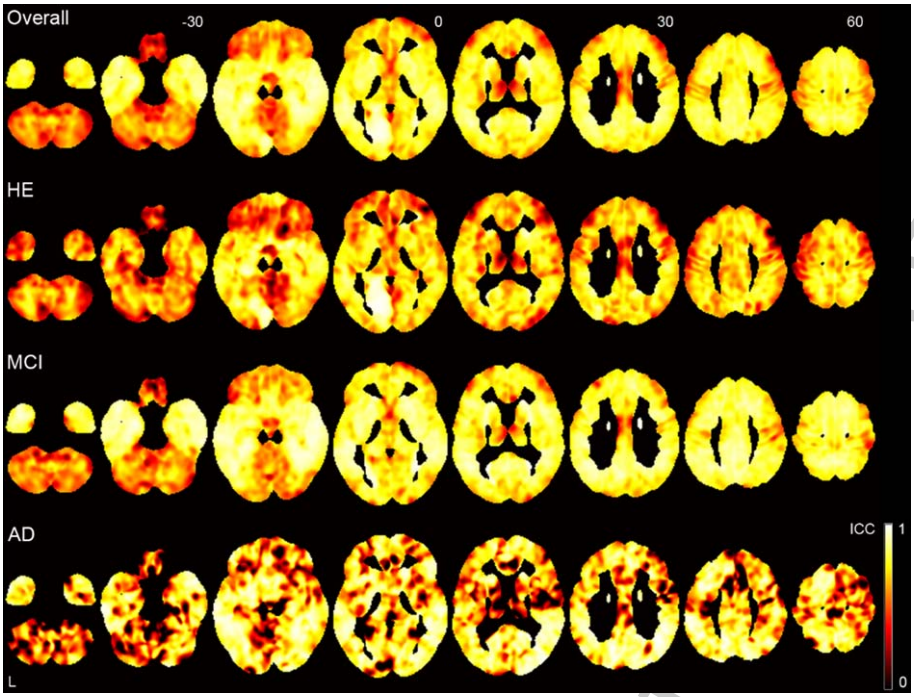

Fig. 2. Voxel-wise ICC maps of the longitudinal changes in gray matter volume over 2 years for reproducibility between non-accelerated and accelerated imaging. ICC, intraclass correlation coefficient; HE, healthy elderly; MCI, mild cognitive impairment; AD, Alzheimer's disease.

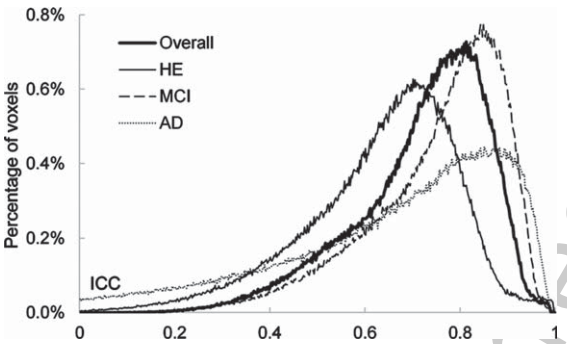

Fig. 3. Histograms (frequency polygons) of the voxel-wise ICC maps of longitudinal changes in gray matter volume over 2 years for reproducibility between non-accelerated and accelerated imaging. ICC, intraclass correlation coefficient; HE, healthy elderly; MCI, mild cognitive impairment; AD, Alzheimer's disease.

Table 1
The distribution of voxel-wise ICC estimates of the longitudinal changes in gray matter volume at 2 years for reproducibility between non-accelerated and accelerated imaging

ICC	Poor (0.00–0.39)	Fair (0.40–0.59)	Good (0.60–0.74)	Excellent (0.75–1.00)
Overall	2.9%	15.7%	32.1%	49.3%
HE	9.2%	26.0%	41.2%	23.6%
MCI	2.5%	12.5%	24.8%	60.2%
AD	18.2%	16.9%	21.7%	43.3%

ICC, intraclass correlation coefficient; HE, healthy elderly; MCI, mild cognitive impairment; AD, Alzheimer's disease.

range = –0.1 to 0.1) are shown in Fig. 5. The patterns of gray matter atrophy that were detected by accelerated imaging closely matched those detected by non-accelerated imaging. The extent of gray matter atrophy over 2 years was greater in patients with MCI than in healthy control subjects, and was also greater in patients with AD than in patients with MCI. In patients with MCI and AD, gray matter atrophy was especially prominent in the temporal lobe, including the hippocampus and parahippocampal cortex, the posterior cingulate cortex, the precuneus, and the orbitofrontal cortex.

The standard deviations and within-subject standard deviations of longitudinal changes in gray matter volume over 2 years are shown for healthy control subjects, patients with MCI, and patients with AD in Fig. 6. As a whole, the variability of longitudinal volume changes was larger in patients with MCI than in healthy control subjects or patients with AD. On the other hand, the within-subject variability was almost the same among the three groups of subjects.

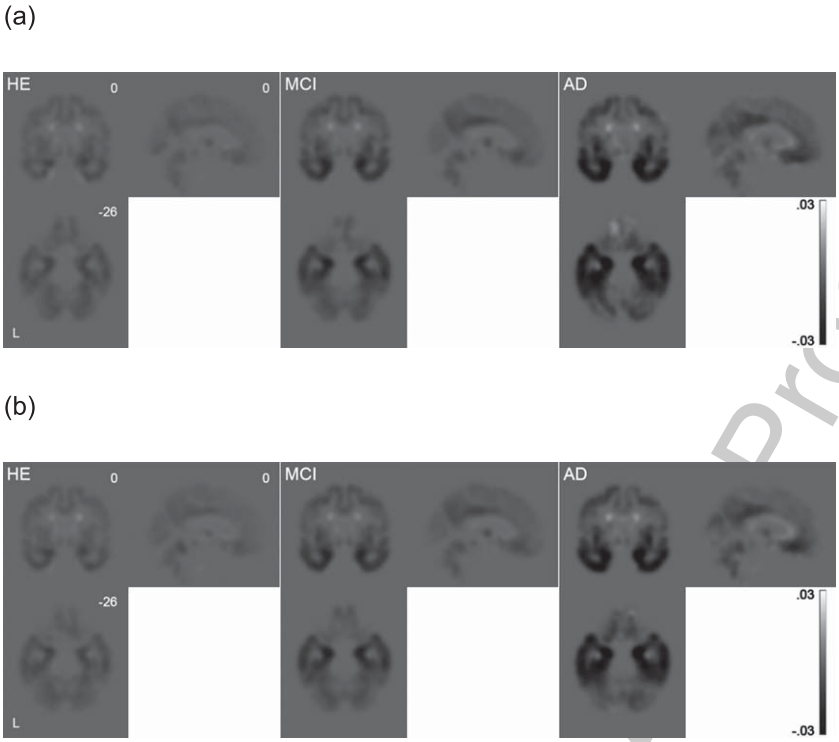

Fig. 4. Mean longitudinal changes in gray matter volume over 2 years derived from (a) non-accelerated and (b) accelerated imaging. HE, healthy elderly; MCI, mild cognitive impairment; AD, Alzheimer's disease.

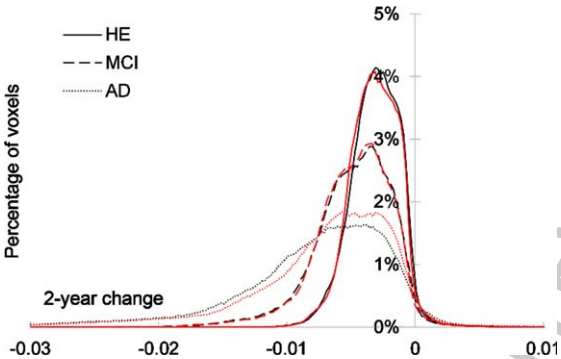

Fig. 5. Histograms (frequency polygons) of the mean longitudinal changes in gray matter volume over 2 years derived from non-accelerated (black) and accelerated (red) imaging. HE, healthy elderly; MCI, mild cognitive impairment; AD, Alzheimer's disease.

Image quality

The distribution of image quality (classified as good, adequate, or poor) in each group of subjects is shown in Fig. 7. The distribution of image quality were not significantly different among the healthy control subjects, patients with MCI, and patients with AD (Kruskal–Wallis test, $p = 0.18$).

DISCUSSION

In this study, we determined the reproducibility of the longitudinal (2-year) changes in brain morphology, measured by longitudinal VBM, between non-accelerated and accelerated structural T1-weighted imaging in healthy elderly subjects, patients with MCI, and patients with AD. The reproducibility of the longitudinal changes in gray matter volume between non-accelerated and accelerated imaging was rated as good to excellent for 81.4% of voxels as a whole. The distribution of image quality was not significantly different among the three groups of subjects, which was possibly due to not much difference in head motion, and the within-subject variability of longitudinal changes in gray matter volume was almost the same among the three groups. The differences in the ICCs among healthy elderly subjects, patients with MCI, and patients with AD were largely due to the differences in the variability of longitudinal changes in gray matter volume because of no significant difference in image quality among the three groups in this study.

Some studies have investigated the effects of using acceleration during structural T1-weighted

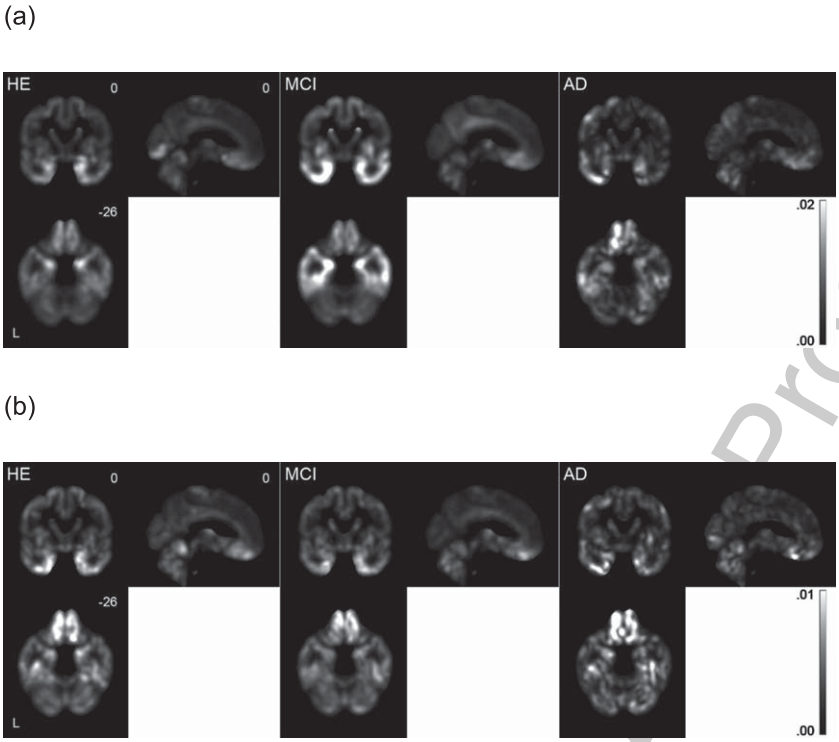

Fig. 6. Standard deviations (a) and within-subject standard deviations (b) of longitudinal changes in gray matter volume over 2 years. HE, healthy elderly; MCI, mild cognitive impairment; AD, Alzheimer's disease.

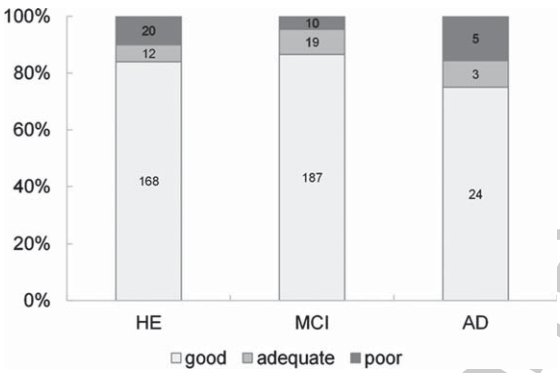

Fig. 7. Distribution of image quality. HE, healthy elderly; MCI, mild cognitive impairment; AD, Alzheimer's disease.

imaging on the estimated longitudinal changes in brain morphology [24–28]. Ching et al. compared the longitudinal brain changes detected by accelerated and non-accelerated scans using tensor-based morphometry and ADNI data [24]. They found no significant difference in the region-of-interest summaries of atrophy rates determined using accelerated and non-accelerated scans taken at 6- and 12-month intervals. Although voxel-wise analysis revealed some apparent regional differences in the atrophy rates at 6 months, there were no differences at 12 months. Leung et al. used ADNI data to investigate the impact of switching from non-accelerated to accelerated MRI over a 12-month interval on whole-brain atrophy measured using the k-means normalized boundary shift integral and deformation-based morphometry [25]. They found that switching from non-accelerated scans at baseline to accelerated scans at follow-up had a relatively minor effect on the computed atrophy rates, although the effect was dependent on the exact sequence details and the scanner manufacturer [25]. Vemuri et al. compared the tensor-based morphometry summary scores between accelerated and non-accelerated scan pairs for the annualized structural changes in a region characteristically affected in AD, also using ADNI data [26]. They found several systematic differences between the summary scores computed from accelerated and non-accelerated scan pairs. However, the accelerated scans showed a comparable performance to non-accelerated scans for discriminating among groups of patients. In this study, we evaluated the reproducibility of the longitudinal changes in brain morphology over 2 years between non-accelerated

and accelerated structural T1-weighted imaging on a voxel-wise basis using longitudinal VBM and data from the ADNI database. The reproducibility between non-accelerated and accelerated imaging was good to excellent for 81.4% of voxels, but differed by diagnosis and by region.

Head motion contributes to the within-subject variability in various neuroimaging settings. Usually, head motion cannot be directly measured during structural T1-weighted imaging; however, it manifests as decreased image quality [38]. In this study, we could not directly measure head motion during imaging and instead evaluated head motion in terms of image quality. The distribution of image quality was not significantly different and the within-subject variability of longitudinal changes in gray matter volume was almost the same among the three groups of subjects. As long as there is no difference in head motion (image quality), the within-subject variability of longitudinal changes in gray matter volume may not be different among healthy elderly subjects, patients with MCI, and patients with AD. To the best of our knowledge, no prior study has evaluated the relationship between head motion (image quality) and the differences in the reproducibility of longitudinal changes in brain morphology between non-accelerated and accelerated imaging among diseases.

We used the longitudinal registration method [32] implemented in the SPM software to register the baseline and follow-up scans, and to calculate the longitudinal changes in brain volume. This method combines rigid alignment, diffeomorphic warping, and differential intensity non-uniformity correction with respect to a within-subject template that evolves into an average of these three aspects, and is constructed in a symmetric, transitive manner. In longitudinal studies of brain morphology, longitudinal image processing, which seeks to reduce the within-subject variability by integrating the information from scans taken at each time-point and calculating within-subject changes, is generally preferable to treating each scan at each time-point independently, an approach that is usually used in cross-sectional studies. However, longitudinal image processing can introduce bias if the scans taken at different time-points are not treated equivalently and symmetrically (i.e., the scans undergo different processing steps). To prevent bias from affecting the estimated longitudinal changes in brain morphology, it is essential to treat the sequential scans symmetrically; otherwise, longitudinal image processing can be damaging rather than useful. In this study using longitudinal VBM, we found evidence of longitudinal gray matter atrophy in regions similar to previous reports [39, 40].

In this study, the scan time was 9 minutes and 6 seconds for non-accelerated imaging and 5 minutes and 34 seconds for accelerated imaging. While scan acceleration, such as parallel imaging, can reduce scan times, shorter scan times may cause a reduced signal-to-noise ratio, which might affect the results of brain morphometry. On the other hand, longer scans may be more subject to the effect of head motion. This study showed that the reproducibility of longitudinal gray matter volume changes determined by VBM between non-accelerated and accelerated imaging was good to excellent for many regions. Accelerated imaging may be preferable to non-accelerated imaging especially in patients unable to tolerate longer scan times.

There are limitations to this study. First, the image quality was not significantly different among the three groups. However, this does not necessarily mean that the image quality was equivalent among the groups. Second, the number of patients with AD was smaller than those of healthy elderly subjects and patients with MCI, while the numbers of healthy elderly subjects and patients with MCI were almost the same. This may make the results in patients with AD somewhat noisier. Finally, various models of scanners at various sites were used in the ADNI. Although each subject underwent scans at screening and follow-up on the same scanner, the effect of site/scanner on longitudinal morphometric changes may exist, but this is somewhat beyond the scope of this study.

CONCLUSIONS

We determined the reproducibility of the longitudinal changes in brain morphology over 2 years, measured by longitudinal VBM, between non-accelerated and accelerated imaging in healthy elderly subjects, patients with MCI, and patients with AD using data from the ADNI database. Our results indicate that the reproducibility of the longitudinal changes in gray matter volume between non-accelerated and accelerated imaging is good to excellent for many regions of the brain but varies by disease and region.

ACKNOWLEDGMENTS

Data collection and sharing for this project was 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: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; 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.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. 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 (https://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-0596).

REFERENCES

[1] Schnack HG, van Haren NE, Hulshoff Pol HE, Picchioni M, Weisbrod M, Sauer H, Cannon T, Huttunen M, Murray R, Kahn RS (2004) Reliability of brain volumes from multicenter MRI acquisition: A calibration study. *Hum Brain Mapp* **22**, 312-320.

[2] Ewers M, Teipel SJ, Dietrich O, Schonberg SO, Jessen F, Heun R, Scheltens P, van de Pol L, Freymann NR, Moeller HJ, Hampel H (2006) Multicenter assessment of reliability of cranial MRI. *Neurobiol Aging* **27**, 1051-1059.

[3] Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *Neuroimage* **32**, 180-194.

[4] Dickerson BC, Fenstermacher E, Salat DH, Wolk DA, Maguire RP, Desikan R, Pacheco J, Quinn BT, Van der Kouwe A, Greve DN, Blacker D, Albert MS, Killiany RJ, Fischl B (2008) Detection of cortical thickness correlates of cognitive performance: Reliability across MRI scan sessions, scanners, and field strengths. *Neuroimage* **39**, 10-18.

[5] Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, Pacheco J, Albert M, Killiany R, Blacker D, Maguire P, Rosas D, Makris N, Gollub R, Dale A, Dickerson BC, Fischl B (2009) MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage* **46**, 177-192.

[6] Huppertz HJ, Kroll-Seger J, Kloppel S, Ganz RE, Kassubek J (2010) Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. *Neuroimage* **49**, 2216-2224.

[7] Takao H, Hayashi N, Ohtomo K (2011) Effect of scanner in longitudinal studies of brain volume changes. *J Magn Reson Imaging* **34**, 438-444.

[8] Takao H, Hayashi N, Ohtomo K (2013) Effects of the use of multiple scanners and of scanner upgrade in longitudinal voxel-based morphometry studies. *J Magn Reson Imaging* **38**, 1283-1291.

[9] Cannon TD, Sun F, McEwen SJ, Papademetris X, He G, van Erp TG, Jacobson A, Bearden CE, Walker E, Hu X, Zhou L, Seidman LJ, Thermenos HW, Cornblatt B, Olvet DM, Perkins D, Belger A, Cadenhead K, Tsuang M, Mirzakhanian H, Addington J, Frayne R, Woods SW, McGlashan TH, Constable RT, Qiu M, Mathalon DH, Thompson P, Toga AW (2014) Reliability of neuroanatomical measurements in a multisite longitudinal study of youth at risk for psychosis. *Hum Brain Mapp* **35**, 2424-2434.

[10] Takao H, Hayashi N, Ohtomo K (2014) Effects of study design in multi-scanner voxel-based morphometry studies. *Neuroimage* **84**, 133-140.

[11] Takao H, Hayashi N, Ohtomo K (2015) Brain morphology is individual-specific information. *Magn Reson Imaging* **33**, 816-821.

[12] Biberacher V, Schmidt P, Keshavan A, Boucard CC, Righart R, Sämann P, Preibisch C, Fröbel D, Aly L, Hemmer B, Zimmer C, Henry RG, Mühlau M (2016) Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. *Neuroimage* **142**, 188-197.

[13] Lee H, Nakamura K, Narayanan S, Brown RA, Arnold DL (2019) Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements. *Neuroimage* **184**, 555-565.

[14] Melzer TR, Keenan RJ, Leeper GJ, Kingston-Smith S, Felton SA, Green SK, Henderson KJ, Palmer NJ, Shoorangiz R, Almuqbel MM, Myall DJ (2020) Test-retest reliability and sample size estimates after MRI scanner relocation. *Neuroimage* **211**, 116608.

[15] Takao H, Amemiya S, Abe O (2021) Reliability of changes in brain volume determined by longitudinal voxel-based morphometry. *J Magn Reson Imaging*, doi: 10.1002/jmri.27568.

[16] Heiervang E, Behrens TE, Mackay CE, Robson MD, Johansen-Berg H (2006) Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage* **33**, 867-877.

[17] Vollmar C, O'Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, Duncan JS, Richardson MP, Koepp MJ (2010) Identical, but not the same: Intra-site and

inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. *Neuroimage* **51**, 1384-1394.

[18] Takao H, Hayashi N, Ohtomo K (2011) Effect of scanner in asymmetry studies using diffusion tensor imaging. *Neuroimage* **54**, 1053-1062.

[19] Zhu T, Hu R, Qiu X, Taylor M, Tso Y, Yiannoutsos C, Navia B, Mori S, Ekholm S, Schifitto G, Zhong J (2011) Quantification of accuracy and precision of multi-center DTI measurements: A diffusion phantom and human brain study. *Neuroimage* **56**, 1398-1411.

[20] Lemkaddem A, Daducci A, Vulliemoz S, O'Brien K, Lazeyras F, Hauf M, Wiest R, Meuli R, Seeck M, Krueger G, Thiran JP (2012) A multi-center study: Intra-scan and inter-scan variability of diffusion spectrum imaging. *Neuroimage* **62**, 87-94.

[21] Takao H, Hayashi N, Kabasawa H, Ohtomo K (2012) Effect of scanner in longitudinal diffusion tensor imaging studies. *Hum Brain Mapp* **33**, 466-477.

[22] Wang JY, Abdi H, Bakhadirov K, Diaz-Arrastia R, Devous MD, Sr. (2012) A comprehensive reliability assessment of quantitative diffusion tensor tractography. *Neuroimage* **60**, 1127-1138.

[23] Takao H, Hayashi N, Ohtomo K (2015) Brain diffusivity pattern is individual-specific information. *Neuroscience* **301**, 395-402.

[24] Ching CR, Hua X, Hibar DP, Ward CP, Gunter JL, Bernstein MA, Jack CR, Jr., Weiner MW, Thompson PM (2015) Does MRI scan acceleration affect power to track brain change? *Neurobiol Aging* **36 Suppl 1**, S167-177.

[25] Leung KK, Malone IM, Ourselin S, Gunter JL, Bernstein MA, Thompson PM, Jack CR, Jr., Weiner MW, Fox NC (2015) Effects of changing from non-accelerated to accelerated MRI for follow-up in brain atrophy measurement. *Neuroimage* **107**, 46-53.

[26] Vemuri P, Senjem ML, Gunter JL, Lundt ES, Tosakulwong N, Weigand SD, Borowski BJ, Bernstein MA, Zuk SM, Lowe VJ, Knopman DS, Petersen RC, Fox NC, Thompson PM, Weiner MW, Jack CR, Jr. (2015) Accelerated vs. unaccelerated serial MRI based TBM-SyN measurements for clinical trials in Alzheimer's disease. *Neuroimage* **113**, 61-69.

[27] Hua X, Ching CRK, Mezher A, Gutman BA, Hibar DP, Bhatt P, Leow AD, Jack CR, Jr., Bernstein MA, Weiner MW, Thompson PM (2016) MRI-based brain atrophy rates in ADNI phase 2: Acceleration and enrichment considerations for clinical trials. *Neurobiol Aging* **37**, 26-37.

[28] Manning EN, Leung KK, Nicholas JM, Malone IB, Cardoso MJ, Schott JM, Fox NC, Barnes J (2017) A comparison of accelerated and non-accelerated MRI scans for brain volume and boundary shift integral measures of volume change: Evidence from the ADNI dataset. *Neuroinformatics* **15**, 215-226.

[29] Sled JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* **17**, 87-97.

[30] Takao H, Abe O, Hayashi N, Kabasawa H, Ohtomo K (2010) Effects of gradient non-linearity correction and intensity non-uniformity correction in longitudinal studies using structural image evaluation using normalization of atrophy (SIENA). *J Magn Reson Imaging* **32**, 489-492.

[31] Takao H, Abe O, Ohtomo K (2010) Computational analysis of cerebral cortex. *Neuroradiology* **52**, 691-698.

[32] Ashburner J, Ridgway GR (2013) Symmetric diffeomorphic modeling of longitudinal structural MRI. *Front Neurosci* **6**, 197-197.

[33] Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* **26**, 839-851.

[34] Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95-113.

[35] Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* **15**, 155-163.

[36] McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* **1**, 30-46.

[37] Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* **6**, 284-290.

[38] Savalia NK, Agres PF, Chan MY, Feczko EJ, Kennedy KM, Wig GS (2017) Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. *Hum Brain Mapp* **38**, 472-492.

[39] Frisoni GB, Prestia A, Rasser PE, Bonetti M, Thompson PM (2009) In vivo mapping of incremental cortical atrophy from incipient to overt Alzheimer's disease. *J Neurol* **256**, 916-924.

[40] Pini L, Pievani M, Bocchetta M, Altomare D, Bosco P, Cavedo E, Galluzzi S, Marizzoni M, Frisoni GB (2016) Brain atrophy in Alzheimer's disease and aging. *Ageing Res Rev* **30**, 25-48.