Contents lists available at ScienceDirect

# NeuroImage



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

# Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: An ADNI study of 515 subjects

Xue Hua<sup>a</sup>, Suh Lee<sup>a</sup>, Igor Yanovsky<sup>c</sup>, Alex D. Leow<sup>a,b</sup>, Yi-Yu Chou<sup>a</sup>, April J. Ho<sup>a</sup>, Boris Gutman<sup>a</sup>, Arthur W. Toga<sup>a</sup>, Clifford R. Jack Jr<sup>d</sup>, Matt A. Bernstein<sup>d</sup>, Eric M. Reiman<sup>e</sup>, Danielle J. Harvey<sup>f</sup>, John Kornak<sup>g,h</sup>, Norbert Schuff<sup>i,j</sup>, Gene E. Alexander<sup>1</sup>, Michael W. Weiner<sup>i,j,k,l</sup>, Paul M. Thompson<sup>a,\*</sup> and the Alzheimer's Disease Neuroimaging Initiative

<sup>f</sup> Department of Public Health Sciences, UCD School of Medicine, Davis, CA, USA

### ARTICLE INFO

Article history: Received 4 May 2009 Revised 1 July 2009 Accepted 3 July 2009 Available online 14 July 2009

### ABSTRACT

Tensor-based morphometry (TBM) is a powerful method to map the 3D profile of brain degeneration in Alzheimer's disease (AD) and mild cognitive impairment (MCI). We optimized a TBM-based image analysis method to determine what methodological factors, and which image-derived measures, maximize statistical power to track brain change. 3D maps, tracking rates of structural atrophy over time, were created from 1030 longitudinal brain MRI scans (1-year follow-up) of 104 AD patients (age:  $75.7 \pm 7.2$  years; MMSE:  $23.3 \pm 1.8$ , at baseline), 254 amnestic MCI subjects (75.0  $\pm$  7.2 years; 27.0  $\pm$  1.8), and 157 healthy elderly subjects (75.9  $\pm$ 5.1 years;  $29.1 \pm 1.0$ ), as part of the Alzheimer's Disease Neuroimaging Initiative (ADNI). To determine which TBM designs gave greatest statistical power, we compared different linear and nonlinear registration parameters (including different regularization functions), and different numerical summary measures derived from the maps. Detection power was greatly enhanced by summarizing changes in a statisticallydefined region-of-interest (ROI) derived from an independent training sample of 22 AD patients. Effect sizes were compared using cumulative distribution function (CDF) plots and false discovery rate methods. In power analyses, the best method required only 48 AD and 88 MCI subjects to give 80% power to detect a 25% reduction in the mean annual change using a two-sided test (at  $\alpha = 0.05$ ). This is a drastic sample size reduction relative to using clinical scores as outcome measures (619 AD/6797 MCI for the ADAS-Cog, and 408 AD/796 MCI for the Clinical Dementia Rating sum-of-boxes scores). TBM offers high statistical power to track brain changes in large, multi-site neuroimaging studies and clinical trials of AD.

© 2009 Elsevier Inc. All rights reserved.

## Introduction

Alzheimer's disease (AD) is the most common form of dementia in people over the age of 60 (Jellinger, 2006). The disease affects more than 26 million people worldwide, including over 5 million in the U.S.

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

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

alone (with an estimated economic cost of 156 billion USD per year; Wimo et al., 2006). From onset to death, AD gradually erodes memory, language, and higher-order cognition over a time course of 10– 15 years (DeKosky and Marek, 2003; Goldman et al., 2001; Jellinger, 2006; Price and Morris, 1999). People with amnestic mild cognitive impairment (MCI)—a preclinical stage of AD—convert to AD at a rate of 10–25% annually (Petersen, 2000; Petersen et al., 2001; Petersen et al., 1994). Increasing efforts are directed towards treating those with MCI (Jack et al., 2008b; Jack et al., 2005) and people at



<sup>&</sup>lt;sup>a</sup> 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

<sup>&</sup>lt;sup>b</sup> Resnick Neuropsychiatric Hospital at UCLA, Los Angeles, CA, USA

<sup>&</sup>lt;sup>c</sup> Department of Mathematics, UCLA, Los Angeles, CA, USA

<sup>&</sup>lt;sup>d</sup> Mayo Clinic, Rochester, MN, USA

<sup>&</sup>lt;sup>e</sup> Banner Alzheimer's Institute, Department of Psychiatry, University of Arizona, Phoenix, AZ, USA

<sup>&</sup>lt;sup>g</sup> Department of Radiology and Biomedical Imaging, UCSF, San Francisco, CA, USA

<sup>&</sup>lt;sup>h</sup> Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA, USA

<sup>&</sup>lt;sup>i</sup> Veterans Affairs Medical Center, San Francisco, CA, USA

<sup>&</sup>lt;sup>j</sup> Department of Radiology and Biomedical Imaging, UCSF, San Francisco, CA, USA

<sup>&</sup>lt;sup>k</sup> Department of Medicine, UCSF, San Francisco, CA, USA

<sup>&</sup>lt;sup>1</sup> Department of Psychiatry, UCSF, San Francisco, CA, USA

<sup>1053-8119/\$ –</sup> see front matter 0 2009 Elsevier Inc. All rights reserved. doi:10.1016/j.neuroimage.2009.07.011

heightened genetic risk, e.g., those with amyloid precursor protein (APP) or presenilin mutations (Goate, 2006; Goate et al., 1991), apolipoprotein E (APOE)  $\varepsilon$ 4 carriers (Reiman et al., 2001; Reiman et al., 1996; Reiman et al., 2004), or Down's syndrome (Haier et al., 2003) (reviewed in Dickerson and Sperling, 2005). As new treatments are being developed, neuroimaging measures are urgently required to assess whether treatments slow disease progression in the brain. Ultimately, therapeutic agents aim to improve cognitive function (reduce symptoms), and also resist the advance of neurodegeneration that can be detected in vivo using neuroimaging (Clark and Karlawish, 2003; Fillit, 2008; Ivinson et al., 2008; Mueller et al., 2006).

Clinical or cognitive measures, such as the Alzheimer's Disease Assessment Scale (ADAS-Cog) (Rosen et al., 1984), are most commonly used to evaluate treatment efficacy in clinical trials, but their low testretest reliability makes it necessary to use large sample sizes and long observation times (Mueller et al., 2006). To address this, several neuroimaging consortia, including the Alzheimer's Disease Neuroimaging Initiative (ADNI), are determining which combination of neuroimaging, CSF, genetic, and cognitive biomarkers can collectively provide a more accurate early diagnosis of AD and monitor disease progression with greatest statistical power (Mueller et al., 2005a; Mueller et al., 2005b). There is great interest in determining which brain imaging methods can track disease progression in AD and MCI most powerfully. Ideal neuroimaging measures, or so-called biomarkers of disease progression, would correlate with cognitive performance, predict future clinical decline, and track longitudinal progression with high reliability and statistical power (Halperin et al., 2009; Mueller et al., 2006; Mueller et al., 2005b; Shaw et al., 2007). Eventually, clinical trials will need to provide the evidence needed to determine whether the effects of established diseaseslowing treatments on MRI or other biomarkers predict a clinical benefit for regulatory agencies to approve future treatments based solely on these biomarker surrogate endpoints (Reiman and Langbaum, in press).

Structural MRI, in particular, has been proposed and tested for monitoring treatment effects in AD (Grundman et al., 2002; Jack et al., 2008b; Jack et al., 2003; Scheltens et al., 2002). Neuroimaging measures, such as hippocampal volumes (Jack et al., 2002; Jack et al., 1999; Morra et al., 2009b; Morra et al., 2009c; Schuff et al., 2009), ventricular volumes (Carmichael et al., 2006; Chou et al., 2008, 2009a,b; Thompson et al., 2004a), and brain boundary shift integral (BBSI) measures (Fox et al., 2000), have been shown to differentiate patients from controls, correlate with changes over time in clinical and cognitive scores, correlate with pathologically confirmed neuronal loss, and predict future conversion from preclinical to symptomatic AD.

All of these neuroimaging approaches typically provide single numeric summaries (e.g., the volume of a structure) from each patient's 3D image set. There is growing interest in whether 3D brain mapping methods, which provide a detailed image of brain differences, can provide greater power to track disease progression than a single number derived from a scan (e.g., hippocampal volume). Tensor-based morphometry (TBM), for example, is a promising image analysis technique that computes the locations and rates of tissue atrophy. Changes are determined by elastically or fluidly aligning successive MRI scans of the same subject, using a registration algorithm. Maps of local expansion or compression factors can be used to estimate the local rates of tissue loss or CSF space expansion at each voxel (Ashburner and Friston, 2003; Chung et al., 2001; Fox et al., 2001; Freeborough and Fox, 1998; Riddle et al., 2004; Studholme et al., 2001; Thompson et al., 2000). TBM can also measure volumetric differences in cross-sectional studies by nonlinearly registering individual brain scans to a common anatomical template (Ashburner and Friston, 2003). TBM-derived measures also have several characteristics of an ideal AD biomarker as they have been shown to correlate with cognitive performance, and predict conversion from MCI to AD, in both cross-sectional (Hua et al., 2008a; Hua et al., 2008b) and longitudinal studies (Leow et al., 2009).

In this study, we investigated which methodological parameters, and which image-derived measures, provided the greatest power for tracking AD and MCI progression using TBM. First, we examined TBM designs with different linear and nonlinear registration parameters (including different regularizing functions (Leow et al., 2005; Yanovsky et al., 2008a)), which control the way in which the images deform. Second, we tested whether statistical power was increased if we summarized the changes detected in the maps using a more restricted region-of-interest (ROI). This ROI was defined statistically as the voxels with highest effect sizes for change over time in an independent training dataset (an approach advocated by Reiman, Chen and their colleagues for PET imaging; (Chen et al., 2009; Reiman et al., 2008). As the ROI approach provided substantial benefits, we subsequently performed exploratory analyses to determine how the minimal sample size estimates depended on how the ROI was thresholded and applied to the data. Differences in the statistical maps due to variations in the TBM design were examined by (1) ranking cumulative distribution function (CDF) plots that assess effect sizes for voxel-based maps, and (2) assessing minimal sample sizes to detect slowing of AD and MCI.

We hypothesized that TBM would lead to greatly reduced sample size estimates for clinical trials compared to existing cognitive measures, and would perform comparably to, and in some cases better than, other imaging measures for which power estimates have been reported. In addition, we expected empirically-defined statistical ROIs to boost power even further.

### Materials and methods

### Subjects

Longitudinal brain MRI scans and associated clinical data were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database (http://www.loni.ucla.edu/ADNI/Data/). ADNI is a large five-year study launched in 2004 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 nonprofit organizations, as a \$60 million public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments acquired at multiple sites (as in a typical clinical trial), can replicate results from smaller single site studies measuring 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 monitor the effectiveness of new treatments, and 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.

Longitudinal brain MRI scans (1-year follow-up) of 515 subjects were analyzed in this study (i.e., 1030 scans total), including 104 AD patients (age at initial scan:  $75.7 \pm 7.2$  years), 254 individuals with amnestic MCI ( $75.0 \pm 7.2$ ), and 157 healthy elderly subjects ( $75.9 \pm 5.1$ ). These subjects are from the same cohort as those in our prior TBM studies of 676 subjects analyzed at baseline (Hua et al., 2008b), and 100 subjects studied longitudinally (Leow et al., 2009). All subjects underwent thorough clinical and cognitive assessment at the time of scan acquisition. Cognitive tests examined here included the Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog), a 70-point scale designed to measure the severity of cognitive measure in AD trials (Rosen et al., 1984). It consists of 11 tasks assessing learning and memory, language production and comprehension, constructional and ideational praxis, and orientation. The

sum-of-boxes Clinical Dementia Rating (CDR-SB), ranging from 0 to 18, measures dementia severity by evaluating patients' performance in six domains: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care (Berg, 1988; Hughes et al., 1982; Morris, 1993). The Mini-Mental State Examination (MMSE) provides a global measure of mental status, evaluating five cognitive domains: orientation, registration, attention and calculation, recall, and language (Cockrell and Folstein, 1988; Folstein et al., 1975). The maximum MMSE score is 30; scores of 24 or lower are generally consistent with dementia. For both ADAS-Cog and CDR-SB, higher scores indicate poorer cognitive function, whereas for MMSE, higher scores denote better cognitive function. All AD patients met NINCDS/ADRDA criteria for probable AD (McKhann et al., 1984). Please refer to the ADNI protocol for detailed inclusion and exclusion criteria (Mueller et al., 2005a; Mueller et al., 2005b).

This dataset was downloaded on or before June 1, 2008, and reflects the status of the database at that point; as data collection is ongoing, we focused on analyzing all available baseline and 1-year follow-up scans, together with associated clinical and cognitive scores. The study was conducted according to the Good Clinical Practice guidelines, the Declaration of Helsinki and U.S. 21 CFR Part 50– Protection of Human Subjects, and Part 56–Institutional Review Boards. Written informed consent was obtained from all participants before experimental procedures, including cognitive tests, were performed.

### MRI acquisition and image correction

All subjects were scanned with a standardized MRI protocol developed for ADNI (Jack et al., 2008a). Briefly, high-resolution structural brain MRI scans were acquired at 59 ADNI sites using 1.5 T MRI scanners. ADNI also collects a smaller subset of data at 3 T but it was not analyzed here. A direct comparison of 3 T versus 1.5 T field strengths, using a smaller subset of subjects scanned at both field strengths, is reported separately in a different publication (Ho et al., in press).

Using a sagittal 3D MP-RAGE scanning protocol, the typical 1.5 T acquisition parameters were repetition time (TR) of 2400 ms, minimum full TE, inversion time (TI) of 1000 ms, flip angle of 8°, 24 cm field of view, and  $192 \times 192 \times 166$  acquisition matrix in the *x*-, *y*-, and *z*-dimensions, yielding a voxel size of  $1.25 \times 1.25 \times 1.2$  mm<sup>3</sup>, later reconstructed to 1 mm isotropic voxels (Chou et al., 2009b).

Image corrections were applied using a processing pipeline at the Mayo Clinic, consisting of: (1) correction of geometric distortion due to gradient nonlinearity (Jovicich et al., 2006), i.e. "gradwarp", (2) B1-correction for adjustment of image intensity inhomogeneity due to B1 nonuniformity (Chou et al., 2009b), (3) N3 bias field correction for reducing residual intensity inhomogeneity (Sled et al., 1998), and (4) geometrical scaling for removing scanner- and potential session-specific calibration errors using a phantom scan acquired for each subject (Gunter et al., 2006). All original image files as well as images with all of these corrections are available to the general scientific community at http://www.loni.ucla.edu/ ADNI/Data/.

### Image pre-processing

To adjust for linear drifts in position and scale within the same subject, the follow-up scan was linearly registered to its matching baseline scan using a 6-parameter (6P) or 9-parameter (9P) registration, driven by a mutual information (MI) cost function (Collins et al., 1994). The 9P registration allows scaling of the scan along 3 orthogonal axes, whereas the 6P registration is a rigid-body alignment and does not allow scaling. To account for global differences in brain scale across subjects, both mutually aligned scans were then linearly registered to the International Consortium for Brain Mapping template (ICBM-53) (Mazziotta et al., 2001), applying the same 9P transformation to both mutually aligned scans. Globally aligned images were resampled in an isotropic space of 220 voxels along *x*-, *y*- and *z*-dimensions with a final voxel size of 1 mm<sup>3</sup>. All scans were original images after image corrections thus no brain mask was used.

# Tensor-based morphometry (TBM) and three-dimensional Jacobian maps

Using high-dimensional nonlinear registration, individual Jacobian maps (i.e., maps of local expansion or compression) were created to quantify 3D patterns of atrophic rates by warping the follow-up scan to match the baseline scan. All the elements of this process are automated; there is no user intervention apart from that required to select scans to analyze and to specify covariates and statistical models. Two separate methods were compared: one method (termed "3DMI") (Leow et al., 2005), was driven by a mutual information cost function, and a regularizing term based on the linear elasticity operator. A second, alternative, algorithm (termed "sKL-MI") used the symmetric Kullback–Leibler (sKL-MI) distance, another measure from information theory, with varying registration parameters, sigma ( $\sigma$ ) and lambda ( $\lambda$ ), denoted as S and L respectively to differentiate from the  $\sigma$  used in power analysis, that control the Jacobian field smoothness and weighting of the regularization, respectively (Yanovsky et al., 2009; Yanovsky et al., 2008a). Briefly, the registration method based on linear elasticity is more commonly used in brain imaging, and has been used in several of our prior TBM studies (e.g., (Gogtay et al., 2008; Hua et al., 2008b)). It models the deforming image as if it were embedded in a deforming physical medium obeying continuum-mechanical laws (Alexander et al., 2001; Broit, 1981; Leow et al., 2006a; Leow et al., 2006b; Shen and Davatzikos, 2002; Thompson et al., 2000). The sKL model causes images to deform in a slightly different way, such that the expansion factor or compression factor (Jacobian determinant) is as spatially uniform as possible in regions where both images have homogeneous intensity. This model, based on information theory, has been advocated as it can avoid statistical bias in the maps of brain change when no true changes are present (Leow et al., 2006a; Leow et al., 2007; Yanovsky et al., 2007a; Yanovsky et al., 2009; Yanovsky et al., 2008b; Yanovsky et al., 2007b). Essentially, each method tends to interpolate the deformation into white matter regions in slightly different ways, and as there is no ground truth (i.e., reference or "gold standard" method) to compute changes in those regions, it is of interest to see which method provides best power for detecting statistical effects on anatomy.

The sKL-MI method solves for the deformation (or, equivalently, for the 3D displacement field) minimizing a compound energy functional consisting of two terms, (1) the image matching term  $F_{MI}$ , and (2) a regularization term based on the symmetric Kullback–Leibler divergence,  $R_{SKL}$ . There are 2 free parameters (*S* and *L*) that can be set by the user, so we also examined their impact on the results. Briefly, the 3D vector-valued flow field  $\boldsymbol{u}$  is obtained by driving image  $I_2$  into registration with image  $I_1$  using a force field (Yanovsky et al., 2009):

$$f(\boldsymbol{u}) = -\frac{\partial F_{\mathrm{MI}}}{\partial \boldsymbol{u}}(I_1, I_2, \boldsymbol{u}) - \lambda \frac{\partial R_{\mathrm{sKL}}}{\partial \boldsymbol{u}}(\boldsymbol{u}),$$

where  $\lambda(L)$  is the weight on the regularization term  $R_{sKL}$ . The instantaneous velocity field v is obtained by convolving f with a Gaussian kernel  $G_{\sigma}$  of variance  $\sigma^2$  ( $S^2$ ):

$$\boldsymbol{v} = G_{\sigma} * f(\boldsymbol{u}).$$

Given the velocity vector field v, the displacement vector field u is obtained by advancing the following partial differential equation in time *t*:

$$\frac{\partial \boldsymbol{u}}{\partial t} = \boldsymbol{v} - \boldsymbol{v} \cdot \nabla \boldsymbol{u}.$$

For both methods, color-coded maps of the Jacobian determinants (the local expansion or compression factors) were used to illustrate regions of volume expansion (i.e., with det J(r)>1), or contraction (with det J(r)<1) (Ashburner and Friston, 2003; Chung et al., 2001; Freeborough and Fox, 1998; Riddle et al., 2004; Thompson et al., 2000;

Toga, 1999) over time. These maps of tissue change rates were also spatially normalized across subjects by nonlinearly aligning all individual Jacobian maps to a minimal deformation template (MDT), for regional comparisons and group statistical analyses. The MDT was constructed based on images from 40 normal controls as detailed elsewhere (Hua et al., 2008a; Hua et al., 2008b).

### Pair-wise group comparisons

To illustrate systematic differences in atrophic rates between groups, we constructed voxel-wise statistical maps based on the Student's *t* statistic. We corrected for the multiple comparisons implicit



**Fig. 1.** These plots show how different analysis choices can boost the statistical power to detect group differences in the rates of brain atrophy. Cumulative distribution functions (CDFs) are based on *p*-value maps detecting group differences in atrophic rates between (a) AD and normal controls, and (b) MCI and normal controls, using a variety of TBM designs with different choices of regularizing functions (SkL-MI or 3DMI). Nonlinear registration parameters were also varied in the case of skL-MI, which has free parameters that can be set by the user (sigma (S) = 6 or 9; lambda (L) = 0.5, 1, 2, 5, 8, or 10). According to the definition of FDR, the rightmost nonzero intersection of the y = 20x line and the CDF for a significance map, is the *q*-value, i.e., the highest *p*-value threshold for which at most 5% false positives are expected in the map. If other factors are equal (such as sample size), a larger *q*-value generally indicates a more powerful statistical effect—in our case, a greater statistical distinction between patient (AD or MCI) and normal groups. As shown by the CDF ranking and the *q*-values, the best TBM design for this dataset used the skL regularizing function, with registration parameters S = 6 and L = 8 (denoted by skL-MI S6L8).

in making a statistical map, by using permutation tests (Bullmore et al., 1999; Chiang et al., 2007a; Chiang et al., 2007b; Nichols and Holmes, 2002; Thompson et al., 2003). In brief, a null distribution for the group differences in Jacobian (atrophic rate) at each voxel was constructed using 5000 random permutations. For each test, the subjects' diagnosis was randomly permuted and voxel-wise t-statistics were calculated. A ratio, describing the fraction of the time the *t*-statistic was more extreme in the randomized tests than the original test, was calculated to give a permutation-based *p*-value for the significance at each voxel. A "global P-value", describing the fraction of time the suprathreshold volume (p < 0.01, uncorrected) was greater in the randomized maps than the real effect (the original labeling), was calculated to determine whether any significant change could be detected across the brain. This procedure has been used in many prior reports (Braskie et al., 2008; Chiang et al., 2007a; Chou et al., 2009b). This is one of several standard ways to set up a permutation test and is sometimes called set-level inference. It deems a map significant when the total quantity of voxels with *p*-values lower than a fixed *a priori* threshold exceeds that obtained in 95% of random simulations.

### Cumulative distribution functions (CDF)

CDF plots of the *p*-values were used to compare the effect sizes of (and therefore the power to detect) statistical effects when using (1) TBM designs with different regularizing functions (3DMI versus sKL-MI), (2) different sKL-MI nonlinear registration parameters (sigma = 6 or 9; lambda = 0.5, 1, 2, 5, 8, or 10), and (3) different choices for the pre-TBM linear registration step (6P versus 9P linear registration of follow-up and baseline scans). 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). CDFs can also be used to compare effect sizes of different methods based on the expected proportions of voxels with significance above any given statistical threshold under

the null hypothesis. We have used this CDF ranking method in several prior reports to compare the power of different methods applied to the same set of images (Chou et al., 2008; Hua et al., 2008a; Lepore et al., 2008; Morra et al., 2009b).

The *q*-value refers to the point (corresponding value from the x axis) at which CDF plot intersects with the line y = 20x, and this represents the highest statistical threshold for which at most 5% false positives are expected in the map. Significance is declared if the CDF intersects the y = 20x line (other than at the origin), as this shows that the volume of supra-threshold statistics is more than 20 times the value that would be expected by chance alone, under null-hypothesis. The *q*-value is a measure of the overall significance of each *p*-value map, with q = 1 indicating full power to always control the false positive rate and q = 0 signifying that no thresholding of the data can adequately control the false positive rate (i.e., the findings are null). The empirical CDF of *p*-values is not to be confused with the more common FDR PP plot, which is the flip of this plot (i.e., the axes are interchanged). This procedure has been used in several of our prior publications (Chou et al., 2009a; Hua et al., 2008a; Morra et al., 2009c).

### Regions of interest (ROIs) and mean atrophy rate

Both anatomically and statistically-defined ROIs were used in this study. First, a temporal lobe ROI, including the temporal lobes of both brain hemispheres, was manually delineated on the MDT template by a trained anatomist using the Brainsuite software program (Shattuck and Leahy, 2002). Secondly, an alternative statistically-defined ROI (subsequently called the "stat-ROI") was defined based on voxels with significant atrophic rates over time (p<0.001) within the temporal lobes, in a nonoverlapping training set of 22 AD patients available at the time of image download, i.e., independent of the testing set used to compute the power numbers (see Fig. 4 for an example). Note that this is a somewhat arbitrary threshold approach to selecting an ROI; in principle the ROI could be chosen specifically to provide maximal



**Fig. 2.** Global scaling improves power. We examined how effect sizes (for detecting group differences in atrophic rates) depended on whether 9-parameter or 6-parameter linear registration was used to align successive scans before performing TBM. Comparisons are based on CDFs of *p*-values within the temporal lobe ROI. The curves show increasing differentiation between groups in rank order from the bottom to top. The 9P linear registration (solid lines) showed superior effect sizes relative to those based on rigid-body (6P) registration (dotted lines in matching colors). This was found regardless of the choice of regularizing function (sKL-MI S9L5 or 3DMI) and patient group (AD versus normal or MCI versus normal).

power (see later, for more on this topic). A separate stat-ROI was generated and applied for each TBM design with different algorithms (3DMI versus sKL-MI) and different parameter settings (6 versus 9P linear registration and different *S* and *L*). A numeric summary—the mean atrophy rate for all voxels with in the ROI—was computed for each person, to summarize annual change within the ROI. We note that ROIs based on independently generated statistical maps have been advocated for use in PET studies by Reiman et al. (Chen et al., 2009; Reiman et al., 2008), and it is also common in fMRI studies to threshold statistical maps using regions identified in orthogonal statistical contrasts or based on independent analyses.

### Sample size calculation and power analysis

A power analysis was defined by the ADNI Biostatistics Core to estimate the sample size required to detect a 25% reduction in the mean annual rate of atrophy, using a two-sided test and standard significance level ( $\alpha$ =0.05) for a hypothetical two-arm study (treatment versus placebo). We used the individual mean atrophic rates computed within the stat-ROI for sample size calculations. In brief,  $\beta$  denotes the estimated annual change (average of the group) and  $\sigma_D$  refers to the standard deviation of the rate of atrophy across subjects. The estimated minimum sample size for each arm is computed from the formula:

$$n = \frac{2\hat{\sigma}_D^2 \left(z_{1-\alpha/2} + z_{\text{power}}\right)^2}{\left(0.25\hat{\beta}\right)^2}$$

Here  $z_{\alpha}$  is the value of the standard normal distribution for which  $P[Z < z_{\alpha}] = \alpha$  and in this case we set  $\alpha$  to its conventional value of 0.05 (Rosner, 1990). We estimated the sample size required to achieve 80% and 90% power (subsequently we will refer to these as n80 and n90).

# Controlling false positives: hypothesis-driven and post hoc exploratory tests

Because we are testing a large number of methodological alternatives, it is necessary to take precautions to avoid inflating the false positive rate due to the large number of comparisons conducted. For example, it is not legitimate to compute effect sizes and minimal sample sizes for a method, and then vary the free parameters of the method until the effect sizes are as large as possible. The latter approach can be used for exploratory purposes but not for drawing statistical inferences regarding the sample sizes. As such, we declared one method at the outset as the "default" TBM method, namely using TBM with our most recent implementation of nonlinear registration. We therefore declared the default method to be the sKL-MI method with parameters set to default values used in our prior papers (S = 9, L=5, after 9P linear registration (Yanovsky et al., 2009); we also thresholded the statistical ROI at an *a priori* statistical threshold of p = 0.001. The estimation of sample sizes for the default method may be considered as appropriate for purposes of statistical inference. We additionally explored the influence of varying free parameters (S=6 or 9; L=0.5, 1, 2, 5, 8, or 10) on sample size estimates, as *post* hoc tests to see if sample sizes could be further improved. We note that these power improvements should be regarded as post hoc findings while the best model would then provide the best parameter set and ROI combination for future studies and clinical trials. Particularly, we chose a subset of parameters that covered a dynamic range with relatively small increment to capture a potential nonlinear trend on statistics derived from the CDF and sample size estimates. We did not attempt to test all possible parameter combinations, as each test run required the computation of N = 515 Jacobian maps and permutationbased *t* tests with an average computation time of 144 h on the LONI computer cluster, which is equipped with 1152 dedicated Sun

### a. AD versus Normal (Global P value < 0.0001)



**Fig. 3.** 3D maps show the mean local atrophic rate in 104 AD subjects (a) and 254 MCI subjects (b) relative to the normal group (157 controls). The left columns of panels a and b show the mean rate of atrophy as a percentage tissue loss, or a percentage increase in ventricular expansion in a year. The right columns of a and b show the voxel-wise significance of these inter-group differences, showing where the group of interest has faster atrophy than normal. Both groups show faster degeneration than that seen in normal aging, with a more widespread pattern in AD and a more anatomically restricted pattern of (significant) active atrophy in MCI. Global *P* values (corrected significance values) showed significant group differences in atrophic rates between AD and normal controls (P<0.0001) and between MCI and controls (P=0.0004).



**Fig. 4.** Statistical ROI. Here the statistically-defined ROI, based on using the TBM design that showed greatest power in the prior experiments (sKL-MI S6L8, p<0.001), is shown overlaid (in red colors) on selected slices of the MDT. The stat-ROI was computed based on an independent training sample of N = 22 AD subjects, considering only voxels that showed progressive atrophy over time, so long as they also fell within the temporal lobe ROI defined earlier. The stat-ROI was used to further improve the detection power by averaging the signals within the regions known to have highest atrophy rates (or, strictly speaking, highest effect sizes for atrophy rates) in the AD training sample.

Microsystems 64-bit 2.2–2.4 GHz processors with 8–16 GB of memory per node.

### Results

### Optimizing the TBM design by ranking CDFs

First, CDF plots were used to compare the effect sizes of brain changes over time, as detected using different TBM designs. In Figs. 1 and 2, the CDFs of the *p*-values observed for the statistical comparison of atrophic rates of patients (AD or MCI) versus normal controls were plotted against the corresponding *p*-value that would be expected under the null hypothesis of no group difference. Greater effect sizes are represented by larger deviations (upswings) near the origin of these CDF plots. The curves show increasing effect sizes, in rank order from bottom to top, for detecting voxels with statistical differences between groups. In other words, the most powerful methods are shown at the top, but all curves that cross the y = 20x line are considered to have detected significant findings. By comparing AD and MCI with the control group, the greatest differentiation was achieved by the nonlinear registration based on the sKL-MI regularizing function with S = 6 and L = 8, though a few TBM designs attained very similar effect sizes, e.g., sKL-MI S9L8 and S9L10 (Fig. 1).

The 9P linear registration, used prior to the steps of TBM, yielded superior effect sizes compared to using a 6P rigid-body registration, regardless of the choice of regularizing function (Fig. 2). Group differences are therefore easier to detect if some global scaling is allowed when aligning the baseline and follow-up MRIs from the same subject. This is consistent with prior observations in long-itudinal MRI studies by Paling et al. (2004). In summary, the optimal TBM design consists of using 9P linear registration of the follow-up to the baseline scans, with subsequent nonlinear registration regularized by sKL-MI, using parameter settings S = 6 and L = 8. The group



**Fig. 5.** Mean atrophy rate and sample size estimates required to detect a 25% slowing of atrophic rates with 80% power (n80), using various TBM designs. See Materials and methods section for caveats on interpreting these samples sizes before and after varying the free parameters of the method. The nonlinear registration parameter *L* controls the level of sKL-MI regularization (spatial smoothness of the recovered changes). For all these experiments, the mean atrophic rate was computed in the statistically-defined ROI shown in Fig. 4, which deliberately focuses on voxels that changed the most in an independent (nonoverlapping) dataset of AD subjects (p < 0.001). As such, the average atrophic rates, reported here, are slightly higher than the mean atrophic rates measured by methods that consider overall anatomical regions of interest defined by an atlas, such as the whole brain or the temporal lobes, for example, which do not deliberately reject voxels where atrophic rates are slower. As expected from theoretical considerations, the average and standard deviation of the mean atrophy rate decreased monotonically when the smoothness parameter *L* was increased. The parameter *S* controls the fluid regularization or the Jacobian field smoothness, to some extent. The effect of varying *S* (from 6 to 9) had relatively little influence on sample size estimates. The 3DMI deformation model showed an inferior ratio of mean to standard deviation (i.e., a lower effect size) and thus led to higher sample size estimates compared to sKL-MI.

### Table 1

Estimated sample sizes needed to detect a 25% reduction in the mean annual change with a two-sided test and  $\alpha$ =0.05 at 80% (n80) and 90% (n90) power, for a two-arm study, using a statistically-defined ROI versus an anatomically-defined temporal lobe ROI.

TBM designs	AD				MCI			
	Stat-I	-ROI Temporal-ROI		Stat-ROI		Temporal-ROI		
	n80	n90	n80	n90	n80	n90	n80	n90
sKL-MI S6L1	60	81	52	70	93	124	74	99
sKL-MI S6L5	49	65	52	70	86	116	89	119
sKL -MI S6L8	48	64	55	74	88	118	99	132
sKL-MI S9L1	52	70	62	83	75	100	98	131
sKL-MI S9L5	52	69	72	96	85	114	132	176
sKL-MI S9L8	55	73	78	104	93	124	148	198
3DMI	125	167	183	244	323	433	401	536

The default method (whose parameters were chosen *a priori*) is shown in bold. The method that subsequently yielded the smallest sample sizes after allowing the free parameters of the method to vary, is shown underlined. There is very little difference between these methods—all giving excellent power—when the statistical ROI is used.

difference maps and sample size estimates, reported from this point onwards, were based on the 9P linear registration, as it was part of the default method.

### 3D profiles of brain atrophy

From the individual Jacobian maps computed with the optimal TBM design (sKL-MI S6L8), we derived maps of mean group differences in atrophic rates, and statistical maps of group differences based on the permutation-based t test, corrected for multiple comparisons. The 3D maps contrasting AD with normal controls revealed much higher rates of ongoing atrophy in the hippocampi, temporal lobes, and almost all of the white matter, accompanied by progressive CSF expansion in the lateral ventricles and circular sulcus of the insula (Fig. 3a). The only region where progressive white matter atrophy was not found to be faster in AD than in healthy subjects was the white matter of the prefrontal cortex, as seen in Fig. 3a. This is consistent with the anatomical trajectory of AD in the cortex, in which temporal and parietal areas typically undergo fastest atrophy in the earliest stages of the disease, and structural atrophy moves into the frontal poles only relatively late in the disease (Braak and Braak, 1991; Braskie et al., 2008; Thompson et al., 2003).

The MCI group displayed similar patterns and locations of ongoing atrophy as in AD, but the regions in which atrophic rates were faster than for controls was more anatomically restricted and was concentrated on the temporal lobes (Fig. 3b). Of course, there may be active atrophy in all brain regions; these maps only emphasize where the patient groups show detectably faster atrophy than normal controls. As expected, AD patients (global P<0.0001) and MCI subjects (global P=0.0004) showed overall faster atrophic rates than controls (both P values were corrected for multiple comparisons using permutation testing).

#### Minimal sample size estimates and power analyses

For a two-arm study, the sample size required to detect a 25% reduction in the mean annual change with a two-sided test and  $\alpha = 0.05$ , was estimated using a power analysis. The stat-ROI was used to further improve the detection power by concentrating on the signals within the regions that were already known to be undergoing detectable atrophy in a nonoverlapping sample of AD training subjects (Fig. 4). All power analyses (based on stat-ROIs) excluded the 22 AD training subjects, to ensure a fair comparison. We used the mean atrophy rates computed within the stat-ROI for sample size calculations.

As indicated by the mathematical formula and illustrated in Fig. 5, the sample size estimate depends on the mean annual change and its

standard deviation in the sample analyzed. Clearly, the required sample sizes are smaller for methods for which the mean rate of change in the parameter measured is higher and/or its standard deviation is lower. The nonlinear registration parameter *L* controls the sKL-MI regularization. Increasing the value of *L* leads to more uniform Jacobian determinant values in regions of homogeneous intensity in the scan (e.g., across white matter regions). As a result, both the mean and standard deviation of the mean atrophy rate decreased monotonically when *L* was increased. The parameter *S* controls the fluid regularization or the Jacobian field smoothness. Increasing the value of *S* makes the Jacobian fields smoother, or less noisy. The effect of switching *S* from 6 to 9 had relatively little influence on sample size estimation, so was not considered further.

Based on the CDF rankings (Fig. 1), and the minimal sample size estimates (Fig. 5), we concluded that the best TBM design is based on sKL regularization with parameters set to S = 6 and L = 8, leading to 48 and 88 as sample size estimates for detecting clinically meaningful changes (25% reductions in the mean atrophic rate) in AD and MCI respectively, with 80% power (Fig. 5 and Tables 1, 2). Detection power was greatly enhanced by using a stat-ROI, defined by an independent training sample of 22 AD patients, compared to an anatomically defined temporal lobe ROI (Table 1).

Although varying the sKL-MI parameters influenced the mean and standard deviation of the average atrophic rate, the sample size estimates remained relatively stable (Fig. 5). The results suggested that sKL-MI is a highly robust method, at least within a plausible range of parameters, and the results were relatively stable regardless of the choice of registration parameters.

The sample size estimates based on TBM (with sKL-MI and 3DMI registration methods) are presented together with those based on standard clinical and cognitive measures (rates of change in ADAS-Cog, MMSE, and CDR-SB) in Table 2. The TBM-derived numeric summaries are about 8 times more powerful than the best clinical assessments (in this case, CDR-SB), in terms of minimal sample size requirements. TBM-derived mean atrophy rates within the stat-ROI offered greatly reduced sample size estimates (i.e., gave superior statistical power) for a hypothetical clinical trial design that assesses reductions in the mean rate of atrophy.

### Influence of the threshold used to define the statistical ROI

In a post-hoc exploratory analysis, we examined how the group mean atrophy rate and sample size estimates (Fig. 6) were affected by

#### Table 2

Sample size required to detect a 25% slowing of atrophic rates with 80% (n80) and 90% (n90) power, using biomarkers defined by TBM versus clinical measures in the same sample of subjects.

	AD n80/n90	MCI N80/n90
sKL-MI S9L5 (a priori) <sup>a</sup>	52/69	85/114
sKL-MI S6L8 <sup>a</sup>	48/64	88/118
3DMI <sup>a</sup>	125/167	323/433
ADAS-Cog	619/828	6797/9092
MMSE	1078/1442	3275/4381
CDR-SB	408/546	796/1065

We note that these may differ slightly from prior reports as they depend to some degree on the cohort studied.

<sup>a</sup> sKL-MI stands for the nonlinear registration algorithm that is driven by a mutual information (MI) cost function and that has a regularizing term based on the symmetric Kullback–Leibler (sKL) distance. *S* and *L* denote registration parameters sigma and lambda, controlling the Jacobian field smoothness and the weighting of the regularization, respectively. 3DMI stands for the alternative nonlinear registration algorithm that is driven by a MI cost function, and a regularizing term based on the linear elasticity operator. Power analysis was based on mean atrophic rate computed within the stat-ROI that was defined as including voxels with a significant atrophy rate (p < 0.001) within the temporal lobes, computed in an independent AD training sample (N = 22), using the TBM design of sKL-MI and 3DMI respectively.



**Fig. 6.** Mean atrophy rate and estimated sample sizes required to detect a 25% slowing of atrophic rates with 80% power (n80), using various levels of statistical thresholding to define the statistical ROI in which mean atrophic rates are computed. As a *post hoc* exploratory experiment, the influence of the stat-ROI threshold was examined for the TBM designs sKL-MI S6L8 (a) and 3DMI (b). The default threshold was p < 0.001. Smaller ROIs improve the power of the 3DMI method but they matter less for sKL-MI, which is the method with the highest power anyway.

the statistical threshold chosen to define the statistical ROI. Clearly, if a region is defined for signal averaging in an image, a very small region can be defined by setting the statistical threshold very high, but this might not give the highest effect size for the mean atrophic rate in new subjects, as different subjects show slight differences in anatomical localization for the location of greatest atrophy. As such, there is no immediate theoretical basis for choosing the extent or threshold for the statistical ROI; these depend on empirical factors (and the sample size of training images).

Reducing the statistical threshold used to define the ROI had little impact on the group mean atrophy rates and sample size estimates when sKL-MI was used to measure atrophic rates (Fig. 6a). However, sample size estimates were gradually improved in the scenario of 3DMI, with stable mean atrophy rate but lower standard deviation as the threshold was decreased and the ROI became smaller (Fig. 6b). This differential influence of the stat-ROI threshold was somewhat expected. In theory, the sKL-MI makes the Jacobian maps more spatially homogeneous in large areas with homogeneous intensity, e.g., the white matter. In a sense, the sKL statistical maps tend to have higher spatial autocorrelation (i.e., are spatially smoother), thus the exact value used for thresholding them has minimal impact. A more detailed understanding of these correlations in deformation maps can be gained by studying the 6-dimension Green's function (spatial covariance function) of the differential operator governing the deformation mappings, a topic that we have investigated previously; the Green's function can also be learned empirically from data (Brun et al., 2009a; Brun et al., 2009b; Fillard et al., 2007; Fillard et al., 2005; Gee, 1999; Grenander and Miller, 1998).

### Discussion

Here we found that a regional numeric summary, derived from TBM, provided a drastically reduced sample size estimate in a power analysis for detecting brain change, with 48 and 88 subjects required in AD and MCI groups respectively for a clinical trial designed to detect 25% improvement in the rate of decline with 80% power. In Table 2, sample size estimates using TBM-based methods were compared with those based on clinical scores in the same group of subjects (N = 515). The best clinical measure was CDR-SB, but even this measure required relatively large sample sizes (n80: 408 for AD, 796 for MCI). These sample sizes are around 8 times higher than those required by the TBM-derived imaging measure. Our power estimates for the clinical

measures (MMSE, ADAS-Cog) are similar to those computed by Schuff et al., 2009, in other ongoing ADNI studies. As confirmed by other independent studies, the required sample sizes were drastically lower for structural MRI versus clinical measures (Fox et al., 2000; Jack et al., 2004; Schuff et al., 2009). Specifically, when TBM was used in combination with an empirically-defined statistical ROI, the resulting numeric summary was comparable to, or better than, other imaging measures for which power estimates have been reported (Beckett et al., 2008; Chen et al., 2009; Tosun et al., 2009). We note that for summarizing our TBM measures, the average of all voxels in the ROI was taken, as it was the simplest and most intuitive numerical summary that can be derived from an ROI. Other approaches are clearly possible, such as taking a weighted average-by weighting the voxels according to their effect sizes in the stat-ROI-or using a machine learning method to select the subset of most informative voxels when used jointly for group discrimination (see Sun et al., (in press) and Ferrarini et al., (2008) for examples of the latter approach). Other map-based numeric summaries will be the topic of future study.

In this large-scale analysis—to our knowledge, the largest longitudinal TBM study to date—we evaluated different TBM designs with optimized linear and nonlinear registration parameters, to maximize the power to detect progressive brain atrophy in AD and MCI. We tested twelve different combinations of TBM modules, with each test involving 515 high-dimensional nonlinear registrations and subsequent CDF-based statistical analyses. Even though computational requirements were immense, our findings of the best parameters and analysis choices can now be used in future studies to expedite them.

These results offer several insights regarding the choice of TBM design, including the choice of the regularizing function (sKL-MI versus 3DMI), selection of nonlinear registration parameters that affect the smoothness of the recovered changes (parameters *S* and *L* for the sKL method), as well as the type of linear registration (rigid or nonrigid) that spatially aligns the scans before nonlinear registration. Using CDF plots, we ranked different TBM designs in terms of their power to detect significant effects and compared *q*-values from an FDR analysis. The best power, represented by the top CDF curve and the highest *q*-value, was achieved with a TBM design with the sKL-MI regularization (with parameters *S* = 6, *L* = 8), and 9P-linear registration. The parameter settings chosen as defaults in our prior work (S9L5) gave very similar sample sizes (52 AD, 85 MCI) (Yanovsky et al., 2009). It is conceivable that the best settings may depend somewhat

on whether AD or MCI is the target of the study, and they may depend somewhat on the inter-scan interval (here 1 year), with shorter follow-up periods giving noisier data and requiring even greater spatial regularization.

From the CDF plots (Fig. 1), we concluded that first, the elastic deformation model (3DMI) performed approximately as well as the deformation model based on information theory (sKL-MI), with sKL-MI performing slightly better or worse depending on how its parameters were set. When using the optimal combination of parameters (S = 6, L = 8), the sKL-based method demonstrated superior statistical power versus the 3DMI-based TBM design. The sKL method has been advocated in our prior work (Yanovsky et al., 2008a) as it avoids some of the drawbacks with standard image warping methods. For example, brain changes over time detected by most of the registration methods in the literature do not have a meanzero null distribution when no true biological changes are present, due to asymmetry in the registration process (see Leow et al., (2007) for a detailed discussion and mathematical derivations). Second, all deformation models successfully detected differences in atrophic rates between AD and controls, and MCI and controls, as is evident from the fact that all the CDFs rise sharply near the origin and all cross the y = 20x line. This means that for all methods and models, the results were significant in this very large study, but the effect sizes at each voxel differed slightly. Put another way, there is a range of statistical thresholds [0,q] that control the expected false discovery rate to be below 5%, and *q* is higher for the more powerful methods, meaning that a greater range of thresholds can be used to control the false discovery rate. Thirdly, and as expected, the curves showing MCIcontrol differences had lesser effect sizes than those showing ADcontrol differences, which make sense intuitively as atrophic rates in AD are the fastest of all, and atrophic rates in MCI are significantly faster than in controls.

The nonlinear registration parameter L controls the sKL-MI regularization. Increasing the value of L leads to lower mean and standard deviation of the atrophy rate (Fig. 5). This is because the heavier sKL-based regularizations (i.e., higher values of L) tend to reduce bias in homogeneous regions by producing uniform Jacobian determinant maps (e.g., in regions of white matter and ventricles where spatially coherent tissue loss and expansion are expected respectively). In a sense, sKL-based regularization is a type of spatial smoothing that is dominant only within inherently homogeneous regions. Therefore, it does not have the disadvantage of general smoothing which tends to average out signals from adjacent regions with distinct features as well as tissue change rates. For example, at the boundary of gray matter and CSF, it would be unfavorable to average or smooth the signals with opposite signs, i.e. loss of GM and expansion of CSF. Whereas in homogeneous regions of white matter and inside the ventricles, it is ideal to apply smoothing to generate uniform tissue change rates, in default of any other information that would lead to greater spatial resolution. This is at least conceptually related to the intensity consistent filtering approach designed to filter contractions and expansions in neighboring tissues separately so that anatomical boundaries are preserved (Studholme et al., 2003).

The type of linear registration (required as a pre-processing step before nonlinearly warping one scan onto the other) also affected the power to detect group differences in atrophic rates. Alignment of scans over time using scaling (9P linear registration) outperformed the rigid-body registration (6P), probably because the 9P registration can correct for scanner voxel size variations in large studies involving multiple sites, scanners and acquisition sequences. Prior longitudinal MRI studies found that using 9P scaling is advantageous for registration over time, even when significant cerebral atrophy is also occurring in the images (Paling et al., 2004). The concept of using 9P registration for serial MRI analysis was counterintuitive at first as one might suspect that important biological changes over time might be eliminated and incorrectly discounted, if scans were registered using 9P scaling. However, for studies of cerebral atrophy in dementia or other neurodegenerative diseases, the population of interest is relatively old and thus the skull size—one of the major driving forces in registration—is relatively stable. In TBM studies of early development (Gogtay et al., 2008; Hua et al., 2009), where there is substantial dynamic growth of the brain and skull, 6P registration may be more suitable, despite not correcting for possible geometric calibration errors due to scanner drift and effects of magnetic field re-shimming over time.

Some of the image correction steps in ADNI (e.g., gradwarp, and phantom-based image scaling) were not implemented in the study by Paling et al. (2004). Here we were able to achieve greater effect sizes using 9-parameter image registration instead of 6-parameter (rigid body) registration (Fig. 2). This suggests that standard image preprocessing pipeline in ADNI might not be sufficient to correct for all scanner variations, and applying 9P registration over time provided additional benefit. It is also worth noting that we did not apply brain mask prior to linear registration, unlike Paling et al. who delineated the brain region in their images using a semi-automated approach. Hence 9P scaling can be applied to longitudinal studies with and without nonbrain tissues, to correct for any remaining scanner voxel size variation that was not accounted for by image correction algorithms. Also, a very recent paper (Clarkson et al., 2009) suggests that 9-parameter registration was equally good or even superior to using phantom-based image corrections in the ADNI data specifically.

There is some debate as to whether it makes sense to compute minimal sample sizes from the changes in the patient group only, without taking into account the aging changes normally occurring in controls. Perhaps controversially, the consensus in the AD literature has been to base sample size estimates on the effect sizes for change computed in the patients only, even though such an approach can be biased in several respects. First, if changes that normally occur with normal aging are not considered, it is assumed that the goal of therapy is to reduce the patients' atrophic rate by 25%, even if a more achievable target would be to reduce the patient-control difference (which is smaller) by 25%. Second, if any apparent shrinkage of the brain occurred due to a systematic (nonrandom) error in scanner calibration at some acquisition sites, or due to phantom errors, geometric drifts over time, or unforeseen biases in image processing, then the cause of the error may not be easy to detect in the absence of a control group, but would be taken into account in the statistics if a control group was also used, as their images would presumably also be prone to the same drift effects. To avoid these issues in the current study, we did not rely solely on sample size estimates (n80 and n90)which were based on the patient groups only-but we also examined effect sizes in comparisons between patients and matched healthy controls. We computed CDFs from statistical maps of group differences between patients and controls, and we used FDR-based q-values to evaluate the power of different TBM designs. As seen in Figs. 1 and 5, the TBM designs offering best sample size estimates (e.g., sKL-MI S9L1 for MCI) may not always best discriminate patients from controls, indicated by inferior CDF ranking and lower q-value. Clearly the former depends solely on the effect sizes for the change in the patients, while the latter also depends on the effect sizes for change observed in controls.

The idea of implementing a statistically-defined ROI based on an independent training sample was advocated by Reiman et al. (Chen et al., 2009; Reiman et al., 2008) for analysis of positron emission tomography (PET) images. This is a variant of a strategy that has long been used in functional MRI studies, where it is common to limit the search region for statistical effects to circumscribed anatomical regions or regions generally implicated in past studies. These regions can be based on maps derived from statistically orthogonal (and therefore independent) contrasts computed from the same sets of images. Here we showed that a statistically-defined ROI can be used to improve the detection power by concentrating the signals that change

most in AD, as determined from a nonoverlapping sample of AD patients (Table 2).

The use of a statistically-based ROI was so much better than using an atlas-based ROI that it motivates further study of how best to use an ROI to optimize power. Different weights could be given to the component voxels, for example; rather than using a simple thresholding of the ROI to compute the mean value, the values in the ROI could be weighted using their effect sizes (e.g., t statistics derived from an independent sample). More sophisticated machine learning procedures could be used to select a subset of voxels in the ROI that are most informative when used jointly (for example using an adaptive boosting technique or a committee based learning method; see (Morra et al., 2009a) for a review of these methods). For example, we recently measured gray matter thickness at 64,000 cortical surface points in MRI data from 36 schizophrenia patients and 36 controls, and a machine learning algorithm (Sun et al., in press) picked a selection of points scattered across the cortex that optimally predicted disease status (patient or control). This achieved an excellent 86.1% classification rate when validated on independent, nonoverlapping, data. By distilling a whole image of features into those that are most predictive, dimension reduction is achieved; these features may also form a spatially distributed anatomical network revealing areas jointly involved in the disease. As the chosen features are all assessed simultaneously, changes in one region that are not necessarily strong enough on their own to be predictive are assessed in conjunction with all others, and systems of co-occurring changes can be identified (Zhang et al., 2008). This is the premise of several machine learning methods that have been applied to brain mapping, such as multivarate network analysis (Alexander et al., 2008), adaptive boosting, online learning, and their variants (see Morra et al., (2009a; 2009b), for a review).

We found that the hippocampus has a high rate of atrophy, and shows faster atrophy in MCI and AD than in controls when the full sample is analyzed (Fig. 3). However, in the stat-ROI (Fig. 4), which was defined in the training set of 22 AD subjects, the effect sizes for atrophic rates were great outside the hippocampus, in the rest of the temporal lobe. The effect size for detecting changes in the hippocampus can be negatively influence by the following three aspects. (1) The hippocampus is immediately adjacent to CSF voxels that are expanding, resulting in high variance and low statistical effect size at the boundaries. (2) Inherent variability in anatomy across subjects, which combined with the smallness of the structure, makes it difficult to align maps of change perfectly across subjects. (3) the effect size of stat-ROI is limited by the size of training set-only 22 AD patients were used to generate the stat-ROI. Adding more subjects to the training set will increase the effect size but at the cost of reducing the sample size of the testing set.

TBM has some specific advantages as a biomarker of disease progression in AD. The serial MRI protocol on which it is based tends to have high test/re-test reliability, and TBM can be used to assess, explore, or confirm the stability of serial MRI by computing changes across short intervals such as 2 weeks (Leow et al., 2006b), where biological changes are minimal. The effects of phantom-based corrections, image inhomogeneity corrections (Boyes et al., 2008), and other image processing steps on longitudinal stability can also be extensively explored, monitored, and optimized, as are effects of the field strength of the scanner (3 versus 1.5 T) (Ho et al., in press) and image reconstruction algorithms on statistical power.

In future studies, we plan to compare the power of several complementary MRI-derived measures, including volumes and maps of hippocampal and ventricular anatomy (Chou et al., 2009a; Morra et al., 2009; Morra et al., 2009c; Schuff et al., 2009), whole-brain and ventricular boundary shift integrals (Fox et al., 2001), automated volumetric parcellations (Fischl et al., 2002), voxel-based morphometry (Alexander et al., 2007; Ashburner and Friston, 2000; Good et al., 2001), abnormality scores derived from high-dimensional pattern

classification (Misra et al., 2009), other tensor-based morphometry methods (Studholme et al., 2005) and cortical thickness mapping (Aganj et al., 2009; Thompson et al., 2003; Thompson et al., 2004b) on the same subset of ADNI data. Additionally, longitudinal declines during the different pre-clinical and clinical stages of AD may not necessarily be linear. Thus, one will need to see the extent to which it affects power in more severely affected AD patients or over different trial durations. At the same time, one could develop optimal statistical ROIs tailored to the subject characteristics of future subjects (e.g., mild versus moderately severe AD and the time-line of the proposed study).

It is unlikely that one single biomarker will outperform all others in terms of its power to track change, correlate with clinical decline, or predict outcomes. Furthermore, different biomarkers may differ in the extent to which they are modified by a treatment, predict a treatment's clinical benefit, or are affected in a confounding way unrelated to the treatment's therapeutic effects (e.g., increased volume loss measured by brain boundary shift integral (BBSI) was found in antibody responders compared with nonresponders in the clinical trial of AN1792 immunotherapy (Fox et al., 2005)). Use of genetic or CSF biomarkers, as well as other imaging measures (from PET or DTI) will be advantageous. In the meantime however, structural MRI measures offer a dramatic reduction in sample sizes required for clinical trials (an 8-fold difference relative to the best clinical measure in this study). Future efforts will focus on combining measures to reduce this sample size still further.

### 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, GlaxoSmithKline, Merck and 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, SL, IY, AL, YC, AH, BG, AT, and PT performed the image analyses; CJ, MB, ER, DH, JK, NS, GA, and MW contributed substantially to the image and data 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.

### References

Aganj, I., Sapiro, G., Parikshak, N., Madsen, S.K., Thompson, P.M., 2009. Measurement of cortical thickness from MRI by minimum line integrals on soft-classified tissue. Hum Perin Many 2000 Eth 12 (algebraic methication about of print).

Hum. Brain Mapp. 2009 Feb. 13 [electronic publication ahead of print].
Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C., 2001. Spatial transformations of diffusion tensor magnetic resonance images. IEEE Trans. Med. Imaging 20, 1131–1139.

- Alexander, G.E., Chen, K., Reiman, E.M., Aschenbrenner, M., Merkley, T.L., Hanson, K.D., Dale, A.M., Bernstein, M.A., Kornak, J., Schuff, N., Fox, N.C., Thompson, P.M., Weiner, M.W., Jack Jr, C.R., 2007. Regional gray matter reductions in Alzheimer's dementia and annestic mild cognitive impairment: preliminary findings from the Alzheimer's Disease Neuroimaging Initiative using voxel-based morphometry. Alzheimer's Association International Conference on the Prevention of Dementia, Washington, DC.
- Alexander, G.E., Hanson, K.D., Chen, K., Reiman, E.M., Bernstein, M.A., Kornak, J., Schuff, N., Fox, N.C., Thompson, P.M., Weiner, M.W., Jack, C.R., 2008. Six month MRI gray matter declines in Alzheimer's dementia evaluated by voxel-based morphometry with multivariate network analysis: preliminary findings from the ADNI study. International Conference on Alzheimer's Disease, Chicago.
- Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. NeuroImage 11, 805–821.

Ashburner, J., Friston, K.J., 2003. Morphometry. Human Brain Function. Academic Press.

- Beckett, L., Harvey, D., Zhang, H., 2008, ADNI preliminary results, July 2008: comparative and longitudinal results. Presentation to Industry Scientific Advisory Board (ISAB) at International Conference on Alzheimer's Disease (ICAD), Chicago, IL.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc., B 57 (1), 289–300.

Berg, L., 1988. Clinical dementia rating (CDR). Psychopharmacol. Bull. 24, 637-639.

- Boyes, R.G., Gunter, J.L., Frost, C., Janke, A.L., Yeatman, T., Hill, D.L., Bernstein, M.A., Thompson, P.M., Weiner, M.W., Schuff, N., Alexander, G.E., Killiany, R.J., DeCarli, C., Jack, C.R., Fox, N.C., 2008. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. NeuroImage 39, 1752–1762.
- Braak, H., Braak, E., 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. (Berl.) 82, 239–259.
- Braskie, M.N., Klunder, A.D., Hayashi, K.M., Protas, H., Kepe, V., Miller, K.J., Huang, S.C., Barrio, J.R., Ercoli, L.M., Siddarth, P., Satyamurthy, N., Liu, J., Toga, A.W., Bookheimer, S.Y., Small, G.W., Thompson, P.M., 2008. Plaque and tangle imaging and cognition in normal aging and Alzheimer's disease. Neurobiol. Aging 2008 Nov. 10 [electronic publication ahead of print].
- Broit, C., 1981. Optimal registration of deformed images (PhD thesis). University of Pennsylvania.
- Brun, C.C., Lepore, N., Pennec, X., Chou, Y.Y., Lee, A.D., Barysheva, M., de Zubicaray, G.I., McMahon, K.L., Wright, M.J., Toga, A.W., Thompson, P.M., 2009a. A Lagrangian formulation for statistical fluid registration. In: Karl, W.C., Rosen, B., Brooks, D. (Eds.), Proceedings of the Sixth IEEE International Symposium on Biomedical Imaging (ISBI2009) Boston, MA, pp. 975–978; IEEE, Piscataway, NJ, 2009, to appear.
- Brun, C.C., Lepore, N., Pennec, X., Chou, Y.Y., Lee, A.D., de Zubicaray, G.I., McMahon, K.L., Wright, M.J., Toga, A.W., Thompson, P.M., 2009b. 3D registration in the Lagrangian framework—application to the genetics of brain shape. Medical Image Computing and Computer Assisted Intervention (MICCAI2009), London, UK.
- Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., Brammer, M.J., 1999. 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 18, 32–42.
- Carmichael, Ö.T., Thompson, P.M., Dutton, R.A., Lu, A., Lee, S.E., Lee, J.Y., Kuller, L.H., Lopez, O.L., Aizenstein, H.J., Meltzer, C.C., Liu, Y., Toga, A.W., Becker, J.T., 2006. Mapping ventricular changes related to dementia and mild cognitive impairment in a large community-based cohort. IEEE ISBI 315–318.
- Chen, K., Reschke, C., Lee, W., Napatkamon, A., Liu, X., Bandy, D., Langbaum, J., Alexander, G.E., Foster, N.L., Koeppe, R.A., Jagust, W.J., Weiner, M.W., Reiman, E.M., 2009. Crosssectional and longitudinal analyses of fluorodeoxyglucose positron emission tomography images from the Alzheimer's disease neuroimaging initiative. ADNI Data Presentations Meeting, Seattle, WA.
- Chiang, M.C., Dutton, R.A., Hayashi, K.M., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M., 2007a. 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. NeuroImage 34, 44–60.
- Chiang, M.C., Reiss, A.L., Lee, A.D., Bellugi, U., Galaburda, A.M., Korenberg, J.R., Mills, D.L., Toga, A.W., Thompson, P.M., 2007b. 3D pattern of brain abnormalities in Williams syndrome visualized using tensor-based morphometry. Neuroimage 36, 1096–1109.
- Chou, Y.Y., Lepore, N., de Zubicaray, G.I., Carmichael, O.T., Becker, J.T., Toga, A.W., Thompson, P.M., 2008. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. NeuroImage 40, 615–630.
- Chou, Y.Y., Lepore, N., Avedissian, C., Madsen, S.K., Parikshak, N., Hua, X., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Toga, A.W., Thompson, P.M., 2009a. Mapping correlations between ventricular expansion and CSF amyloid and tau biomarkers in 240 subjects with Alzheimer's disease, mild cognitive impairment and elderly controls. NeuroImage 46, 394–410.
- Chou, Y.Y., Leporé, N., Avedissian, C., Madsen, S.K., Barysheva, M., Hua, X., Jack Jr, C.R., Weiner, M.W., Toga, A.W., Thompson, P.M., and the Alzheimer's Disease Neuroimaging Initiative, 2009b. Mapping ventricular expansion and its clinical correlates in Alzheimer's disease and mild cognitive impairment using multi-atlas fluid image alignment. SPIE Medical Imaging, Lake Buena Vista, Florida, February 7–12.
- Chung, M.K., Worsley, K.J., Paus, T., Cherif, C., Collins, D.L., Giedd, J.N., Rapoport, J.L., Evans, A.C., 2001. A unified statistical approach to deformation-based morphometry. NeuroImage 14, 595–606.
- Clark, C.M., Karlawish, J.H., 2003. Alzheimer disease: current concepts and emerging diagnostic and therapeutic strategies. Ann. Intern. Med. 138, 400–410.
- Cockrell, J.R., Folstein, M.F., 1988. Mini-mental state examination (MMSE). Psychopharmacol. Bull. 24, 689–692.
- Clarkson, M.J., Ourselin, S., Nielsen, C., Leung, K.K., Barnes, J., Whitwell, J.L., Gunter, J.L., Hill, D.L., Weiner, M.W., Jack Jr., C.R., Fox, N.C., 2009. Comparison of phantom and registration scaling corrections using the ADNI cohort. Neuroimage [electronic publication online before print].

- Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.
- DeKosky, S.T., Marek, K., 2003. Looking backward to move forward: early detection of neurodegenerative disorders. Science 302, 830–834.
- Dickerson, B.C., Sperling, R.A., 2005. Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease. NeuroRx 2, 348–360.
- Ferrarini, L., Palm, W.M., Olofsen, H., van der Landen, R., van Buchem, M.A., Reiber, J.H., Admiraal-Behloul, F., 2008. Ventricular shape biomarkers for Alzheimer's disease in clinical MR images. Magn. Reson. Med. 59, 260–267.
- Fillard, P., Arsigny, V., Pennec, X., Thompson, P.M., Ayache, N., 2005. Extrapolation of sparse tensor fields: application to the modeling of brain variability. Information Processing in Medical Imaging (IPMI). Glenwood Springs, Colorado, pp. 644–652.
- Fillard, P., Arsigny, V., Pennec, X., Hayashi, K.M., Thompson, P.M., Ayache, N., 2007. Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. NeuroImage 34, 639–650.
- Fillit, H., 2008. Drug discovery and the prevention of Alzheimer's disease. Alzheimers Dement. 4, S26-28.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
- Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.
- Fox, N.C., Cousens, S., Scahill, R., Harvey, R.J., Rossor, M.N., 2000. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. Arch. Neurol. 57, 339–344.
- Fox, N.C., Crum, W.R., Scahill, R.I., Stevens, J.M., Janssen, J.C., Rossor, M.N., 2001. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. Lancet 358, 201–205.
- Fox, N.C., Black, R.S., Gilman, S., Rossor, M.N., Griffith, S.G., Jenkins, L., Koller, M., 2005. Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. Neurology 64, 1563–1572.
- Freeborough, P.A., Fox, N.C., 1998. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. J. Comput. Assist. Tomogr. 22, 838–843. Gee, J.C., 1999. On matching brain volumes. Pattern Recogn. 32, 99–111.
- Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.
- Goate, A., 2006. Segregation of a missense mutation in the amyloid beta-protein precursor gene with familial Alzheimer's disease. J. Alzheimers. Dis. 9, 341–347.
- Goate, A., Chartier-Harlin, M.C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., et al., 1991. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349, 704–706.
- Gogtay, N., Lu, A., Leow, A.D., Klunder, A.D., Lee, A.D., Chavez, A., Greenstein, D., Giedd, J.N., Toga, A.W., Rapoport, J.L., Thompson, P.M., 2008. Three-dimensional brain growth abnormalities in childhood-onset schizophrenia visualized by using tensor-based morphometry. Proc. Natl. Acad. Sci. U. S. A. 105, 15979–15984.
- Goldman, W.P., Price, J.L., Storandt, M., Grant, E.A., McKeel Jr., D.W., Rubin, E.H., Morris, J.C., 2001. Absence of cognitive impairment or decline in preclinical Alzheimer's disease. Neurology 56, 361–367.
- Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14, 21–36.
- Grenander, U., Miller, M.I., 1998. Computational anatomy: an emerging discipline. Q. Appl. Math. LVI, 617–694.
- Grundman, M., Sencakova, D., Jack Jr., C.R., Petersen, R.C., Kim, H.T., Schultz, A., Weiner, M.F., DeCarli, C., DeKosky, S.T., van Dyck, C., Thomas, R.G., Thal, L.J., 2002. Brain MRI hippocampal volume and prediction of clinical status in a mild cognitive impairment trial. J. Mol. Neurosci. 19, 23–27.
- Gunter, J., Bernstein, M., Borowski, B., Felmlee, J., Blezek, D., Mallozzi, R., 2006. Validation testing of the MRI calibration phantom for the Alzheimer's Disease Neuroimaging Initiative study. ISMRM 14th Scientific Meeting and Exhibition.
- Haier, R.J., Alkire, M.T., White, N.S., Uncapher, M.R., Head, E., Lott, I.T., Cotman, C.W., 2003. Temporal cortex hypermetabolism in Down syndrome prior to the onset of dementia. Neurology 61, 1673–1679.
- Halperin, I., Morelli, M., Korczyn, A.D., Youdim, M.B., Mandel, S.A., 2009. Biomarkers for evaluation of clinical efficacy of multipotential neuroprotective drugs for Alzheimer's and Parkinson's diseases. Neurotherapeutics 6, 128–140.
- Ho, A.J., Hua, X., Lee, S., Yanovsky, I., Leow, A.D., Gutman, B., Dinov, I.D., Toga, A.W., Jack, C.R., Jr., Bernstein, M.A., Reiman, E.M., Harvey, D., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., in press. Comparing 3 T and 1.5 T MRI for tracking AD progression with tensor-based morphometry. NeuroImage.
- Hua, X., Leow, A.D., Lee, S., Klunder, A.D., Toga, A.W., Lepore, N., Chou, Y.Y., Brun, C., Chiang, M.C., Barysheva, M., Jack Jr., C.R., Bernstein, M.A., Britson, P.J., Ward, C.P., Whitwell, J.L., Borowski, B., Fleisher, A.S., Fox, N.C., Boyes, R.G., Barnes, J., Harvey, D., Kornak, J., Schuff, N., Boreta, L., Alexander, G.E., Weiner, M.W., Thompson, P.M., Alzheimer's Disease Neuroimaging, I, 2008a. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage 41, 19–34.
- Hua, X., Leow, A.D., Parikshak, N., Lee, S., Chiang, M.C., Toga, A.W., Jack Jr., C.R., Weiner, M.W., Thompson, P.M., 2008b. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects. NeuroImage 43, 458–469.

- Hua, X., Leow, A.D., Levitt, J.G., Caplan, R., Thompson, P.M., Toga, A.W., 2009. Detecting brain growth patterns in normal children using tensor-based morphometry. Hum. Brain Mapp. 30, 209–219.
- Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A., Martin, R.L., 1982. A new clinical scale for the staging of dementia. Br. J. Psychiatry 140, 566–572.
- Ivinson, A.J., Lane, R., May, P.C., Hosford, D.A., Carrillo, M.C., Siemers, E.R., 2008. Partnership between academia and industry for drug discovery in Alzheimer's disease. Alzheimers. Dement. 4, 80–88.
- Jack Jr., C.R., Petersen, R.C., Xu, Y.C., O Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Waring, S.C., Tangalos, E.G., Kokmen, E., 1999. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. Neurology 52, 1397–1403.
- Jack Jr., C.R., Dickson, D.W., Parisi, J.E., Xu, Y.C., Cha, R.H., O Brien, P.C., Edland, S.D., Smith, G.E., Boeve, B.F., Tangalos, E.G., Kokmen, E., Petersen, R.C., 2002. Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. Neurology 58, 750–757.
- Jack Jr., C.R., Slomkowski, M., Gracon, S., Hoover, T.M., Felmlee, J.P., Stewart, K., Xu, Y., Shiung, M., O Brien, P.C., Cha, R., Knopman, D., Petersen, R.C., 2003. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. Neurology 60, 253–260.
- Jack Jr., C.R., Shiung, M.M., Gunter, J.L., O Brien, P.C., Weigand, S.D., Knopman, D.S., Boeve, B.F., Ivnik, R.J., Smith, G.E., Cha, R.H., Tangalos, E.G., Petersen, R.C., 2004. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurology 62, 591–600.
- Jack Jr., C.R., Shiung, M.M., Weigand, S.D., O Brien, P.C., Gunter, J.L., Boeve, B.F., Knopman, D.S., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Petersen, R.C., 2005. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. Neurology 65, 1227–1231.
- Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., J, L.W., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008a. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.
- Jack Jr., C.R., Petersen, R.C., Grundman, M., Jin, S., Gamst, A., Ward, C.P., Sencakova, D., Doody, R.S., Thal, L.J., 2008b. Longitudinal MRI findings from the vitamin E and donepezil treatment study for MCI. Neurobiol. Aging 29, 1285–1295.
- Jellinger, K.A., 2006. Alzheimer 100-highlights in the history of Alzheimer research. J. Neural. Transm. 113, 1603-1623.
- 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., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436–443.
- Leow, A., Huang, S.C., Geng, A., Becker, J.T., Davis, S., Toga, A.W., Thompson, P.M., 2005. Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties. Information Processing in Medical Imaging. Glenwood Springs, Colorado, USA, pp. 493–503.
- Leow, A.D., Chiang, M.C., Becker, J.T., Davis, S.W., Toga, A.W., Thompson, P.M., 2006a. Realizing unbiased deformation: a theoretical consideration. 9th IEEE Conference on Medical Image Computer and Computer Assisted Intervention (MICCAI), MICCAI2006 Workshop on Mathematical Foundations of Computational Anatomy, Copenhagen, Denmark.
- Leow, A.D., Klunder, A.D., Jack Jr., C.R., Toga, A.W., Dale, A.M., Bernstein, M.A., Britson, P.J., Gunter, J.L., Ward, C.P., Whitwell, J.L., Borowski, B.J., Fleisher, A.S., Fox, N.C., Harvey, D., Kornak, J., Schuff, N., Studholme, C., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2006b. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. NeuroImage 31, 627–640.
- Leow, A.D., Yanovsky, I., Chiang, M.C., Lee, A.D., Klunder, A.D., Lu, A., Becker, J.T., Davis, S.W., Toga, A.W., Thompson, P.M., 2007. Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration. IEEE Trans. Med. Imaging 26, 822–832.
- Leow, A.D., Yanovsky, I., Parikshak, N., Hua, X., Lee, S., Toga, A.W., Jack Jr., C.R., Bernstein, M.A., Britson, P.J., Gunter, J.L., Ward, C.P., Borowski, B., Shaw, L.M., Trojanowski, J.Q., Fleisher, A.S., Harvey, D., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2009. Alzheimer's disease neuroimaging initiative: a one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. NeuroImage 45, 645–655.
- Lepore, N., Brun, C., Chou, Y.Y., Chiang, M.C., Dutton, R.A., Hayashi, K.M., Luders, E., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M., 2008. Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. IEEE Trans. Med. Imaging 27, 129–141.
- 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., 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos. Trans. R. Soc. Lond., B. Biol. Sci. 356, 1293–1322.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. 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 34, 939–944.
- Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. NeuroImage 44, 1415–1422.
- Morra, J.H., Tu, Z., Apostolova, L.C., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Hua, X., Toga, A.W., Jack Jr., C.R., Weiner, M.W., Thompson, P.M., 2008. Validation

of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. NeuroImage 43, 59–68.

- Morra, J., Tu, Z., Toga, A.W., Thompson, P.M., 2009a. Machine learning for brain image segmentation. In: Gonzalez, F., Romero, E. (Eds.), Tutorial Chapter in Textbook: Biomedical Image Analysis and Machine Learning Technologies, to appear: 2009.
- Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Hua, X., Toga, A.W., Jack Jr., C.R., Schuff, N., Weiner, M.W., Thompson, P.M., 2009b. Automated 3D mapping of hippocampal atrophy and its clinical correlates in 400 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Hum. Brain Mapp.
- Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Toga, A.W., Jack Jr., C.R., Schuff, N., Weiner, M.W., Thompson, P.M., 2009c. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. NeuroImage 45, S3–15.
- Morris, J.C., 1993. The clinical dementia rating (CDR): current version and scoring rules. Neurology 43, 2412–2414.
- Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005a. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin. N. Am. 15, 869–877 xi–xii.
- Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005b. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement. 1, 55–66.
- Mueller, S.G., Schuff, N., Weiner, M.W., 2006. Evaluation of treatment effects in Alzheimer's and other neurodegenerative diseases by MRI and MRS. NMR Biomed. 19, 655–668.
- Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.
- Paling, S.M., Williams, E.D., Barber, R., Burton, E.J., Crum, W.R., Fox, N.C., O Brien, J.T., 2004. The application of serial MRI analysis techniques to the study of cerebral atrophy in late-onset dementia. Med. Image Anal. 8, 69–79.
- Petersen, R.C., 2000. Aging, mild cognitive impairment, and Alzheimer's disease. Neurol. Clin. 18, 789–806.
- Petersen, R.C., Smith, G.E., Ivnik, R.J., Kokmen, E., Tangalos, E.G., 1994. Memory function in very early Alzheimer's disease. Neurology 44, 867–872.
- Petersen, R.C., Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., Ritchie, K., Rossor, M., Thal, L., Winblad, B., 2001. Current concepts in mild cognitive impairment. Arch. Neurol. 58, 1985–1992.
- Price, J.L., Morris, J.C., 1999. Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. Ann. Neurol. 45, 358–368.
- Reiman, E.M., Langbaum, J.B.S., in press. Brain imaging in the evaluation of putative Alzheimer's disease slowing, risk-reducing and prevention therapies. In: Jagust, W., D'Esposito, M. (Eds.), Imaging and the Aging Brain. Oxford University Press, New York.
- Reiman, E.M., Caselli, R.J., Yun, L.S., Chen, K., Bandy, D., Minoshima, S., Thibodeau, S.N., Osborne, D., 1996. Preclinical evidence of Alzheimer's disease in persons homozygous for the epsilon 4 allele for apolipoprotein E. N. Engl. J. Med. 334, 752–758.
- Reiman, E.M., Caselli, R.J., Chen, K., Alexander, G.E., Bandy, D., Frost, J., 2001. Declining brain activity in cognitively normal apolipoprotein E epsilon 4 heterozygotes: a foundation for using positron emission tomography to efficiently test treatments to prevent Alzheimer's disease. Proc. Natl. Acad. Sci. U. S. A. 98, 3334–3339.
- Reiman, E.M., Chen, K., Alexander, G.E., Caselli, R.J., Bandy, D., Osborne, D., Saunders, A. M., Hardy, J., 2004. Functional brain abnormalities in young adults at genetic risk for late-onset Alzheimer's dementia. Proc. Natl. Acad. Sci. U. S. A. 101, 284–289.
- Reiman, E.M., Chen, K., Ayutyanont, N., Bandy, D., Lee, W., Reschke, C., Alexander, G.E., Weiner, M.W., Koeppe, R.A., Jagust, W.J., 2008. Twelve-Month Cerebral Metabolic Declines in Probable Alzheimer's Disease and Amnestic Mild Cognitive Impairment: Preliminary Findings from the Alzheimer's Disease Neuroimaging Initiative. ICAD, Chicago, Illinois.
- Riddle, W.R., Li, R., Fitzpatrick, J.M., DonLevy, S.C., Dawant, B.M., Price, R.R., 2004. Characterizing changes in MR images with color-coded Jacobians. Magn. Reson. Imaging 22, 769–777.
- Rosen, W.G., Mohs, R.C., Davis, K.L., 1984. A new rating scale for Alzheimer's disease. Am. J. Psychiatry 141, 1356–1364.

Rosner, B., 1990. Fundamentals of Biostatistics. PWS-Kent Publishing Company, Boston.

- Scheltens, P., Fox, N., Barkhof, F., De Carli, C., 2002. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. Lancet Neurol. 1, 13–21.
- Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L.M., Trojanowski, J.Q., Thompson, P.M., Jack Jr., C.R., Weiner, M.W., 2009. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132 (Pt 4), 1067–1077.
- Shattuck, D.W., Leahy, R.M., 2002. BrainSuite: an automated cortical surface identification tool. Med. Image Anal. 6, 129–142.
- Shaw, L.M., Korecka, M., Clark, C.M., Lee, V.M., Trojanowski, J.Q., 2007. Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics. Nat. Rev. Drug Discov. 6, 295–303.
- Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging 21, 1421–1439.
- Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.
- Storey, J.D., 2002. A direct approach to false discovery rates. J. R. Stat. Soc., B 64 (Pt. 3), 479–498.

- Studholme, C., Cardenas, V., Schuff, N., Rosen, H., Miller, B., Weiner, M., 2001. Detecting spatially consistent structural differences in Alzheimer's and fronto temporal dementia using deformation morphometry. MICCAI 41–48.
- Studholme, C., Cardenas, V., Maudsley, A., Weiner, M., 2003. An intensity consistent filtering approach to the analysis of deformation tensor derived maps of brain shape. NeuroImage 19, 1638–1649.
- Studholme, C., Drapaca, C., Cardenas, V., 2005. Intensity robust viscous fluid deformation based morphometry using regionally adapted mutual information. Conf. Proc. IEEE Eng. Med. Biol. Soc. 1, 470–473.
- Sun, D., van Erp, T.G.M., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K., Toga, A.W., Cannon, T.D., in press. Elucidating an MRI-based biomarker for psychosis: classification using probabilistic brain atlas and machine learning algorithms. Biological Psychiatry.
- Thompson, P.M., Giedd, J.N., Woods, R.P., MacDonald, D., Evans, A.C., Toga, A.W., 2000. Growth patterns in the developing brain detected by using continuum mechanical tensor maps. Nature 404, 190–193.
- Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W., 2003. Dynamics of gray matter loss in Alzheimer's disease. J. Neurosci. 23, 994–1005.
- Thompson, P.M., Hayashi, K.M., De Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M., Toga, A.W., 2004a. Mapping hippocampal and ventricular change in Alzheimer disease. NeuroImage 22, 1754–1766.
- Thompson, P.M., Hayashi, K.M., Sowell, E.R., Gogtay, N., Giedd, J.N., Rapoport, J.L., de Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Doddrell, D.M., Wang, Y., van Erp, T.G., Cannon, T.D., Toga, A.W., 2004b. Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. NeuroImage 23 (Suppl. 1), S2–S18.
- Toga, A.W., 1999. Brain Warping, 1st ed. Academic Press, San Diego.

- Tosun, D., Schuff, N., Raptentsetsang, S., Truran-Sacrey, D., Weiner, M.W., 2009. Regional Rates of Brain Atrophy in Correlation to Cognitive Decline and Biomarker Concentrations. Seattle, WA, ADNI Data Presentations Meeting.
- Wimo, A., Jonsson, L., Winblad, B., 2006. An estimate of the worldwide prevalence and direct costs of dementia in 2003. Dement. Geriatr. Cogn. Disord. 21, 175–181.
- Yanovsky, I., Chiang, M.C., Thompson, P.M., Klunder, A.D., Becker, J.T., Davis, S.W., Toga, A.W., Leow, A.D., 2007a. Quantifying deformation using information theory: the logunbiased nonlinear registration. IEEE Int. Symp. Biomed. Imaging (ISBI) 13–16.
- Yanovsky, I., Thompson, P.M., Osher, S.J., Leow, A.D., 2007b. Topology preserving logunbiased nonlinear image registration: theory and implementation. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).
- Yanovsky, I., Thompson, P., Osher, S., Leow, A.D., 2008a. Asymmetric and symmetric unbiased image registration: statistical assessment of performance. IEEE Computer Society Workshop On Mathematical Methods in Biomedical Image Analysis. 1, 1–8.
- Yanovsky, I., Thompson, P.M., Osher, S.J., Hua, X., Shattuck, D.W., Toga, A.W., Leow, A.D., 2008b. Validating unbiased registration on longitudinal MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI). IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1091–1094.
- Yanovsky, I., Leow, A.D., Lee, S., Osher, S.J., Thompson, P.M., 2009. Comparing registration methods for mapping brain change using tensor-based morphometry. Medical Image Analysis. doi:10.1016/j.media.2009.06.002.
- Zhang, H., Wu, T., Bae, M., Reiman, E.M., Alexander, G.E., Jack Jr., C.R., Thompson, P.M., Chen, K., 2008. Use of the Support Vector Machine and Sensitivity of an AD-Related Region-of-Interest Gray Matter Classifier in Identifying Amnestic MCI Subjects Who Convert to AD: Preliminary Findings from the AD Neuroimaging Initiative. Chicago, International Conference on Alzheimer's Disease.