

Published in final edited form as:

Neuroimage. 2008 May 15; 41(1): 19–34. doi:10.1016/j.neuroimage.2008.02.010.

3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry

Xue Hua^a, Alex D. Leow^a, Suh Lee^a, Andrea D. Klunder^a, Arthur W. Toga^a, Natasha Lepore^a, Yi-Yu Chou^a, Caroline Brun^a, Ming-Chang Chiang^a, Marina Barysheva^a, Clifford R. Jack Jr.^b, Matt A. Bernstein^b, Paula J. Britson^b, Chadwick P. Ward^b, Jennifer L. Whitwell^b, Bret Borowski^b, Adam S. Fleisher^c, Nick C. Fox^d, Richard G. Boyes^d, Josephine Barnes^d, Danielle Harvey^e, John Kornak^f, Norbert Schuff^{g,i}, Lauren Boreta^g, Gene E. Alexander^j, Michael W. Weiner^{g,h,i}, Paul M. Thompson^{a,*}, and Alzheimer's Disease Neuroimaging Initiative

*a*Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Neuroscience Research Building 225E, 635 Charles Young Drive, Los Angeles, CA 90095-1769, USA

*b*Mayo Clinic College of Medicine, Rochester, MN, USA

*c*Department of Neurosciences, UC San Diego, La Jolla, CA, USA

*d*Dementia Research Centre, University College London, Institute of Neurology, London, UK

*e*Department of Public Health Sciences, UC Davis School of Medicine, Davis, CA, USA

*f*Department of Radiology and Department of Epidemiology and Biostatistics, UC San Francisco, San Francisco, CA, USA

*g*Department of Radiology, UC San Francisco, San Francisco, CA, USA

*h*Department of Medicine and Psychiatry, UC San Francisco, San Francisco, CA, USA

*i*Department of Veterans Affairs Medical Center, San Francisco

*j*Department of Psychology and Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, USA

Abstract

Tensor-based morphometry (TBM) creates three-dimensional maps of disease-related differences in brain structure, based on nonlinearly registering brain MRI scans to a common image template. Using two different TBM designs (averaging individual differences versus aligning group average templates), we compared the anatomical distribution of brain atrophy in 40 patients with Alzheimer's disease (AD), 40 healthy elderly controls, and 40 individuals with amnesic mild cognitive impairment (aMCI), a condition conferring increased risk for AD. We created an unbiased geometrical average image template for each of the three groups, which were matched for sex and age (mean age: 76.1 years+/-7.7 SD). We warped each individual brain image ($N=120$) to the control group average template to create Jacobian maps, which show the local expansion or compression factor at each point in the image, reflecting individual volumetric differences. Statistical maps of group differences revealed widespread medial temporal and limbic atrophy in AD, with a lesser, more restricted distribution in MCI. Atrophy and CSF space expansion both correlated strongly with Mini-Mental State Exam (MMSE) scores and Clinical Dementia Rating (CDR). Using cumulative

*Corresponding author. Fax: +1 310 206 5518.

E-mail address: thompson@loni.ucla.edu (P.M. Thompson).

Publisher's Disclaimer: This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

p-value plots, we investigated how detection sensitivity was influenced by the sample size, the choice of search region (whole brain, temporal lobe, hippocampus), the initial linear registration method (9- versus 12-parameter), and the type of TBM design. In the future, TBM may help to (1) identify factors that resist or accelerate the disease process, and (2) measure disease burden in treatment trials.

Introduction

Alzheimer's disease (AD) is the commonest form of dementia worldwide, afflicting over 5 million people in the United States alone. In early AD, memory is typically among the first functions to be impaired, followed by a progressive decline in executive function, language, affect, and other cognitive and behavioral domains. It would be beneficial to prevent AD progression before widespread neurodegeneration has occurred, so recent therapeutic efforts have also focused on individuals with mild cognitive impairment (MCI), a transitional state between normal aging and dementia that carries a 4–6-fold increased risk, relative to the general population, of future diagnosis of dementia (Petersen et al., 1999; Petersen, 2000; Petersen et al., 2001). Early detection requires innovations in tracking disease burden *in vivo* (Fleisher et al., 2007). Magnetic resonance imaging (MRI) and MRI-based image analysis methods have the potential to track brain atrophy automatically at multiple time-points. MRI has revealed fine-scale anatomical changes which are associated with cognitive decline and which occur in a spreading pattern that mirrors the advance of pathology (Thompson and Apostolova, in press). MRI-based maps of brain degeneration are beginning to reveal the distribution and evolution of cerebral volume losses, how brain changes in AD and other dementias relate to behavior, and which brain changes predict imminent decline (Scahill et al., 2003; Apostolova et al., 2006; Apostolova and Thompson, 2007).

Tensor-based morphometry (TBM) is a relatively new image analysis technique that identifies regional structural differences from the gradients of the deformation fields that align, or ‘warp’, images to a common anatomical template (reviewed in (Ashburner and Friston, 2003)). Highly automated methods such as TBM are being tested to examine their utility in large-scale clinical trials, and in studies to identify factors that influence disease onset, progression (Leow et al., 2005b; Cardenas et al., 2007), or normal development (Thompson et al., 2000a; Chung et al., 2001; Hua et al., in press). In TBM, a nonlinear registration algorithm reshapes each 3D structural image to match a target brain image – either based on an individual subject, or specially constructed to reflect the mean anatomy of a population (Kochunov et al., 2001, 2002; Lepore et al., 2007). Color-coded Jacobian maps – which show the local expansion or compression factor at each point in the image – indicate local volume loss or gain relative to a reference image (Freeborough and Fox, 1998; Chung et al., 2001; Fox et al., 2001; Ashburner and Friston, 2003; Riddle et al., 2004). TBM may also be used to map systematic anatomic differences between different patient groups using cross-sectional data (Davatzikos et al., 2003; Shen and Davatzikos, 2003; Studholme et al., 2004; Dubb et al., 2005; Brun et al., 2007; Chiang et al., 2007a,b; Lee et al., 2007; Lepore et al., 2008).

The traditional TBM design (Ashburner, 2007; Chiang et al., 2007a,b) computes individual Jacobian maps, i.e. “expansion factor maps”, from the non-linear registrations that align each subject's MRI image to a reference brain. Distinguishing features of group morphometry emerge after the maps of individual anatomical differences from the template are compared statistically across groups, or correlated with relevant clinical measures. This scheme may be called ‘*averaging individual differences*’ in the sense that the signal analyzed is based on maps of anatomical differences computed for every individual separately (Rohlfing et al., 2005). We use this term to distinguish it from an approach that directly aligns mean anatomical templates representing each group (Rohlfing et al., 2005; Aljabar et al., 2008). By contrast, when a Jacobian map is created for each subject – which is the standard TBM approach that we use to

report findings in this paper – correlations may be assessed between the detected individual differences and individual factors such as age, sex and clinical scores. We compare the standard and direct approaches later in this paper.

3D maps that define the level of atrophy (relative to appropriate controls) at a certain disease stage (Jack et al., 2005), may have value in staging the degenerative process, predicting outcomes, and understanding atrophic patterns characteristic of different dementia subtypes or stages, e.g. when individuals transition from MCI and AD. In this study, we examined the level of atrophy in AD and MCI relative to controls; we studied how specific methodological choices (e.g., sample size, initial linear registration) affected the statistical power to detect these differences; and we also investigated, at a voxelwise level, how brain atrophy correlated with clinical measures such as MMSE, and global Clinical Dementia Rating (CDR). Finally, we compared our results using the traditional TBM design with ones from directly aligning group average images – a relatively new concept in deformation-based group morphometry, which has been advocated recently in the literature (Rohlfing et al., 2005; Aljabar et al., 2006, 2008).

Materials and methods

Subjects

The Alzheimer's Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005a,b) is a large multi-site longitudinal MRI and FDG-PET (fluorodeoxyglucose positron emission tomography) study of 800 adults, ages 55 to 90, including 200 elderly controls, 400 subjects with mild cognitive impairment, and 200 patients with AD. 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 MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early 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, M.D., VA Medical Center and University of California – San Francisco.

At the time of writing this report, data collection for the ADNI project is in progress. Here we performed an initial analysis of the screening MRI scans of 120 subjects, divided into 3 groups: 40 healthy elderly individuals, 40 individuals with amnesic MCI, and 40 individuals with probable AD. Each group of 40 subjects was well matched in terms of gender and age: each group included 21 males and 19 females; mean ages for the control, MCI and AD groups were, respectively, 76.2 years (standard deviation (SD)=6.9 years), 75.9 years (SD=8.3), and 76.0 years (SD=8.5), with no significant age differences among the three groups (one-way ANOVA p -value=0.98).

To test whether each type of TBM design correctly detects no differences when no true differences are present, we selected an independent (second) group of normal subjects ($N=40$, mean age=76.0 years, SD=4.5 years), age- and gender-matched to the first group of controls. There was no overlap between this group and the initial normal group described above.

All subjects underwent thorough clinical/cognitive assessment at the time of scan acquisition. As part of each subject's cognitive evaluation, the Mini-Mental State Examination (MMSE) was administered to provide a global measure of mental status based on evaluation of five cognitive domains (Folstein et al., 1975; Cockrell and Folstein, 1988); scores of 24 or less (out

of a maximum of 30) are generally consistent with dementia. The Clinical Dementia Rating (CDR) was also assessed as a measure of dementia severity (Hughes et al., 1982; Morris, 1993). A global CDR of 0, 0.5, 1, 2 and 3, respectively, indicate no dementia, very mild, mild, moderate, and severe dementia. The elderly normal subjects had MMSE scores between 28 and 30 (inclusive), a global CDR of 0, and no symptoms of depression, MCI, or other forms of dementia. The MCI subjects had MMSE scores in the range of 24 to 30, a global CDR of 0.5, and mild memory complaints, with memory impairment assessed via education-adjusted scores on the Wechsler Memory Scale - Logical Memory II (Wechsler, 1987). All AD patients met NINCDS/ADRDA criteria for probable AD (McKhann et al., 1984) with an MMSE score between 20 and 23. As such, these subjects would be considered as having mild to moderate, but not severe, AD. Overall, ADNI included AD subjects with MMSE scores as high as 26, and a lower limit of 20, but we focused here on the 20–23 range of MMSE scores to identify a specific stage of AD at which a somewhat consistent level of atrophy might be identified. 16 AD patients had a CDR of 0.5, and the rest had a CDR of 1. Detailed exclusion criteria, e.g., regarding concurrent use of psychoactive medications, may be found in the ADNI protocol (Mueller et al., 2005a,b). Briefly, subjects were excluded if they had any serious neurological disease other than incipient AD, any history of brain lesions or head trauma, or psychoactive medication use (including antidepressants, neuroleptics, chronic anxiolytics or sedative hypnotics, etc.).

The study was conducted according to Good Clinical Practice, the Declaration of Helsinki and U.S. 21 CFR Part 50-Protection of Human Subjects, and Part 56-Institutional Review Boards. Written informed consent for the study was obtained from all participants before protocol-specific procedures, including cognitive testing, were performed.

MRI acquisition and image correction

All subjects were scanned with a standardized MRI protocol, developed after a major effort evaluating and comparing 3D T1-weighted sequences for morphometric analyses (Leow et al., 2006; Jack et al., in press).

High-resolution structural brain MRI scans were acquired at multiple ADNI sites using 1.5 Tesla MRI scanners from General Electric Healthcare and Siemens Medical Solutions. All scans were collected according to the standard ADNI MRI protocol. For each subject, two T1-weighted MRI scans were collected using a sagittal 3D MP-RAGE sequence. As described in (Jack et al., in press) typical 1.5T acquisition parameters are repetition time (TR) of 2400 ms, minimum full TE, inversion time (TI) of 1000 ms, flip angle of 8°, 24 cm field of view, acquisition matrix was 192×192×166 in the x-, y-, and z- dimensions yielding a voxel size of 1.25×1.25×1.2 mm³. In plane, zero-filled reconstruction (i.e., sinc interpolation) yielded a 256×256 matrix for a reconstructed voxel size of 0.9375×0.9375×1.2 mm³. The images were calibrated with phantom-based geometric corrections to ensure consistency among scans acquired at different sites (Gunter et al., 2006).

Additional image corrections were also applied, using a processing pipeline at the Mayo Clinic, consisting of: (1) a procedure termed *GradWarp* for correction of geometric distortion due to gradient non-linearity (Jovicich et al., 2006), (2) a “B1-correction”, to adjust for image intensity non-uniformity using B1 calibration scans (Jack et al., in press), (3) “N3” bias field correction, for reducing intensity inhomogeneity caused by non-uniformities in the radio frequency (RF) receiver coils (Sled et al., 1998), and (4) geometrical scaling, according to a phantom scan acquired for each subject (Jack et al., in press), to adjust for scanner- and session-specific calibration errors. In addition to the original uncorrected image files, images with all of these corrections already applied (*GradWarp*, B1, phantom scaling, and N3) are available to the general scientific community.

Image pre-processing

To adjust for global differences in brain positioning and scale across individuals, all scans were linearly registered to the stereotactic space defined by the International Consortium for Brain Mapping (ICBM-53) (Mazziotta et al., 2001) with a 9-parameter (9P) transformation (3 translations, 3 rotations, 3 scales) using the Minctracc algorithm (Collins et al., 1994). The Results section reports separate tests based on using 12-parameter affine registrations for the initial (global) component of the registration (which also allows shearing along x , y , and z axes). Globally aligned images were resampled in an isotropic space of 220 voxels along each axis (x , y , and z) with a final voxel size of 1 mm^3 .

Unbiased group average template - Minimal Deformation Target (MDT)

A minimal deformation target (MDT) is an unbiased average template image created to represent common anatomical features for a group of subjects, typically with a mathematically-defined mean geometry for a population (Good et al., 2001; Kochunov et al., 2002; Joshi et al., 2004; Studholme and Cardenas, 2004; Kovacevic et al., 2005; Christensen et al., 2006; Lorenzen et al., 2006; Lepore et al., 2007).

The motivation for constructing a mean geometric template, or ‘customized template’, based on subjects in the study is to make it easier to automatically register new scans to the template, to reduce bias in the registrations (using a template that deviates least from the anatomy of the subjects), and to improve statistical power, which has been shown to be slightly higher if a customized template is used (Lepore et al., 2007). To construct an MDT for the normal subject group, the 9-parameter globally aligned brain scans ($N=40$) were averaged voxel-by-voxel after intensity normalization to create an initial affine average template. Next, the aligned individual scans were non-linearly registered to the affine average template using a non-linear inverse consistent elastic intensity-based registration algorithm (Leow et al., 2005a,b). Satisfactory registration was achieved when a joint cost function was optimized, based on a linear combination of the mutual information (MI) between the deforming image and the target (affine average template) and the elastic energy of the deformation, which quantifies the irregularity of the deformation field. The deformation field was computed using a spectral method to implement the Cauchy-Navier elasticity operator (Marsden and Hughes, 1983; Thompson et al., 2000b) using a Fast Fourier Transform (FFT) resolution of $32 \times 32 \times 32$. This corresponds to an effective voxel size of 6.875 mm in the x , y , and z dimensions ($220 \text{ mm}/32=6.875 \text{ mm}$). The non-linear average image was then derived from the mean of the 40 individual scans that were non-linearly registered to the affine average template. Finally, we created the MDT for the normal group by applying inverse geometric centering of the displacement fields to the non-linear average (Kochunov et al., 2002, 2005). With the same procedure, we constructed a separate MDT for the MCI and AD groups. These MDTs are obtainable at: <http://www.loni.ucla.edu/~thompson/XUE/MDT/>.

In addition to the MDTs created based on a sample of size $N=40$, and 9P linear registration, we also investigated the effects of reducing the sample size on the statistical maps of group differences ($N=10, 20, 30$ subjects per group) using either 9-parameter (9P) or 12-parameter (12P) affine registration. For comparisons in reduced samples, the same MDT was still used (based on 40 subjects) to make sure that the results were spatially registered with each other.

Three-dimensional Jacobian maps

To quantify 3D patterns of volumetric brain atrophy in MCI and AD based on the method of “averaging individual differences” (Fig. 1a), all individual brains ($N=120$) were non-linearly aligned to the MDT for the normal group (Leow et al., 2005a). Subsequently, a separate Jacobian map was created for each subject to characterize the local volume differences between that individual and the normal group anatomical mean template. The determinant (local

expansion factor) of the local Jacobian matrix was derived from the forward deformation field (see (Lepore et al., 2008), for a more complex approach analyzing the full tensor). Color-coded Jacobian determinants were used to illustrate regions of volume expansion, i.e. those with $\det J(r) > 1$, or contraction, i.e., $J(r) < 1$ (Freeborough and Fox, 1998;Toga, 1999;Thompson et al., 2000a;Chung et al., 2001;Ashburner and Friston, 2003;Riddle et al., 2004) relative to the normal group template. Negative or zero-valued Jacobians are not obtained using this method, as the inverse-consistent implementation regularizes the inverse deformation mapping and causes the resulting Jacobian determinants to cluster quite tightly around zero after log transformation, as well as removing the skew and bias from their distribution (see Leow et al., 2005a,b,2007 for examination of the Jacobian distributions). As all images were registered to the same template, these Jacobian maps share a common anatomical coordinate defined by the normal template. Individual Jacobian maps within each group were averaged across subjects and compared statistically at each voxel to assess the magnitude and significance of deficits in MCI and AD versus the healthy controls.

We also examined group differences by directly aligning group average images (Fig. 1b), a concept first introduced by Rohlfing (Rohlfing et al., 2005). In this approach, an unbiased geometrical average template was created for each of the three groups, and each disease group average template was directly aligned to the control group average template to create single Jacobian maps from which to quantify inter-group local volume differences.

Subject-template alignment methods follow a similar pattern to MDT construction, and for elastic registration we used a FFT resolution of $32 \times 32 \times 32$; this corresponds to an effective size of 6.875 mm ($220 \text{ mm}/32 = 6.875 \text{ mm}$) in each of the x-, y-, and z- dimensions; for template-template registrations, we ran the deformations at a higher FFT resolution of 64, since the MDTs share anatomical features with very similar resolution and contrast. The choice of FFT grid size depends on the expected spatial coherence (autocorrelation) of features to be detected, and could be modeled, in a more complex approach, by empirical estimation of the bivariate Green's function or 6D Lambda-tensor ((Fillard et al., 2005); this approach will be tested once these covariance functions are estimable from a very large image database of images).

Statistical tests

The first approach generated 120 Jacobian maps, which encode individual differences with respect to the normal template. This enabled us to carry out voxel-wise statistical tests between the individual Jacobian maps in each group within a common coordinate system. The Jacobian maps in MCI and AD were compared to those from normal controls. At each voxel, we evaluated the significance level of group differences using a two-sample *t* test with unequal variance. The resulting *p*-values were displayed as maps to allow visualization of the patterns of significant differences throughout the brain.

In addition, we used permutation testing to assess the overall significance of group differences, corrected for multiple comparisons [see, e.g., Bullmore et al., 1999; Nichols and Holmes 2002; Thompson et al., 2003; Chiang et al., 2007a,b]. A null distribution for the group differences in Jacobian at each voxel was constructed using 10,000 random permutations of the data. For each test, the subjects' diagnosis was randomly permuted and voxel-wise *t* tests were conducted to identify voxels more significant than $p=0.05$. The volume of voxels in the brain more significant than $p=0.05$ was computed for the real experiment and for the random assignments. Finally, a ratio, describing the fraction of the time the suprathreshold volume was greater in the randomized maps than the real effect (the original labeling), was calculated to give an overall *P*-value for the significance of the map (corrected for multiple comparisons by permutation). The correction is for the number of tests, so it quantifies the level of surprise in seeing the overall map. The number of permutations *N* was chosen to be 10,000, to control the standard error *SE_p* of the omnibus probability *p*, which follows a binomial distribution $B(N,$

p) with known standard error (Edgington, 1995). When $N=10,000$, the approximate margin of error (95% confidence interval) for p is around 5% of p .

CDF plots

Cumulative distribution function (CDF) plots were used to compare the power of detecting significant effects when using the TBM design of averaging individual differences, with sample sizes varying from 10 to 40 per group, and the two different linear registration schemes. These CDF plots are commonly generated when using false discovery rate methods to assign overall significance values to statistical maps (Benjamini and Hochberg, 1995; Genovese et al., 2002; Storey, 2002); they may also be used to compare effect sizes of different methods, subject to certain caveats (Lepore et al., 2007), as they show the proportion of supra-threshold voxels in a statistical map, for a range of thresholds. A cumulative plot of p -values in a statistical map, after the p -values have been sorted into numerical order, can compare the proportion of suprathreshold statistics with null data, or between one method and another, to assess their power to detect statistical differences that survive thresholding at both weak and strict thresholds (in fact at any threshold in the range 0 to 1). In the examples shown here, the cumulative distribution function of the p -values observed for the statistical comparison of patients versus controls is plotted against the corresponding p -value that would be expected, under the null hypothesis of no group difference. For null distributions (comparing two independent normal groups), the cumulative distribution of p -values is expected to fall approximately along the diagonal line $y=x$, because a proportion y of voxels in a null p -value map will, on average, fall below the threshold y ; large upswings of the CDF from that diagonal line are associated with significant signal. Greater effect sizes are represented by larger deviations in these CDF plots (and the theory of false discovery rates gives formulae for thresholds that control false positives at a known rate).

Using the results of the above two-sample t -tests, Fig. 3 shows the cumulative histograms (CDF plots) of the probability maps for voxel-wise differences in mean Jacobian between the MCI and AD groups and normal controls. Within each CDF plot, the curves show increasing effect sizes, in rank order from bottom to top, for detecting voxels with statistical differences between groups.

Regions of interest (ROIs)

Regions of interest, including frontal, parietal, temporal, and occipital lobes, were defined by manually labeling the normal group MDT. The MDT was traced by a trained anatomist to generate binary masks for each lobe, which were subsequently used to summarize brain atrophy at a regional level in each group. Within each lobe, tissue types were distinguished by creating maps of gray and white matter, CSF, and non-brain tissues using the partial volume classification (PVC) algorithm from the *BrainSuite* software package (Shattuck and Leahy, 2002). CSF was excluded from the masks as the trend for CSF differences is typically opposite to cerebral differences in subjects with varying levels of brain atrophy, i.e., greater CSF space expansions are typically associated with greater atrophy. While these CSF signals are potentially of diagnostic interest (Carmichael et al., 2006, 2007), they were excluded to avoid confounding the average values in regions where tissue atrophy was assessed.

The hippocampus was delineated on the control ($N=40$) average template by investigators at the University College London (J.B.). The ROI tracing was performed using MIDAS (Medical Image Display and Analysis System) software (Freeborough et al., 1997). This delineation included the hippocampus proper, dentate gyrus, subiculum, and alveus (Fox et al., 1996; Scahill et al., 2003).

Correlations of structural brain differences (Jacobian Values) with clinical measurements and genetic variants

At each voxel, correlations were assessed, using the general linear model, between the Jacobian value and several clinical measures - the MMSE, and Clinical Dementia Rating summary scores (Morris, 1993). The CDR assesses a patient's cognitive and functional performance in six areas on a scale of 0 (no impairment) to 3 (impaired): memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care. As there is a significant range restriction with global CDR scores, we also assessed correlations with the CDR 'sum-of-boxes' scores, which have a greater dynamic range (0-18), and arguably provides more useful information than the CDR global score, especially in mild cases (Lynch et al., 2006). In the Jacobian maps, CSF regions typically show 'expansion' as AD progresses (for example, due to lateral ventricle enlargement), so we performed separate evaluations of the positive, negative and two-sided associations between the Jacobian and diagnostic group. The results of voxel-wise correlations were corrected for multiple comparisons by permutation testing. Clinical scores were randomly assigned to each subject and the number of voxels with significant correlations ($p \leq 0.01$) was recorded. After 10,000 permutations, a ratio was calculated describing the fraction of the null simulations in which a statistical effect (defined here in advance as the total supra-threshold volume) had occurred with similar or greater magnitude than the real effects. The primary threshold of 0.01 has been used in our past studies and is based on setting a moderately strong threshold at the voxel level (alternatively, FDR could be used); the total supra-threshold volume is often used to assess the magnitude of an anatomically distributed effect, giving in general higher statistical power but a lesser ability to spatially localize the signal than tests based on cluster extent or peak height (Frackowiak et al., 2003). This ratio served as an estimate of the overall significance of the correlations, corrected for multiple comparisons, as performed in many prior studies (Nichols and Holmes, 2002).

Results

3D maps of brain atrophy in MCI and AD

We first examined the level of brain atrophy using the method of averaging individual differences. The resulting statistical maps (Fig. 2) detected the known characteristic patterns of atrophy in AD, revealing profound tissue loss in the temporal lobes bilaterally, the hippocampus, thalamus, widening of the bodies of the lateral ventricles and expansion of the circular sulcus of the insula.

Permutation tests were conducted to assess the overall significance of the maps, corrected for multiple comparisons. The permutation tests confirmed that there were significant tissue changes in the MCI (two tailed: $P=0.04$; negative one tail: $P=0.02$, ROI: left temporal lobe) and AD (two tailed: $P=0.002$, ROI: whole brain) when compared to the normal group respectively, corrected for multiple comparisons.

Power to detect brain atrophy in MCI and AD

The cumulative distribution function (CDF) curves (Fig. 3) illustrate the power to detect significant brain atrophy in MCI and AD using the method of averaging individual differences. Eight different experiments are shown, comparing various sample sizes ($N=10, 20, 30,$ or 40 per group) and different linear registration schemes (9P vs. 12P). The null distribution is confirmed to be correct based on aligning the second group of normal individuals to the initial control template. In other words, although FDR methods assume that a null effect would have a CDF that is a diagonal line (see Fig. 3), we also confirmed that this is indeed the case using empirical data from two groups of controls. The blackline in Fig. 3 falls almost exactly on the diagonal, confirming that this TBM design controls for false positives at the appropriate rate for all thresholds (it is not exactly diagonal). If CDFs from many independent samples were

averaged, the population mean CDF should tend towards a diagonal line. As expected, regardless of the method used, there are more significant voxels (at any given threshold such as $p < 0.01$) detected in the AD versus Control comparison, relative to the MCI versus Control comparison. More interestingly, however, the cumulative p -value plots obtained when 9P linear registration is used (*solid lines*) are mostly situated above the ones from 12P linear registration (*dotted lines*), suggesting that the 9P registration scheme may have superior power for detecting atrophy in MCI and AD and differentiating these groups from normal subjects. As might be expected, sample size greatly influences the power to detect brain atrophy in MCI and AD, with effect sizes increasing monotonically with sample size.

Correlations with clinical measurements

Any quantitative measure of brain atrophy has greater value if it can be shown to correlate with established measures of cognitive or clinical decline, or with future outcome measures, such as imminent conversion to AD. We found strong correlations between the Jacobian values derived from the standard method ($N=120$) and the clinical measures (MMSE, CDR summary and sum-of-boxes scores; Table 1). This table reports corrected p -values for the correlations with voxel-level TBM values, rather than with a global summary value from TBM. To avoid reductions in power due to restricting the range to the AD or MCI groups separately, these correlations are reported for the entire sample of 120. As such, the normal subjects, who tend to score in the normal range on all the clinical measures, drive these associations to some extent.

Within the whole brain, the P values represent the overall significance level of correlations between the Jacobian maps and the clinical measures (corrected for multiple comparisons). The significance level is based on the number of suprathreshold voxels in the ROI, rather than their average or maximum. This method is sometimes known as set-level inference, which generally has greatest power (relative to other tests such as peak height or cluster size) for detecting a spatially distributed effect. Since there are two types of signals in the Jacobian maps: regional expansion (e.g., in the ventricles) and regional atrophy (e.g., in gray and white matter), positive and negative correlations are tested separately. Two-tailed tests detect any consistent structural differences without an emphasis on the sign of the changes (gain or loss).

Average brain changes within each regions of interest (ROI)

In voxel-based studies, such as TBM, there is interest in reducing detailed 3D maps to simpler numeric summaries that may be more convenient to use as outcome measures in a clinical trial, especially when a small number of outcome measures must be agreed in advance. To summarize group differences or other statistical effects detected by TBM in a lobe, hemisphere, or in a region of interest computed from an independent experiment, several different numeric summaries are possible, such as the number of suprathreshold voxels in an ROI, the maximum statistic within an ROI, or some weighted average of the Jacobian values within the ROI. For simplicity, we summarize the Jacobian values by averaging them within several ROIs traced on the control MDT. While not necessarily the optimal summary in terms of power, these results may at least be compared with the results of automated volumetric parcellation methods. This equivalence occurs because the average Jacobian in a region would be proportional to the overall volume of that region if it were labeled automatically by transferring atlas labels onto the individual using the deformation field.

We computed the spatial mean of the Jacobian within each ROI for every subject from the individual Jacobian maps ($N=120$; Fig. 4). In the white matter, and in frontal, parietal, and temporal gray matter, there is a consistent trend for tissue reduction: AD < MCI < Normal. The result from a T test (two-tailed, unequal variance) detects significant atrophy only in AD and only in the temporal lobe (marked with a * in Fig. 4). The occipital lobe, which is typically

one of the last areas to be affected by AD (Delacourte et al., 1999;Thompson et al., 2003), shows no tissue loss.

As a *post hoc* test, we investigated whether the use of a hippocampal region of interest would detect group differences better than using the whole temporal lobe (Fig. 5). This type of test is exploratory only, with the goal of finding the best region for averaging the Jacobian values, if a single numerical score is derived from TBM. The hippocampal ROI was delineated on the Control $N=40$ average template by investigators at the University College London (J.B.). In the AD group, atrophy was detected only in the left hippocampal ROI ($p=0.04$). There is a visually apparent trend for a left versus right asymmetry in the degree of atrophy, but it is not significant in either MCI or AD samples.

The outcome of these analyses suggests that using a hippocampal or temporal lobe ROI to summarize the effects in TBM maps may be inferior to using pFDR to quantify suprathreshold statistics within the same ROI. This is because the effects within each ROI are spatially heterogeneous, and numerical averages across spatial regions necessarily deplete the power of local tests by averaging all voxels equally. By contrast, pFDR can measure the quantity of non-null statistical events in an ROI, which may detect effects that are focused on a relatively small region of an ROI, or only partially overlapping with it.

Comparison with the method of directly aligning mean anatomical templates

Fig. 6 shows the mean level of volumetric atrophy, in AD and MCI, relative to controls, as a percentage, using the method of directly aligning group templates, which was advocated in (Rohlfing et al., 2005;Aljabar et al., 2006,2008). This method suggests that there is widespread atrophy in AD, in agreement with both the standard method and with visual inspection of the MDT templates. Atrophy of 20–30% was detected throughout the temporal lobe in AD, with moderate atrophy (10–20%) in the superior and middle frontal gyri, superior frontal sulcus, and corona radiata. The MCI pattern suggests atrophy of around 5% throughout the white matter, with deficits reaching 10–15% in the temporal lobes and hippocampus. A mean Jacobian was calculated within each ROI to show the computed overall volume differences for each lobe (Table 2). When compared to the normal group, the AD group shows the greatest volumetric deficit loss in the white matter, a reduction of 6.62%, and in temporal lobe gray matter, a volume deficit of 5.79%. In line with the literature, frontal and parietal gray matter show smaller proportional deficits, and tissue loss is not detected in the occipital lobes.

However, the direct alignment method has a serious limitation. When computing a group difference based on aligning group average images, there is no convenient way to conduct voxel-wise statistical tests to establish the significance of the observed differences (as noted in (Rohlfing et al., 2005)) since only one Jacobian map is derived to identify differences between the two group templates. In principle, a null distribution for the group-to-group deformation may be computed by permuting the assignment of subjects to groups, constructing mean anatomical templates for each permutation, and assessing the statistical distribution of deformation maps that would arise between these templates. As thousands of independent MDTs would be required to assemble this reference distribution, and each would require two rounds of nonlinear registration in groups of 40 subjects, this is computationally prohibitive (requiring around 80,000 CPU hours). If an omnibus probability (i.e., corrected for multiple comparisons) is determined by comparing the number of suprathreshold voxels in the true labeling to the permutation distribution, the number of permutations N must be chosen to control the standard error SE_p of omnibus probability p , which follows a binomial distribution $B(N, p)$ with $SE_p = \sqrt{p(1-p)/N}$ (Edgington, 1995). To adequately control the standard error of the resulting p -values derived from the permutation distribution, $N=8,000$ randomizations

are required to ensure that approximate margin of error (95% confidence interval) for p is around 5% of p , when 0.05 is chosen as the significance level.

As an approximation, we conducted voxel-wise two-sample t tests using the variance term obtained from the first approach as an estimate of the group variance between MCI/AD and control, subject to checking (below) that this did not inflate Type I error when truly null groups were compared. Using the estimated variance from the individual Jacobian maps, this alternative TBM design appeared to detect substantial atrophy in regions degenerating both early and late in AD. However, when applied to compare two different groups of normal subjects, the direct method did not control for false positives at the conventional rate, showing widespread “differences” even after multiple comparisons correction (Fig. 7; FDR q -value=0.0001). This problem occurs because the variance in the template-to-template registration is not simply related to the variance in the individual-to-template case; it depends on the geometry of the registration algorithm's cost function landscape with respect to the transformation parameters. One might expect the averaging of individual differences to be a slightly conservative approach as the variance in individual-to-template registrations is typically much higher than the variance in template-to-template registrations, as the cost function landscape is much smoother with respect to the alignment parameters when aligning two template images of very similar contrast and geometry. The registration error in individual registrations may be greater than that observed in template-to-template registrations, and this source of variance works against finding systematic group differences in volume, and may therefore underestimate the true reduction in volume in AD and MCI. This seems to be supported by the finding that the estimated volume differences for each tissue type in each lobe are around 50% greater for the TBM method based on directly aligning group averages, than for the TBM method based on averaging individual differences. A similar pattern was observed in a recent study (Chou et al., in press), in which anatomical labeling of the ventricles based on a single registration was more error-prone than combining multiple images to derive a segmentation, which led to better effect sizes in discriminating AD from controls (see (Twining et al., 2005), for related work). Even so, the lack of a computable null distribution for the direct method means that differences it detects cannot be regarded as statistically established. Using the variance of the individual mappings is not appropriate, as it leads to false positives.

A second argument may also be made that the direct method is inherently more prone to registration error and than the averaging of results from many registrations. Regardless of the algorithm used, both linear and nonlinear registration are imperfect and registration errors are not simply Gaussian at each voxel. When each subject is registered individually to a template, these errors are not likely to be compounded, as each subject has slightly different error maps that are likely to cancel out to some degree. However, when the non-linear averages are directly registered to each other, the registration errors will be compounded (as the same registration error is found in all subjects of the group after they have been aligned to the group template). This is likely to induce “spatial shifts” that may appear as (false) group differences.

Global differences

Finally some comment is necessary regarding the discounting of global anatomical differences in TBM. The maps reported here assessed residual anatomical differences after an initial 9-parameter global scaling of all AD, MCI, and control subjects to match an anatomical template. This scaling was performed in the automated registration step, and, in our cohort, the degrees of scaling (mean global expansion factors) for groups of controls, MCI and AD patients were 1.35 (SD = 0.14), 1.35 (0.14) and 1.32 (0.15) respectively, and there was no significant difference among the three groups (single factor ANOVA p -value = 0.62). As such, we did not adjust for group differences in overall brain scaling in our analyses, as no such differences were detected.

Single-subject analysis

Some comment is warranted regarding the possible value of TBM to assess of atrophy in individual subjects, which is closer to the problem faced in a clinical setting when evaluating disease burden. While we do not attempt a comprehensive analysis of this question here, Fig. 8 shows a map comparing brain structure in a single subject against a group mean. Relative to the mean template from the control subjects, this individual has 30% lower regional volumes throughout much of the white matter (*blue colors*), clear CSF space expansion in the Sylvian fissures (*red colors*) and in cortical regions, where sulcal spaces are enlarged. By comparing this with the standard deviation of the normal group, the significance map shows widespread regions with abnormally low tissue volumes (in the white matter) or abnormal expansions (in the perisylvian CSF). These effects are not focused in the cortex, suggesting that elastic registration has higher power to resolve white matter atrophy, perhaps because (1) registration is typically more accurate in the deep white matter than in the cortical gray matter, and (2) normal structural variation in subcortical regions is less than at the cortex, so abnormalities are easier to detect.

Discussion

This study had four main findings. First, a TBM method based on directly aligning group averaged images was found to be problematic, as it did not correctly control for false positives. This problem was solved by aligning each subject to a single template, and analyzing individual maps. Second, we showed a CDF-based method that can help to decide which methodological choices affect power in TBM; linear (9 parameter) initial registration and larger samples were found to give higher effect sizes, and the dependency on sample size was explored. Third, analysis of voxels in large regions such as the temporal lobe was more powerful than using small regions such as the hippocampus, confirming that TBM is better for resolving distributed atrophy rather than very small-scale changes, at least when used in a cross-sectional design. Fourthly, clinical measures of deterioration in brain function (MMSE, CDR scores) were tightly linked with both atrophy and ventricular expansion, but the atrophy measures gave higher effect sizes. The best TBM-based marker of neurodegeneration was temporal lobe atrophy, as this distinguished AD from controls better than other measures.

In our comparison of two types of TBM design, we first used the traditional method, which creates individual Jacobian maps for each subject by non-linearly aligning their MRIs to the normal MDT template. All the Jacobian maps share a common coordinate system defined by the normal MDT, so an average map of the group (normal, MCI or AD) was created by taking the arithmetic mean at each voxel (other possible approaches include using the geometric mean, matrix logarithm mean, Fréchet mean, or geodesic metrics on the deformation velocity (Woods, 2003; Avants and Gee, 2004; Leow et al., 2006; Aljabar et al., 2008; Lepore et al., 2008). Statistical parametric maps may then be computed to associate regional atrophy with predictors measured in each individual (diagnosis, clinical scores, etc.). By contrast, the direct method uses geometric centering to construct an average template that conforms to the group mean geometry, and then a single non-rigid transformation quantifies group differences. The two methods both detect tissue loss in temporal lobes, hippocampus, the thalamus and widespread widening of sulcal and ventricular CSF spaces, congruent with prior studies (Baron et al., 2001; Callen et al., 2001; Frisoni et al., 2002; Busatto et al., 2003; Gee et al., 2003; Thompson et al., 2003; Karas et al., 2004; Testa et al., 2004; Teipel et al., 2007; Whitwell et al., 2007).

The direct method has several limitations. First, it is difficult to covary for other variables measured at the individual level, such as age or sex, although this could be circumvented to some degree by matching samples for these variables. Second, it is computationally prohibitive to compute an empirical null distribution for deformations between group average templates, unless tens of thousands of templates are generated from permuted datasets. Null distributions

for Jacobian maps based on individual registrations are faster to compute, but do not adequately control for false positives when null experiments are performed (such as aligning two control MDTs with no true difference). Further study is necessary to clarify how registration errors compare when registering individuals and templates to other templates. In a recent study, Aljabar et al. (2008) computed maps of brain growth in 25 infants scanned one year apart, at one and two years of age, based on creating a mean template for baseline scans and directly aligning it to a mean template from follow-up scans. While they were not able to provide significance measures for the mapped changes, the overall growth factors for gray and white matter, computed from this direct registration, agreed with measures from independent segmentations, and the results were visually reasonable and in line with the neurodevelopmental literature. This suggests that the change rates observed with the direct method may be accurate, at least in a longitudinal study, but their significance is difficult to assess. If the direct method is used in a longitudinal study, it may be more robust than in a cross-sectional study, as the cohorts at each time point are by definition matched on all demographic variables other than time. In a cross-sectional study, any confounds in demographic matching of the groups may enter the maps of group differences, without a statistical means to adjust for them or estimate their effects.

Any TBM study is limited by the accuracy with which deformable registration can match anatomical boundaries between individual brains and corresponding regions on the template. Our mean deformation template (MDT) was created after rigorous nonlinear registration, and geometric centering. Several studies have suggested that registration bias can be reduced, and effect sizes increased, by using an unbiased group-average template of this kind (Kovacevic et al., 2005; Kochunov et al., 2002; Good et al., 2001; Lepore et al., 2007). Most anatomical features and boundaries are well-preserved in the MDT, and the hippocampus is sufficiently discernible to be labeled by hand on the MDT. Even so, it may not be possible to achieve accurate regional measurements of atrophy, especially in small regions such as the hippocampus, since that would assume a locally highly accurate registration. TBM is best for assessing differences with at a scale greater than 3–4 mm (the resolution of the FFT used to compute the deformation field). For smaller-scale effects, direct modeling of the structure, e.g. using surface-based geometrical methods, may offer additional statistical power to detect subregional differences (e.g., Morra et al., submitted for publication).

As the ADNI initiative is a study of 200 AD, 400 MCI, and 200 controls, this study focused not just on AD but also on MCI. The focus in the AD field has shifted to MCI in recent years, in the hope of tracking disease progression and ultimately resisting it, before individuals progress to AD. It is useful to know what factors affect detection power or link with cognition in MCI versus AD, as factors that can enhance power in MCI may not be so relevant in a study of AD, and regions in which atrophy correlates with cognition in MCI may not be so relevant to cognition in AD, or in healthy aging. In this study, we therefore included power estimates and measures of effect sizes for TBM studies of both MCI and AD, revealing that sample requirements differ greatly for different effects of interest.

In this study, we did not (beyond multiple pair-wise comparisons) attempt to gain any insight into the shift in morphological changes from normal controls to MCI to AD. A strength of a TBM analysis would be to map all subjects to a common template, and then track the distribution of atrophy it spreads anatomically over time (e.g. Thompson et al., 2001) or with clinical progression (Janke et al., 2001). As ADNI is a longitudinal study, we plan to fit longitudinal models to detect the shift in the location of greatest atrophy as the longitudinal data (e.g., 1 year follow-up scans) become available. This will require repeated-measures methods, which have not yet been validated for TBM, and specialized methods for creating longitudinal mean templates, which are emerging in the literature (see Lorenzen et al., 2004, 2006).

The ROI-based analyses (Figs. 4 and 5) revealed patterns of atrophy in MCI and AD, but with relatively low significance levels. In future, we will see if statistical power can be improved by adjusting for the effects of the CSF signals on the overall estimates of atrophy, as the effects of CSF expansion partially oppose the contraction signal. Due to potential biases, we avoided analyzing effects from the contracting voxels only (i.e., voxels with Jacobian less than one), such as taking the average Jacobian in the contracting regions, or counting the numbers of contracting voxels. Such an approach could be biased, in that a group with greater variance in the Jacobian could have more contracting voxels while having the same mean level of atrophy. Also, an analysis of contracting voxels could be biased towards a group with a very small region of very high atrophy, which could occur, at least in principle.

Use of CDF plots

In neuroscientific studies using TBM, it is vital to optimize statistical power for detecting anatomical differences, especially when evaluating the power of treatment to counteract degeneration, as in a drug trial, or in an epidemiological study to identify neuroprotective factors (Lopez et al., 2007). Comparison of power across image analysis methods is of great interest, but some caveats are necessary regarding the use of CDF-based approaches, in which the ordered p -values are plotted and compared to the expected 45-degree line under the null hypothesis of “no effect”. In highly sensitive methods, the departure of the early part of the curve from a 45-degree line will be large (showing a positive upswing). This assumption is supported by our plots (Fig. 3), in which successively larger sample sizes boost the effect size in statistical maps identifying group differences, for both MCI and AD. As shown in the CDF plots (Fig. 3), for all significance thresholds (values on the x axis), the proportion of significant voxels, detecting group differences, increases dramatically as the group size is enlarged from $N=10$ to 40. In prior work (Lepore et al., 2008), we used this same CDF approach to note that the deviation of the statistics from the null distribution generally increases with the number of parameters included in the statistics, with multivariate TBM statistics on the full tensor typically outperforming scalar summaries of the deformation based on the eigenvalues, trace, or the Jacobian determinant. With this approach, we also found that effect sizes in TBM may be boosted, at least in some contexts, by using mean anatomical templates based on Lie group averaging (Lepore et al., 2007) or by using deformation models based on information-theoretic Kullback-Leibler distances (Leow et al., 2007), or using Riemannian fluid models, which regularize the deformation in a log-Euclidean manifold (Brun et al., 2007).

Even so, we do not have ground truth regarding the extent and degree of atrophy or neurodegeneration in AD or MCI. So, although an approach that finds greater disease effect sizes is likely to be more accurate than one that fails to detect disease, it would be better to compare these models in a predictive design where ground truth regarding the dependent measure is known (i.e., morphometry predicting cognitive scores or future atrophic change; see e.g., (Grundman et al., 2002)). We are collecting this data at present, and any increase in power for a predictive model may allow a stronger statement regarding the relative power of different models in TBM, or the relative power of one image analysis method versus another for tracking brain disease.

A second caveat is that just because a CDF curve is higher for one method than another in one experiment, it may not be true of all experiments. Without confirmation on multiple samples, it may not reflect a reproducible difference between methods. FDR and its variants (Storey, 2002; Langers et al., 2007) declare that a CDF shows evidence of a signal if it rises greater than 20 times more sharply than a null distribution, so a related criterion could be developed to compare two empirical mean CDFs after multiple experiments. As simple numeric summaries sacrifice much of the power of maps, and provide a rather limited view of the

differences in sensitivity among voxel-based methods, additional work on CDF-based comparisons of methods seems warranted.

Correlations with clinical measures

The corrected P values signify the overall significance levels of the correlations between atrophy and clinical scores within the whole brain. For MMSE, both the positive and two-tailed tests are significant, suggesting a correlation between the regions of volume reduction and lower MMSE scores. For global CDR and sum-of-boxes, we obtain robust results in both negative and two-tailed correlations. As higher CDR scores denote greater impairment, the negative correlation links lower brain volume with greater CDR scores. Based on Table 1, *atrophy* of brain tissue (gray and white matter) detected by TBM links better with cognition than volume expansion (e.g., of the ventricles), although each is significantly associated with both MMSE and CDR. Strictly speaking, the CSF expansion signal may offer less signal to noise than the atrophic signal as we are using statistical tests that depend on the total volume of regions that reach a certain threshold (supra-threshold volume and corrected q -values from FDR). It may be that, if the statistical tests had been formulated differently, e.g., as strict voxel-level comparisons (e.g., maximal t -statistics), they would detect CSF differences with greater effect sizes than atrophic effects.

Analysis of group size

It may seem odd to assess effect size in groups as small as 10 to 40 subjects per group when imaging studies such as ADNI now assess 200 or 400 subjects per group. Here a sample as low as 10 is merely included to show how power completely breaks down when the sample is minimal and not sufficiently powered to detect an effect with reasonable confidence. Although morphometric studies of 10–20 subjects per group were more common in studies five to ten years ago (e.g. Thompson et al., 2001), most current MRI studies are designed to contrast patients in several categories (treatment versus placebo, MCI converters versus non-converters, ApoE4 carriers versus non-carriers), so it is common to have groups containing as few as 10 subjects for some statistical contrasts (given the low annual rate of conversion from MCI to AD, and the low incidence of certain risk genotypes). As seen with our CDF approach, for contrasts that are underpowered, it may have merit to plot the CDFs based on pilot samples, and assess the rate at which the CDFs are increasing (or not) with successive increments in the sample size. Although there is no widely accepted power analysis for morphometric studies using statistical maps as outcome measures, the CDF based methods, such as those advocated here, offer a means to study whether incrementing a small sample could yield sufficient power to reject a null hypothesis.

Single-subject analysis

Although these maps (Fig. 8) are clearly of interest, several caveats are needed in interpreting them. First, in this case all of the variance used to assess abnormality comes from a statistic comparing the single subject with the normal group, so some covariation for age, sex, and possibly other factors, ideally based on multiple regression in a large sample, would be more appropriate to calibrate the level of age-adjusted atrophy. Second, lower tissue volumes in an individual are not always a sign of disease, so plotting regional volumes as a percentile relative to a normative population (which is essentially what the significance map is) may reflect a combination of disease-related atrophy, and some natural variation in brain volumes. These factors could be easier to disentangle in a longitudinal evaluation of the same patient over time. Finally, as noted by Salmond et al. (2002), if a Gaussian distribution is assumed for the Jacobian statistics at each voxel, a significant number of false positives may still arise purely due to non-Gaussianity when comparing a single subject to a group. To ensure that the data are smooth enough for the residuals to be regarded as normally distributed, Salmond et al. suggested that

the data be first heavily smoothed (using a 12mm FWHM kernel); alternatively, a large control population could be used to establish a non-parametric reference distribution at each voxel, which is essentially the permutation approach taken here.

Anatomical maps and prior work

The main contribution of this paper, relative to prior work using voxel-based morphometry (VBM) and tensor-based morphometry in AD or MCI, is to study the effects of different analysis choices within the framework of TBM, and how they affect the sensitivity for detecting disease effects. Our anatomical findings are largely in line with prior work using automated techniques to map patterns of brain atrophy at voxel-level. Initial formulations of VBM derived maps of structural differences by comparing the local composition of brain tissue types after global position and volumetric differences had been removed through spatial normalization (Ishii et al., 2005; Shiino et al., 2006; Davatzikos et al., 2008; Fan et al., 2008; Karas et al., 2007; Smith et al., 2007; Vemuri et al., 2008). In contrast, TBM is a method based on high-dimensional image registration, which derives information on regional volumetric differences from the deformation field that aligns the images. Recent reformulations of VBM, termed 'optimized VBM' (Davatzikos et al., 2001; Good et al., 2001) modulate the voxel intensity of the spatially normalized gray matter maps by the local expansion factor of a 3D deformation field that aligns each brain to a standard brain template. As a result, the final modulated voxel contains the same amount of gray matter as in the native pre-registered gray matter map. Chetelat et al. (2002) and Karas et al. (2004) used VBM to analyze patterns of gray matter loss in MCI and AD. Relative to normal subjects, Chetelat et al., (2002) found that MCI subjects showed significant atrophy in the hippocampus, temporal cortices, and cingulate gyri. Gray matter density in the posterior association cortex was significantly higher in MCI than AD. Karas et al. (2004) found similar patterns of parietal atrophy in AD and MCI, but found active hippocampal atrophy in the transitional stage from MCI to AD. The author suggested this discrepancy could be due to borderline significance or difference in disease severity of MCI populations. A very recent study by Teipel et al. (Teipel et al., 2007) used the TBM method to study brain degeneration in MCI and AD. They used principal component analysis to extract spatially distributed anatomical features associated with the diagnosis of AD, and they focused on identifying features that may be useful in predicting the transition from MCI to AD. Future longitudinal TBM studies with the ADNI data are likely to reveal which aspects of atrophy are most predictive of future conversion to AD, and which voxel-based methods are optimal for detecting progression or correlations with cognition. As the sample size increases, it may be possible to detect and model effects of the MRI platform, field strength, or acquisition site, to determine whether the multi-site and dual MRI platform acquisition of the data contributed to reduced effect sizes, especially for the MCI group. Comparisons distinguishing MCI from controls may be more sensitive to these effects, whereas the AD versus control group comparison has an effect size so great that it overwhelms any increased variability due to multicenter acquisition. This potential source of variability, that is perhaps not typical of studies in general, will be evaluated in future.

Acknowledgments

Data used in preparing this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database (www.loni.ucla.edu/ADNI). Many ADNI investigators therefore contributed to the design and implementation of ADNI or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators is available at www.loni.ucla.edu/ADNI/Collaboration/ADNI_Citation.shtml. This work was primarily funded by the ADNI (Principal Investigator: Michael Weiner; NIH grant number U01 AG024904). ADNI is funded by the National Institute of Aging, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and the Foundation for the National Institutes of Health, through generous contributions from the following companies and organizations: Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, Glaxo-SmithKline, Merck & Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, the Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories, and the Institute for the Study of Aging (ISOA),

with participation from the U.S. Food and Drug Administration. 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. Algorithm development for this study was also funded by the NIA, NIBIB, the National Library of Medicine, and the National Center for Research Resources (AG016570, EB01651, LM05639, RR019771 to PT). Author contributions were as follows: XH, AL, SL, AK, AT, NL, YC, MC, MB, RB, JB, NS, LB, and PT performed the image analyses; CJ, AD, MAB, PB, JG, CW, JW, BB, AF, NF, DH, JK, CS, GA, and MW contributed substantially to the image acquisition, study design, quality control, calibration and pre-processing, databasing and image analysis. We thank Anders Dale for his contributions to the image pre-processing and the ADNI project. Part of this work was undertaken at UCLH/UCL, which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme.

References

- Aljabar P, Bhatia KK, Hajnal JV, Boardman JP, Srinivasan L, Rutherford MA, Dyet LE, Edwards AD, Rueckert D. Analysis of Growth in the Developing Brain Using Non-Rigid Registration. *IEEE Int. Symp. Biomed. Imaging* 2006:201–204.
- Aljabar P, Bhatia KK, Murgasova M, Hajnal JV, Boardman JP, Srinivasan L, Rutherford MA, Dyet LE, Edwards AD, Rueckert D. Assessment of brain growth in early childhood using deformation-based morphometry. *Neuroimage* 2008;39:348–358. [PubMed: 17919930]
- Apostolova LG, Thompson PM. Brain mapping as a tool to study neurodegeneration. *Neurotherapeutics* 2007;4:387–400. [PubMed: 17599704]
- Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, Thompson PM. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Arch. Neurol* 2006;63:693–699. [PubMed: 16682538]
- Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage* 2007;38:95–113. [PubMed: 17761438]
- Ashburner, J.; Friston, KJ. Morphometry.. In: Ashburner, J.; Friston, KJ.; Penny, W., editors. *Human Brain Function*. Academic Press; 2003.
- Avants B, Gee JC. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage* 2004;23(Suppl 1):S139–S150. [PubMed: 15501083]
- Baron JC, Chetelat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage* 2001;14:298–309. [PubMed: 11467904]
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., B* 1995;57(1):289–300.
- Brun, C.; Lepore, N.; Pennec, X.; Chou, Y.; Lopez, O.; Aizenstein, H.; Becker, J.; Toga, A.; Thompson, P. Proc. of MICCAI'07 Workshop on Statistical Registration: Pair-wise and Group-wise Alignment and Atlas Formation. Springer; Berlin, Brisbane, Australia: 2007. Comparison of Standard and Riemannian Elasticity for Tensor-Based Morphometry in HIV/AIDS.
- Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Trans. Med. Imaging* 1999;18(1):32–42. [PubMed: 10193695]
- Busatto GF, Garrido GE, Almeida OP, Castro CC, Camargo CH, Cid CG, Buchpiguel CA, Furuie S, Bottino CM. A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease. *Neurobiol. Aging* 2003;24:221–231. [PubMed: 12498956]
- Callen DJ, Black SE, Gao F, Caldwell CB, Szalai JP. Beyond the hippocampus: MRI volumetry confirms widespread limbic atrophy in AD. *Neurology* 2001;57:1669–1674. [PubMed: 11706109]
- Cardenas VA, Studholme C, Gazdzinski S, Durazzo TC, Meyerhoff DJ. Deformation-based morphometry of brain changes in alcohol dependence and abstinence. *Neuroimage* 2007;34:879–887. [PubMed: 17127079]
- Carmichael OT, Thompson PM, Dutton RA, Lu A, Lee SE, Lee JY, Kuller LH, Lopez OL, Aizenstein HJ, Meltzer CC, Liu Y, Toga AW, Becker JT. Mapping ventricular changes related to dementia and mild cognitive impairment in a large community-based cohort. *IEEE ISBI* 2006:315–318.
- Carmichael OT, Kuller LH, Lopez OL, Thompson PM, Dutton RA, Lu A, Lee SE, Lee JY, Aizenstein HJ, Meltzer CC, Liu Y, Toga AW, Becker JT. Ventricular volume and dementia progression in the Cardiovascular Health Study. *Neurobiol. Aging* 2007;28:389–397. [PubMed: 16504345]

- Chetelat G, Desgranges B, De La Sayette V, Viader F, Eustache F, Baron JC. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. *Neuroreport* 2002;13:1939–1943. [PubMed: 12395096]
- Chiang MC, Dutton RA, Hayashi KM, Lopez OL, Aizenstein HJ, Toga AW, Becker JT, Thompson PM. 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. *Neuroimage* 2007a;34:44–60. [PubMed: 17035049]
- Chiang MC, Reiss AL, Lee AD, Bellugi U, Galaburda AM, Korenberg JR, Mills DL, Toga AW, Thompson PM. 3D pattern of brain abnormalities in Williams syndrome visualized using tensor-based morphometry. *Neuroimage* 2007b;36:1096–1109. [PubMed: 17512756]
- Chou YY, Lepore N, de Zubicaray GI, Carmichael OT, Becker JT, Toga AW, Thompson PM. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. *Neuroimage*. in press
- Christensen GE, Johnson HJ, Vannier MW. Synthesizing average 3D anatomical shapes. *Neuroimage* 2006;32:146–158. [PubMed: 16697223]
- Chung MK, Worsley KJ, Paus T, Cherif C, Collins DL, Giedd JN, Rapoport JL, Evans AC. A unified statistical approach to deformation-based morphometry. *Neuroimage* 2001;14:595–606. [PubMed: 11506533]
- Cockrell JR, Folstein MF. Mini-Mental State Examination (MMSE). *Psychopharmacol. Bull* 1988;24:689–692. [PubMed: 3249771]
- Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr* 1994;18:192–205. [PubMed: 8126267]
- Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *Neuroimage* 2001;14:1361–1369. [PubMed: 11707092]
- Davatzikos C, Barzi A, Lawrie T, Hoon AH Jr, Melhem ER. Correlation of corpus callosal morphometry with cognitive and motor function in periventricular leukomalacia. *Neuropediatrics* 2003;34:247–252. [PubMed: 14598230]
- Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol. Aging* 2008;29:514–523. [PubMed: 17174012]
- Delacourte A, David JP, Sergeant N, Buee L, Wattez A, Vermersch P, Ghazali F, Fallet-Bianco C, Pasquier F, Lebert F, Petit H, Di Menza C. The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer's disease. *Neurology* 1999;52:1158–1165. [PubMed: 10214737]
- Dubb A, Xie Z, Gur R, Gee J. Characterization of brain plasticity in schizophrenia using template deformation. *Acad. Radiol* 2005;12:3–9. [PubMed: 15691720]
- Edgington, ES. Randomization tests. 3rd. Marcel Dekker; New York: 1995.
- Fan Y, Batmanghelich N, Clark CM, Davatzikos C. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* 2008;39(4):1731–1743. [PubMed: 18053747]
- Fillard, P.; Arsigny, V.; Pennec, X.; Thompson, P.; Ayache, N. Information Processing in Medical Imaging (IPMI). Springer; Berlin: 2005. Extrapolation of Sparse Tensor Fields: Application to the Modeling of Brain Variability; p. 644–652. Glenwood Springs, Colorado
- Fleisher A, Fennema-Notestine C, Hagler D, Podraza K, Wu E, Taylor C, Karow D, Dale A. Baseline structural MRI correlates of clinical measures in the Alzheimer's disease neuroimaging initiative. *Alzheimer's & Dementia: Journal of the Alzheimer's Association* July 2007 2007;3(3):S107–S108.
- Folstein MF, Folstein SE, McHugh PR. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res* 1975;12:189–198. [PubMed: 1202204]
- Fox NC, Warrington EK, Freeborough PA, Hartikainen P, Kennedy AM, Stevens JM, Rossor MN. Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. *Brain* 1996;119(Pt 6):2001–2007. [PubMed: 9010004]
- Fox NC, Crum WR, Scahill RI, Stevens JM, Janssen JC, Rossor MN. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet* 2001;358:201–205. [PubMed: 11476837]

- Frackowiak, RSJ.; Friston, KJ.; Frith, C.; Dolan, R.; Friston, KJ.; C, JP.; Zeki, S.; Ashburner, J.; Penny, WD., editors. *Human Brain Function*. 2nd. Academic Press; USA: 2003.
- Freeborough PA, Fox NC. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *J. Comput. Assist. Tomogr* 1998;22:838–843. [PubMed: 9754126]
- Freeborough PA, Fox NC, Kitney RI. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Comput. Methods Programs Biomed* 1997;53:15–25. [PubMed: 9113464]
- Frisoni GB, Testa C, Zorzan A, Sabattoli F, Beltramello A, Soininen H, Laakso MP. Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. *J. Neurol. Neurosurg. Psychiatry* 2002;73:657–664. [PubMed: 12438466]
- Gee J, Ding L, Xie Z, Lin M, DeVita C, Grossman M. Alzheimer's disease and frontotemporal dementia exhibit distinct atrophy-behavior correlates. *Acad. Radiol* 2003;10:1392–1401. [PubMed: 14697007]
- Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 2002;15:870–878. [PubMed: 11906227]
- Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 2001;14:21–36. [PubMed: 11525331]
- Grundman M, Sencakova D, Jack CR Jr. Petersen RC, Kim HT, Schultz A, Weiner MF, DeCarli C, DeKosky ST, van Dyck C, Thomas RG, Thal LJ. Brain MRI hippocampal volume and prediction of clinical status in a mild cognitive impairment trial. *J. Mol. Neurosci* 2002;19:23–27. [PubMed: 12212787]
- Gunter, J.; Bernstein, M.; Borowski, B.; Felmlee, J.; Blezek, D.; Mallozzi, R. Validation testing of the MRI calibration phantom for the Alzheimer's Disease Neuroimaging Initiative Study; ISMRM 14th Scientific Meeting and Exhibition; 2006.
- Hua X, Leow AD, Levitt JG, Caplan R, Thompson PM, Toga AW. Detecting brain growth patterns in normal children using tensor-based morphometry. *Hum. Brain. Mapp.* in press
- Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br. J. Psychiatry* 1982;140:566–572. [PubMed: 7104545]
- Ishii K, Kawachi T, Sasaki H, Kono AK, Fukuda T, Kojima Y, Mori E. Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of z score images. *AJNR Am. J. Neuroradiol* 2005;26:333–340. [PubMed: 15709131]
- Jack CR Jr. Shiung MM, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, Knopman DS, Smith GE, Ivnik RJ, Tangalos EG, Petersen RC. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic MCI. *Neurology* 2005;65:1227–1231. [PubMed: 16247049]
- Jack CR Jr. Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DL, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, Decarli CS, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging.* in press
- Janke AL, de Zubicaray G, Rose SE, Griffin M, Chalk JB, Galloway GJ. 4D deformation modeling of cortical disease progression in Alzheimer's dementia. *Magn. Reson. Med* 2001;46:661–666. [PubMed: 11590641]
- Joshi S, Davis B, Jomier M, Gerig G. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage* 2004;23(Suppl 1):S151–S160. [PubMed: 15501084]
- Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R, Kennedy D, Schmitt F, Brown G, Macfall J, Fischl B, Dale A. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 2006;30:436–443. [PubMed: 16300968]
- Karas GB, Scheltens P, Rombouts SA, Visser PJ, van Schijndel RA, Fox NC, Barkhof F. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 2004;23:708–716. [PubMed: 15488420]
- Karas G, Scheltens P, Rombouts S, van Schijndel R, Klein M, Jones B, van der Flier W, Vrenken H, Barkhof F. Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study. *Neuroradiology* 2007;49:967–976. [PubMed: 17955233]

- Kochunov P, Lancaster JL, Thompson P, Woods R, Mazziotta J, Hardies J, Fox P. Regional spatial normalization: toward an optimal target. *J. Comput. Assist. Tomogr* 2001;25:805–816. [PubMed: 11584245]
- Kochunov P, Lancaster J, Thompson P, Toga AW, Brewer P, Hardies J, Fox P. An optimized individual target brain in the Talairach coordinate system. *Neuroimage* 2002;17:922–927. [PubMed: 12377166]
- Kochunov P, Lancaster J, Hardies J, Thompson PM, Woods RP, Cody JD, Hale DE, Laird A, Fox PT. Mapping structural differences of the corpus callosum in individuals with 18q deletions using targetless regional spatial normalization. *Hum. Brain Mapp* 2005;24:325–331. [PubMed: 15704090]
- Kovacevic N, Henderson JT, Chan E, Lifshitz N, Bishop J, Evans AC, Henkelman RM, Chen XJ. A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability. *Cereb. Cortex* 2005;15:639–645. [PubMed: 15342433]
- Langers DR, Jansen JF, Backes WH. Enhanced signal detection in neuroimaging by means of regional control of the global false discovery rate. *Neuroimage* 2007;38:43–56. [PubMed: 17825583]
- Lee AD, Leow AD, Lu A, Reiss AL, Hall S, Chiang MC, Toga AW, Thompson PM. 3D pattern of brain abnormalities in Fragile X syndrome visualized using tensor-based morphometry. *Neuroimage* 2007;34:924–938. [PubMed: 17161622]
- Leow, A.; Huang, SC.; Geng, A.; Becker, JT.; Davis, S.; Toga, AW.; Thompson, PM. Information Processing in Medical Imaging. Springer Berlin, Glenwood Springs; Colorado, USA: 2005a. Inverse Consistent Mapping in 3D Deformable Image Registration: Its Construction and Statistical Properties; p. 493-503.
- Leow, AD.; Thompson, PM.; Hayashi, KM.; Bearden, C.; Nicoletti, MA.; Monkul, SE.; Brambilla, P.; Sassi, RB.; Mallinger, AG.; Soares, JC. Society for Neuroscience. Washington, DC: 2005b. Lithium Effects on Human Brain Structure Mapped Using Longitudinal MRI.
- Leow AD, Klunder AD, Jack CR Jr, Toga AW, Dale AM, Bernstein MA, Britson PJ, Gunter JL, Ward CP, Whitwell JL, Borowski BJ, Fleisher AS, Fox NC, Harvey D, Kornak J, Schuff N, Studholme C, Alexander GE, Weiner MW, Thompson PM. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *Neuroimage* 2006;31:627–640. [PubMed: 16480900]
- Leow A, Yanovsky I, Chiang M, Lee A, Klunder A, Lu A, Becker J, Davis S, Toga A, Thompson P. Statistical properties of Jacobian maps and inverse-consistent deformations in non-linear image registration. *IEEE Trans. Med. Imaging* 2007:822–832. [PubMed: 17679333]
- Lepore, N.; Brun, C.; Pennec, X.; Chou, Y.; Lopez, O.; Aizenstein, H.; Becker, J.; Toga, A.; Thompson, P. MICCAI, Springer; Berlin, Brisbane, Australia: 2007. Mean Template for Tensor-Based Morphometry using Deformation Tensors.
- Lepore N, Brun C, Chou YY, Chiang MC, Dutton RA, Hayashi KM, Luders E, Lopez OL, Aizenstein HJ, Toga AW, Becker JT, Thompson PM. Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE Trans. Med. Imag* 2008;27:129–141.
- Lopez OL, Kuller LH, Becker JT, Dulberg C, Sweet RA, Gach HM, Dekosky ST. Incidence of dementia in mild cognitive impairment in the cardiovascular health study cognition study. *Arch. Neurol* 2007;64:416–420. [PubMed: 17353386]
- Lorenzen P, et al. Multi-class posterior atlas formation via unbiased Kullback-Leibler template estimation. *MICCAI* 2004:95–102.
- Lorenzen P, Prastawa M, Davis B, Gerig G, Bullitt E, Joshi S. Multi-modal image set registration and atlas formation. *Med. Image Anal* 2006;10:440–451. [PubMed: 15919231]
- Lynch CA, Walsh C, Blanco A, Moran M, Coen RF, Walsh JB, Lawlor BA. The clinical dementia rating sum of box score in mild dementia. *Dement. Geriatr. Cogn. Disord* 2006;21:40–43. [PubMed: 16254429]
- Marsden, J.; Hughes, T. Mathematical foundations of elasticity. Prentice-Hall; 1983.
- Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond., B Biol. Sci* 2001;356:1293–1322. [PubMed: 11545704]

- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939–944. [PubMed: 6610841]
- Morra J, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR, Schuff N, Weiner MW, Thompson PM. Automated 3D Mapping of Hippocampal Atrophy and its Clinical Correlates in 400 Subjects with Alzheimer's Disease, Mild Cognitive Impairment, and Elderly Controls, ISBI 2008. submitted for publication
- Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;43:2412–2414. [PubMed: 8232972]
- Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, Trojanowski JQ, Toga AW, Beckett L. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin. North Am* 2005a;15:869–877.xi-xii
- Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers. Dement* 2005b;1:55–66. [PubMed: 17476317]
- Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp* 2002;15:1–25. [PubMed: 11747097]
- Petersen RC. Aging, mild cognitive impairment, and Alzheimer's disease. *Neurol. Clin* 2000;18:789–806. [PubMed: 11072261]
- Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol* 1999;56:303–308. [PubMed: 10190820]
- Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, Ritchie K, Rossor M, Thal L, Winblad B. Current concepts in mild cognitive impairment. *Arch. Neurol* 2001;58:1985–1992. [PubMed: 11735772]
- Riddle WR, Li R, Fitzpatrick JM, DonLevy SC, Dawant BM, Price RR. Characterizing changes in MR images with color-coded Jacobians. *Magn. Reson. Imaging* 2004;22:769–777. [PubMed: 15234445]
- Rohlfing T, Pfefferbaum A, Sullivan EV, Maurer CR. Information fusion in biomedical image analysis: combination of data vs. combination of interpretations. *Inf. Process. Med. Imaging* 2005;19:150–161. [PubMed: 17354692]
- Salmond CH, Ashburner J, Vargha-Khadem F, Connelly A, Gadian DG, Friston KJ. Distributional assumptions in voxel-based morphometry. *Neuroimage* 2002;17:1027–1030. [PubMed: 12377176]
- Scahill RI, Frost C, Jenkins R, Whitwell JL, Rossor MN, Fox NC. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Arch. Neurol* 2003;60:989–994. [PubMed: 12873856]
- Shattuck DW, Leahy RM. BrainSuite: an automated cortical surface identification tool. *Med. Image Anal* 2002;6:129–142. [PubMed: 12045000]
- Shen D, Davatzikos C. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. *Neuroimage* 2003;18:28–41. [PubMed: 12507441]
- Shiino A, Watanabe T, Maeda K, Kotani E, Akguchi I, Matsuda M. Four subgroups of Alzheimer's disease based on patterns of atrophy using VBM and a unique pattern for early onset disease. *Neuroimage* 2006;33:17–26. [PubMed: 16904912]
- Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 1998;17:87–97. [PubMed: 9617910]
- Smith CD, Chebrolu H, Wekstein DR, Schmitt FA, Jicha GA, Cooper G, Markesbery WR. Brain structural alterations before mild cognitive impairment. *Neurology* 2007;68:1268–1273. [PubMed: 17438217]
- Storey JD. A direct approach to false discovery rates. *J. R. Statist. Soc. B* 2002;64(Pt 3):479–498.
- Studholme C, Cardenas V. A template free approach to volumetric spatial normalization of brain anatomy. *Pattern Recogn. Lett* 2004;25:1191–1202.
- Studholme C, Cardenas V, Blumenfeld R, Schuff N, Rosen HJ, Miller B, Weiner M. Deformation tensor morphometry of semantic dementia with quantitative validation. *Neuroimage* 2004;21:1387–1398. [PubMed: 15050564]

- Teipel SJ, Born C, Ewers M, Bokde AL, Reiser MF, Moller HJ, Hampel H. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *Neuroimage* 2007;38:13–24. [PubMed: 17827035]
- Testa C, Laakso MP, Sabattoli F, Rossi R, Beltramello A, Soininen H, Frisoni GB. A comparison between the accuracy of voxel-based morphometry and hippocampal volumetry in Alzheimer's disease. *J. Magn. Reson. Imaging* 2004;19:274–282. [PubMed: 14994294]
- Thompson P, Apostolova L. Computational anatomical methods as applied to aging and dementia. *Br. J. Radiol. Dec.*;2007 in press Invited Article
- Thompson PM, Giedd JN, Woods RP, MacDonald D, Evans AC, Toga AW. Growth patterns in the developing brain detected by using continuum mechanical tensor maps. *Nature* 2000a;404:190–193. [PubMed: 10724172]
- Thompson PM, Woods RP, Mega MS, Toga AW. Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. *Hum. Brain Mapp* 2000b;9:81–92. [PubMed: 10680765]
- Thompson PM, Vidal C, Giedd JN, Gochman P, Blumenthal J, Nicolson R, Toga AW, Rapoport JL. Mapping adolescent brain change reveals dynamic wave of accelerated gray matter loss in very early-onset schizophrenia. *Proc. Natl. Acad. Sci. U. S. A* 2001;98:650–11655.
- Thompson PM, Hayashi KM, de Zubicaray G, Janke AL, Rose SE, Semple J, Herman D, Hong MS, Dittmer SS, Doodrell DM, Toga AW. Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci* 2003;23:994–1005. [PubMed: 12574429]
- Toga, AW. *Brain Warping*. 1st. Academic Press; San Diego: 1999.
- Twining, C.; Cootes, T.; Marsland, S.; Petrovic, V.; Schestowitz, R.; Taylor, C. *A Unified Information-Theoretic Approach to Groupwise Non-rigid Registration and Model Building; Information Processing in Medical Imaging: 19th International Conference, IPMI; Springer, Berlin. Glenwood Springs, CO.; 2005. p. 1-14.*
- Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage* 2008;39:1186–1197. [PubMed: 18054253]
- Wechsler, D. *Wechsler Memory Scale*. Psychological Corp/Harcourt Brace Jovanovich; New York: 1987.
- Whitwell JL, Przybelski SA, Weigand SD, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr. 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain* 2007;130:1777–1786. [PubMed: 17533169]
- Woods RP. Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation. *Neuroimage* 2003;18:769–788. [PubMed: 12667854]

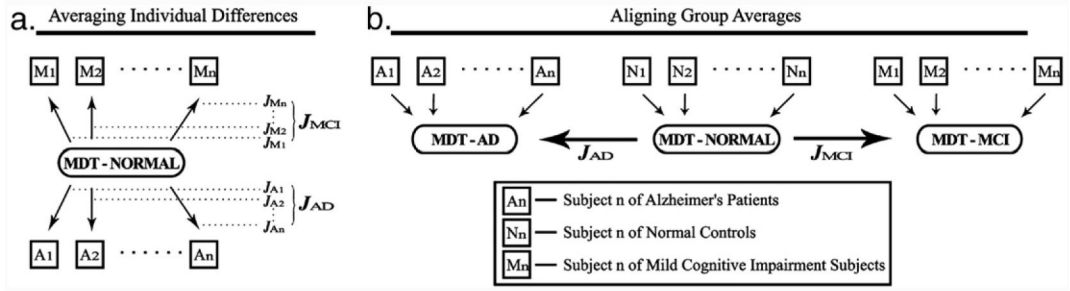


Fig. 1.

Two Alternative TBM Designs. Inter-group differences in brain structure may be assessed with TBM, using two alternative designs, which differ in terms of which images are registered to each other. The first approach, termed “Averaging Individual Differences”, (a), a mean template is created for the control subjects. Then, every image in the study is nonlinearly aligned to the control average template and the set of individual differences between each subject and the template is analyzed statistically. In panel a, the mean template was built from controls only; arguably, it could instead be re-built each time based on all subjects relevant to a specific hypothesis (e.g., Controls and MCI only, for an MCI-control comparison), but this would mean that the results of different contrasts would not be spatially registered with each other. The second approach, termed “Aligning Group Averages”, reflects a relatively new concept in TBM design (Rohlfing et al., 2005; Aljabar et al., 2006, 2008), (b), and creates minimal deformation targets (MDTs) for each diagnostic group separately using nonlinear registrations of subjects within the group (A1, A2, ...; N1, N2, ...; etc.) to create a template reflecting the group's mean geometry. Systematic anatomical differences between groups may then be assessed using direct alignment of these group-specific templates. In panel b, J_{AD} and J_{MCI} denote the Jacobians of these mappings, which contain information on the level of atrophy in the MCI and AD groups versus controls. We compare these two methods at the end of the Results section; for all statistical tests, the standard method of “averaging individual differences” is used.

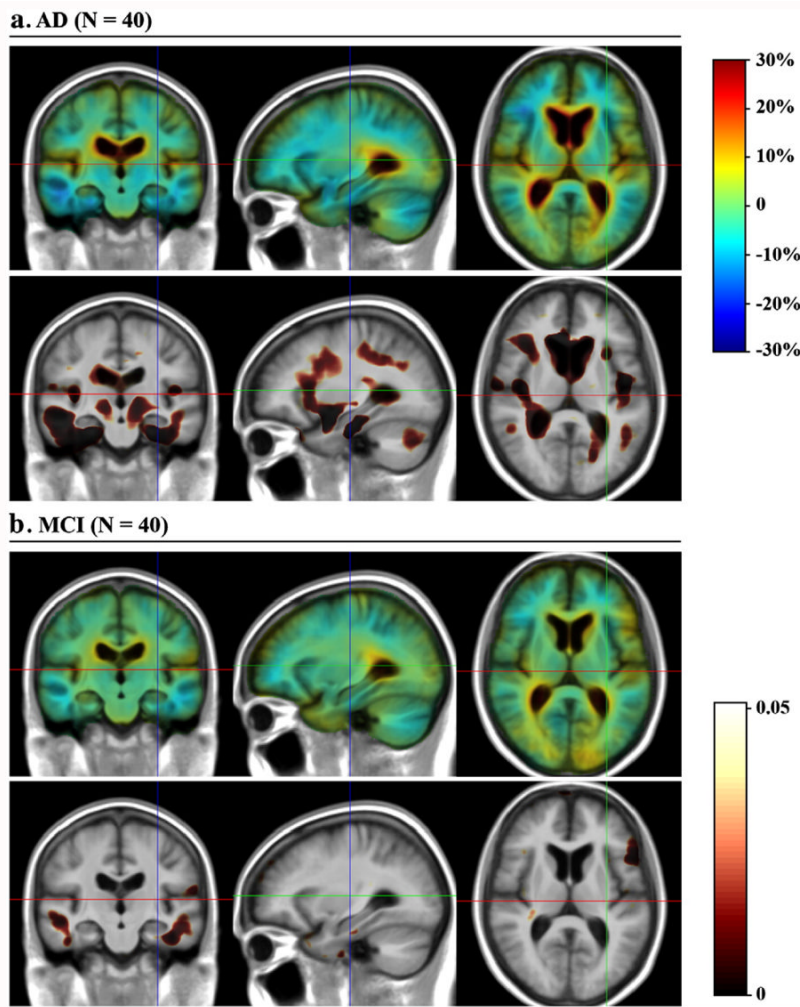


Fig. 2. *3D Maps of Brain Atrophy* (based on the method of averaging individual differences). The top rows of panel a and b show the level of atrophy in 40 AD patients and 40 MCI subjects as a percentage reduction in volume relative to controls, respectively. The bottom rows show the significance of these reductions, revealing highly significant atrophy in AD but a more anatomically restricted pattern of atrophy in MCI. The high Jacobian values immediately adjacent to the ventricle result from the limited spatial resolution of the deformation fields, which are computed via a Fourier transform on a 32^3 grid. The ventricles expand presumably due to cell or myelin loss in broad areas overlying the ventricular surface; not in the narrow band of tissue immediately adjacent to the ventricular surface, in which voxels are partial volumed with voxels where expansion is detected. This is also the case with PET and fMRI imaging, where imaging signals ‘bleed’ into the CSF space where the signal differs (and typically there is no CSF signal). As with those modalities, sharp boundaries are not found in group averaged deformation maps, so the image resolution should be interpreted with this in mind. Higher resolution morphometry is possible in longitudinal studies, where it is feasible (and makes sense) to perform image registration at a finer spatial scale.

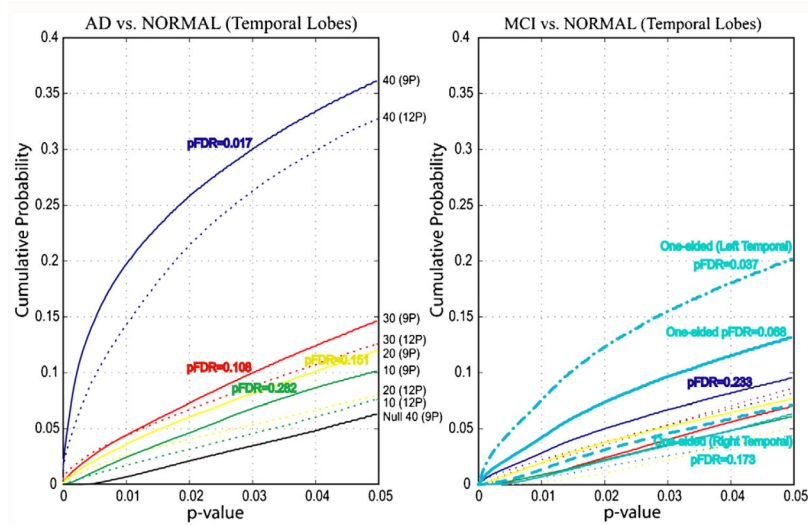


Fig. 3. Comparing Effect Sizes with CDF Plots: Influences of Sample Size and Registration Model (9 vs. 12 parameter linear registration) on the Statistical Effect size in Discriminating MCI and AD from Normal Subjects, using the TBM design of averaging individual differences. With some caveats (see Discussion), higher CDF curves denote greater effect sizes, i.e., a large number of significant voxels detected within the temporal lobes. Three aspects are notable: (1) Atrophy is detected with a greater effect size in the AD group compared to the MCI group, regardless of the method; (2) in AD, effect sizes are greater when larger samples are used (e.g., 40 per group versus 30, 20 or 10); (3) Using 9-parameter versus 12-parameter initial registration may provide slight gains in power, but these are no greater than the improvements gained by adding 10 subjects to the sample. The omnibus (corrected) significance for each experiment is the q -value obtained from the positive FDR method. 40 subjects per group are needed to detect atrophy significantly in AD and at trend level in MCI, using the entire temporal lobe as an ROI. Atrophy is detected in MCI with 40 subjects per group, but only in the left temporal lobe. Perhaps surprisingly, the jump in effect size as one goes from a sample size of 30 to 40 is much greater for normal vs AD than normal vs MCI. This may be because the true effect size in AD is greater, but it may also be because CDFs rely on thresholdings of the statistical maps, leading to many voxels reaching significance within a small range of sample sizes.

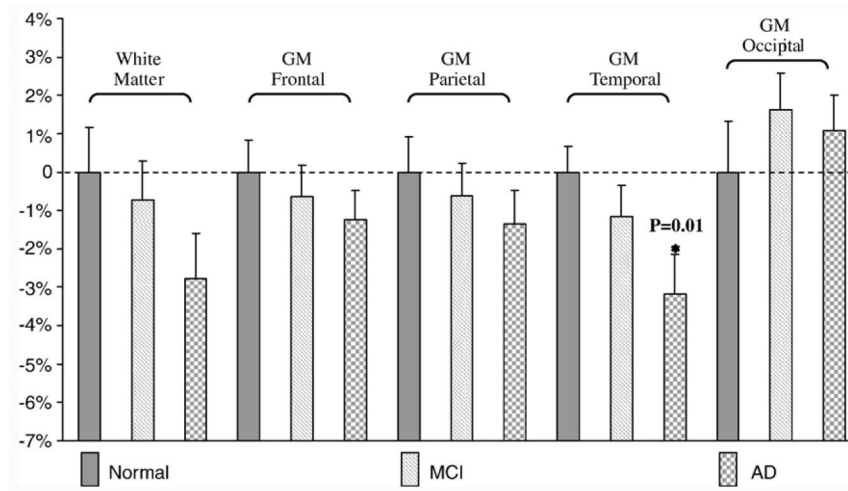


Fig. 4. Mean Jacobian of Normal, MCI and AD in each lobar ROI ($N=40$ for each group, 9P linear registration). This plot shows TBM-based volume estimates (relative to the normal subjects) for each tissue type in each lobe. To ease comparison of volumes across lobes, values are expressed as a proportion of the control average volume. The * indicates that there is significant atrophy relative to normal subjects.

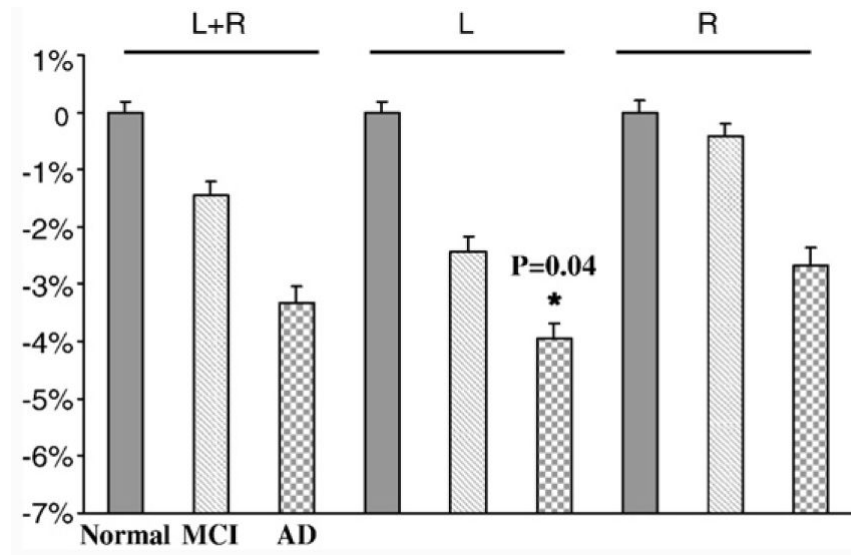


Fig. 5. Mean Jacobian Values within the Hippocampus, for Normal, MCI and AD Groups, using a Manually-Defined Traces of the Hippocampal Formation Delineated on the Control Average Template ($N=40$ for each group, 9P linear registration). This plot shows TBM-based volume estimates (relative to the normal subjects) for each tissue type in ROIs of the left and right hippocampi. The * indicates that there is significant atrophy relative to normal subjects.

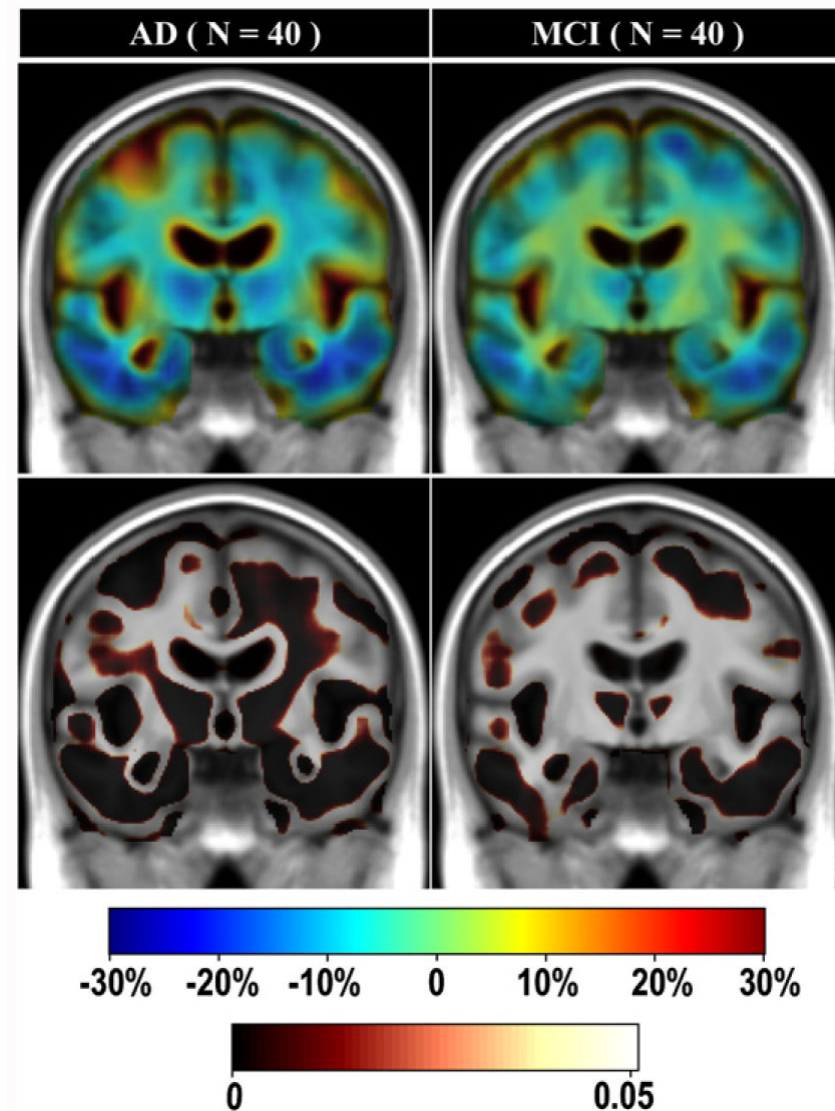


Fig. 6. 3D maps of brain atrophy (based on aligning group averages). These maps show the level of volumetric atrophy in AD and MCI relative to controls. In the top left panel, AD patients show prominent atrophy of up to 30% regionally in the temporal lobes, widespread reductions in the white matter, and notable expansion of the interhemispheric fissure (*coded in red colors*). MCI subjects show atrophy in the same regions but to a lesser degree. These maps are computed after spatial normalization of all brains to the same global scale, so regions with apparent excess tissue indicate regions with either absolute volumetric gains, or relatively greater tissue in proportion to overall brain scale. The bottom row shows the significance of these changes, computed using the variance of the individual deformation mappings, and color-coded according to the scale at the bottom.

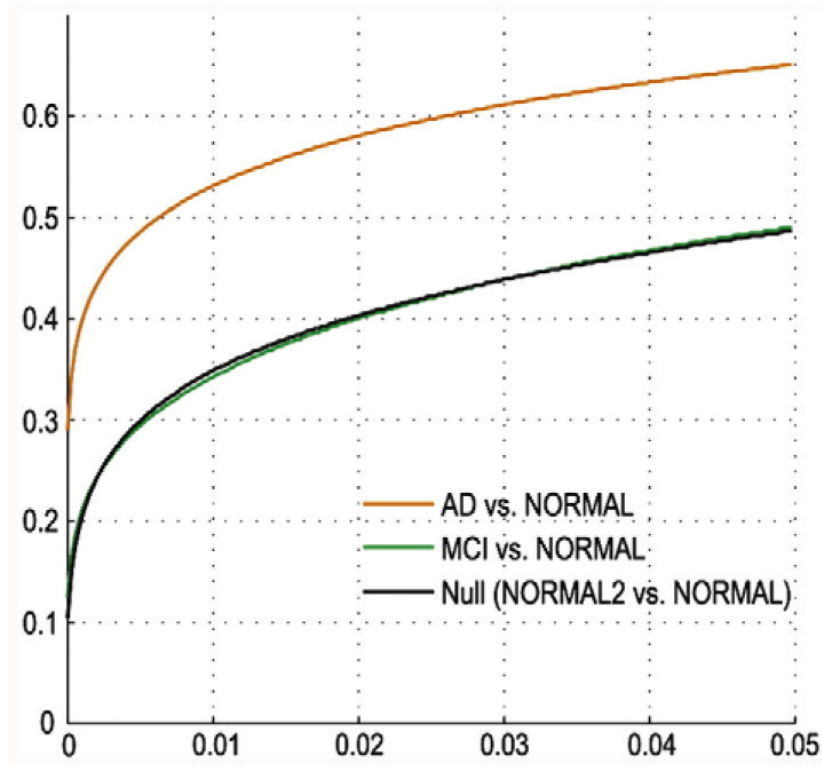


Fig. 7.

CDF Plots of Group Differences based on Aligning Group Average Images ($N=40$; 9P linear registration). Voxel-wise two-sample t tests were conducted using the Jacobians derived from direct alignment of group average templates and the variance term obtained from the method of averaging individual differences. The green curve (MCI vs. Normal) overlaps with the black curve which represents the empirically confirmed null distribution of statistics (registering one normal group template to the other). These lines are far from diagonal, and the method of aligning group averages, when used with a variance term from individual registrations, does not control false positives properly (FDR q -value for controls: 0.0001).

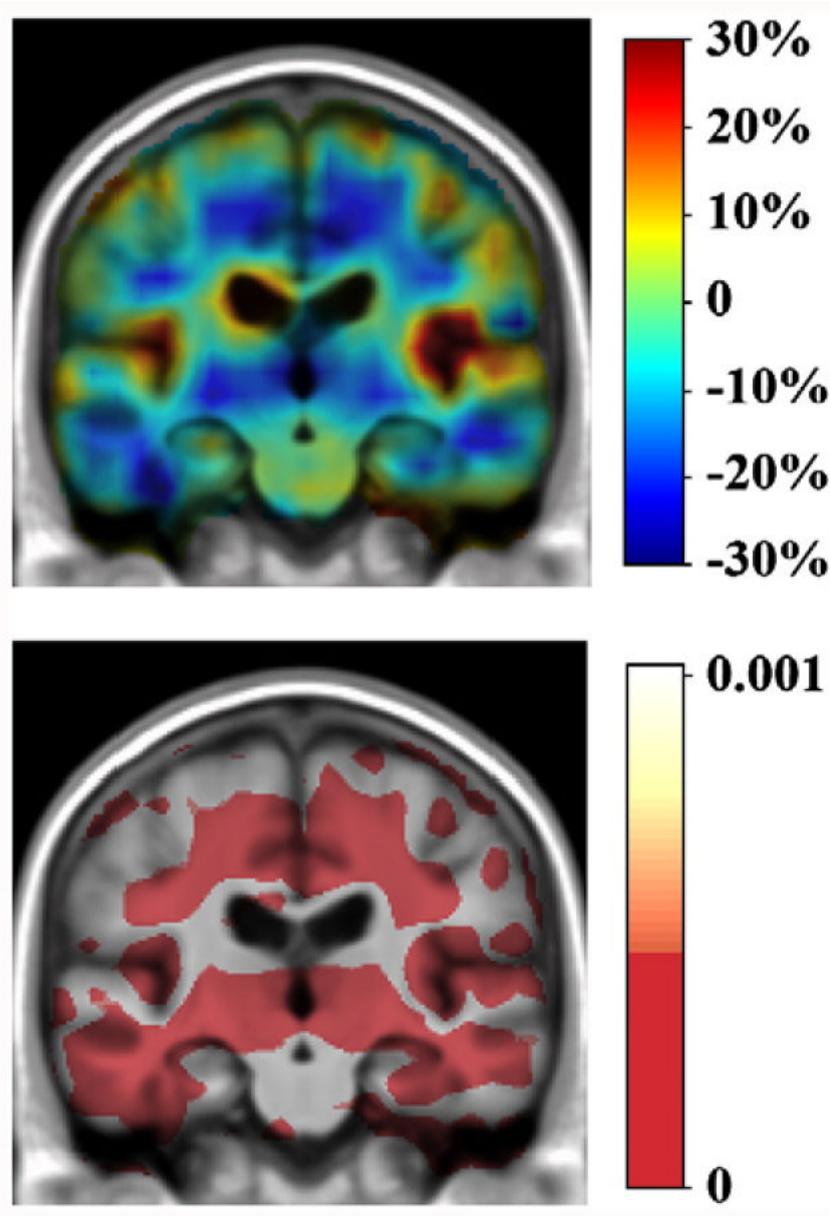


Fig. 8. Single-subject analysis. Here regional brain volumes in a single subject with AD are locally 20–30% lower than the control group average (top panel, blue colors) and 20–30% larger in some of the CSF spaces (red colors). Using the variance in the control group to assess the percentile, for regional volumes, at which this subject would fall relative to normal controls, most regions are well outside the confidence limits for normal volumes (lower panel, red colors). For caveats regarding the significance and interpretation of single-subject TBM maps, see the main text.

Table 1
Correlations with clinical measures of dementia severity

	MMSE Score	Global CDR	Sum-of-Boxes
Negative correlations (-1)	0.079	0.006	0.002
Two-tailed correlations (0)	0.001	0.007	0.001
Positive correlations (+1)	0.001	0.048	0.011

P values (corrected for multiple comparisons by permutation testing) for correlations between Jacobian maps and clinical measures. If negative correlations are significant, there are regions of the image (e.g., the ventricles) where greater expansion is correlated with lower MMSE scores. If positive correlations are significant, there are regions of the image (e.g., in gray and white matter) where volume reductions are correlated with lower MMSE scores. If two-tailed correlations are significant, there is evidence that structural differences in the group, whether they are contractions or expansions, are linked with cognition. For MMSE, the positive or two-tailed correlations – which are sensitive to atrophy – are more robust. Higher global CDR and sum-of-boxes scores denote greater impairment; in that case, the negative or two-tailed correlations are more reliable. Put simply, the atrophy (volume contraction) detected by TBM links better with cognition than volume expansions do (in the CSF spaces), although each is significantly associated with both MMSE and CDR. Strictly speaking, it may be that the CSF expansion signal has less signal to noise than the atrophic signal as we are using a statistical tests that depend on the total volume of regions that reach a certain threshold (supra-threshold volume and corrected *q*-values from FDR). It may be that, if the statistical tests had been formulated differently, e.g., as strict voxel-level comparisons (e.g. maximal *t*-statistics), they would detect CSF differences with greater effect sizes than atrophic effects.

Table 2
Mean Jacobian within each ROI (based on aligning group averages)

	White Matter	Frontal GM	Parietal GM	Temporal GM	Occipital GM
AD	-6.62%	-2.19%	-2.26%	-5.79%	1.52%
MCI	-3.06%	-4.61%	-3.14%	-4.32%	-0.17%

When aligning group averages, the tissue reductions, as a percentage relative to controls, are shown here for the AD and MCI groups.