

An automated algorithm for the computation of brain volume change from sequential MRIs using an iterative principal component analysis and its evaluation for the assessment of whole-brain atrophy rates in patients with probable Alzheimer's disease[☆]

Kewei Chen,^{a,b,c,d,*} Eric M. Reiman,^{a,d,e} Gene E. Alexander,^{d,f} Daniel Bandy,^{a,d} Rosemary Renaut,^{b,d} William R. Crum,^g Nick C. Fox,^h and Martin N. Rossor^h

^aPositron Emission Tomography Center, Banner Good Samaritan Medical Center, Phoenix, AZ 85006, USA

^bDepartment of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804, USA

^cDepartment of Radiology, University of Arizona, Tucson, AZ 85724, USA

^dArizona Alzheimer's Research Center and the Alzheimer's Disease Core Center, Phoenix, AZ 85006, USA

^eDepartment Psychiatry, University of Arizona, Phoenix, AZ 85006, USA

^fDepartment of Psychology, Arizona State University, Tempe, AZ 85287, USA

^gDivision of Imaging Sciences, The Guy's, King's and St. Thomas' School of Medicine, Kings College London, London, UK

^hDementia Research Group, National Hospital for Neurology and Neurosurgery, London, UK

Received 24 June 2003; revised 15 October 2003; accepted 8 January 2004

This article introduces an automated method for the computation of changes in brain volume from sequential magnetic resonance images (MRIs) using an iterative principal component analysis (IPCA) and demonstrates its ability to characterize whole-brain atrophy rates in patients with Alzheimer's disease (AD). The IPCA considers the voxel intensity pairs from coregistered MRIs and identifies those pairs a sufficiently large distance away from the iteratively determined PCA major axis. Analyses of simulated and real MRI data support the underlying assumption of a linear relationship in paired voxel intensities, identify an outlier distance threshold that optimizes the trade-off between sensitivity and specificity in the detection of small volume changes while accounting for global intensity changes, and demonstrate an ability to detect changes as small as 0.04% of brain volume without confounding effects of between-scan shifts in voxel intensity. In eight patients with probable AD and eight age-matched normal control subjects, the IPCA was comparable to the established but partly manual digital subtraction (DS) method in characterizing annual rates of whole-brain atrophy: resulting rates were correlated (Spearman rank correlation = 0.94, $P < 0.0005$) and comparable in distinguishing probable AD from normal aging (IPCA-detected atrophy rates: $2.17 \pm 0.52\%$ per year in the patients vs. $0.41 \pm 0.22\%$ per year in the controls [Wilcoxon–Mann–Whitney test $P = 7.8 \times 10^{-4}$]; DS-detected atrophy rates: $3.51 \pm 1.31\%$ per year in the patients vs. $0.48 \pm 0.29\%$ per year in the controls [$P = 7.8 \times 10^{-4}$]). The IPCA could be used in tracking the progression of AD, evaluating

the disease-modifying effects of putative treatments, and investigating the course of other normal and pathological changes in brain morphology.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Iterative principal component analysis; MRI; Alzheimer's disease; Atrophy

Introduction

Patients with probable Alzheimer's dementia (AD) have an abnormally high rate of whole-brain atrophy, a potential indicator of disease progression, which might help in the early detection and tracking of Alzheimer's disease and in assessing the disease-modifying efficacy of candidate treatments. In their original studies, Fox et al. developed a semiautomated digital subtraction (DS, also known as the brain boundary shift integral [BBSI]) method for the measurement of whole-brain atrophy in individual human subjects using sequential magnetic resonance images (MRIs) (Freeborough and Fox, 1997; Fox and Freeborough, 1997). They demonstrated significantly higher rates of whole-brain atrophy in patients with probable AD than those associated with normal aging (Fox et al., 1996), and significantly higher rates of whole-brain atrophy shortly before the onset of dementia in persons at risk of AD (Fox et al., 1999, 2001). They computed the statistical power of this method to test the efficacy of candidate treatments to retard the progression of the disease (Fox et al., 2000).

While the DS method provides a promising marker of disease progression for the study of AD and the evaluation of disease-

[☆] The method described in this article was initially presented in part at the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences, Las Vegas, NV, June 26–29, 2000.

* Corresponding author. PET Center, Banner Good Samaritan Regional Medical Center, 1111 E. McDowell Road, Phoenix, AZ 85006. Fax: +1-602-239-3073.

E-mail address: kchen@math.la.asu.edu (K. Chen).

Available online on ScienceDirect (www.sciencedirect.com.)

modifying candidate treatments, it has a couple of limitations. First, the DS method requires some manual effort to identify voxels corresponding to brain tissue. (This procedure is used to normalize each of the individual's MRIs for differences in mean brain signal intensity.) This manual effort could limit implementation of the DS method in other laboratories and, in the absence of a rigorous quality adherence procedure, it could limit the comparability of measurements among different raters and findings among different laboratories. Second, as discussed below, there may be instances in which the DS method is vulnerable to the potentially confounding effects of between-scan differences in voxel intensity (called voxel-intensity shift) that are unrelated to changes in brain volume and cannot be corrected using image normalization for mean brain signal intensity.

In this article, we introduce a simple and automated method for the measurement of changes in brain volume from an individual's sequential MRIs using an iterative principal component analysis (IPCA). This method uses the widely available neuroimaging software package SPM99 (<http://www.fil.ion.ucl.ac.uk/spm/>) for its preprocessing steps, capitalizing on this package's ability to coregister images (Friston et al., 1995), correct them for image inhomogeneities (Ashburner and Friston, 1997), and provide baseline measurements of whole-brain or intracranial volume (to compute changes in brain volume as a percentage of the baseline volume). Simulated and real MRI data pairs were used to (i) validate this method's underlying assumption of a linear relationship in paired voxel intensities; (ii) demonstrate its ability to measure changes in brain volume even in the presence of nonzero voxel intensity shift, which could confound the results obtained by the DS method; (iii) identify "a paired voxel-intensity threshold distance" that optimizes the trade-off between sensitivity and specificity in the detection of small volume changes; and (iv) demonstrate its ability to detect changes as small as 0.04% of baseline brain volume. Sequential MRIs acquired approximately 1 year apart in eight patients with probable AD and eight age-matched normal control subjects were used to (v) provide evidence that the IPCA method is comparable to the well-established DS method in its ability to characterize and contrast rates of whole-brain atrophy in these two subject groups and (vi) investigate the IPCA's ability to characterize abnormally high rates of whole-brain atrophy in patients with probable AD using any combination of SPM99 image preprocessing settings (including three different image-alignment techniques, three levels of inhomogeneity correction, and two ways to characterize baseline whole-brain or intracranial volume).

Methods

Subjects and data

(a) Baseline and follow-up MRIs from 23 cognitively, neurologically, and psychiatrically normal human subjects in Arizona were used to help validate the IPCA method's underlying assumption of a linear relationship in paired voxel intensities, demonstrate its ability to measure changes in brain volume even in the presence of voxel-intensity shift, identify "a paired voxel-intensity threshold distance" that optimizes the trade-off between sensitivity and specificity in the detection of small volume changes, and demonstrate its ability to detect small changes in brain volume. The subjects included 10 males and 13 females,

had a mean age of 48 (range 34–65), and were studied under guidelines approved by Institutional Review Board at Banner Good Samaritan Medical Center (Phoenix, AZ, USA). Thirteen of the subjects had their baseline and follow-up MRI scans performed during the same day, were removed from the scanner and repositioned between scans, and had a between-scan interval of approximately 30 min. The other 10 subjects had a between-scan interval of approximately 2 years. The baseline and follow-up MRIs were acquired on a 1.5-T Signa system (General Electric, Milwaukee, WI, USA) using a volumetric T_1 -weighted pulse sequence (radio-frequency-spoiled gradient recall acquisition in the steady state [SPGR], TR = 33, TE = 5, $\alpha = 30^\circ$, NEX = 1, field of view = 24 cm, imaging matrix = 256×192 , slice thickness = 1.5 mm, and scan time = 13:36 min). The MRI data set consisted of 124 contiguous horizontal slices with in-plane voxel dimensions of 0.94×1.25 mm.

(b) Four sets of simulated MRI data sets from a normal brain phantom, each seeded with a different random number generator, two with 3% and two with 5% measurement noise were used for the same purposes noted above. These data sets were downloaded from <http://www.bic.mni.mcgill.ca/brainweb> (Collins et al., 1998; Kwan et al., 1999), were generated using a volumetric pulse sequence (TR = 18, TE = 10, slice thickness = 1 mm), and consisted of 181 contiguous horizontal sections.

(c) Baseline and 1-year follow-up MRIs from eight patients who satisfied previously published criteria for probable AD (McKhann et al., 1984) and eight age-matched cognitively normal control subjects in London were used to compare the IPCA method to the DS method in the ability to characterize and contrast rates of whole-brain atrophy and to investigate the effects of different combinations of SPM99 image preprocessing settings on the IPCA's ability to characterize abnormally high rates of whole-brain atrophy in patients with probable AD. The patients were 54 ± 8 (mean \pm SD) years of age, included three males and five females, had scores of 20.7 ± 7.2 on the Mini-Mental State Examination (MMSE; Folstein et al., 1975) at the time of their baseline scans, and had an interval of 1.1 ± 0.5 years between their MRIs; and the normal control subjects were also 54 ± 8 years of age, included two males and six females, had scores of 29.5 ± 0.8 on the MMSE, and had an interval of 1.1 ± 0.1 years between their MRIs. All of these subjects provided their informed consent and were studied under guidelines approved by the Local Research Ethics Committee at the National Hospital for Neurology and Neurosurgery in London. The baseline and follow-up MRIs were acquired on a 1.5-T Signa system (General Electric) using a volumetric T_1 -weighted pulse sequence (radio-frequency-spoiled gradient recall acquisition in the steady state [SPGR], TR = 35, TE = 5, $\alpha = 35^\circ$, NEX = 1, field of view = $24 \times 24 \times 19.2$ cm, imaging matrix = 256×128 , and slice thickness = 1.5 mm). The MRI data set consisted of 124 contiguous coronal sections with in-plane voxel dimensions of 0.9375×0.9375 mm.

BBSI (digital subtraction) procedure

The DS method, also known as the BBSI method, was described previously (Fox and Freeborough, 1997) and is summarized here: (i) An automated algorithm coregisters the individual's follow-up MRI to his or her baseline MRI. (ii) Using a semi-automated, interactive method, and manual editing to classify brain volumes, voxel intensities in the two images are normalized by dividing the signal intensity in each brain voxel by the mean signal

intensity over the brain volume. (iii) Low- and high-intensity thresholds, I_1 and I_2 , respectively, are used to distinguish voxels outside the cortical brain surface from those within it. (Voxel values that are less than I_1 or greater than I_2 are reassigned to I_1 or I_2 , respectively.) (iv) The change in brain volume is then calculated as the integrated intensity differences between the baseline and follow-up image through the brain boundary relative to $I_2 - I_1$. When Fox and Freeborough (1997) validated the procedure and applied it to the study of patients with probable AD and age-matched controls, values of 0.25 and 0.75 were assigned to parameters I_1 and I_2 , respectively.

IPCA and image preprocessing procedure

A flow diagram of the IPCA method is given in Appendix A. Let x_i and y_i be the voxel intensities at brain location i in the baseline MRI₁ and coregistered follow-up MRI₂, respectively ($i = 1, \dots, N$, where N is the number of brain voxels). Each of the (x_i, y_i) pairs can be plotted as a point in an x - y coordinate system. Since most of voxels in the two MRIs are expected to correspond to the same kind of brain tissue (gray matter, white matter, or cerebrospinal fluid [CSF] respectively), most of the (x_i, y_i) points will be scattered around a ray—a straight line (the IPCA *linearity assumption*) with a positive slope. We refer to instances in which this ray does not pass through the origin in the x - y coordinate system [i.e., it does not have a zero intercept] as a linear *voxel-intensity shift* from the baseline MRI to the follow-up image. Whereas most of (x_i, y_i) points distributed around this line are encompassed within a relatively narrow band, those points that are a sufficient (i.e., *threshold*) distance away from the line may be considered *outliers*, reflecting between-scan differences in voxel intensity.

Outliers may reflect between-scan differences in the brain tissue type (e.g., a voxel corresponding to gray matter in MRI₁ and CSF in MRI₂ in cases of brain atrophy), changes in white matter intensity, or noise. As indicated below, we have identified an outlier threshold that optimizes the trade-off between sensitivity and specificity in the detection of small volume changes (i.e., one that permits us to detect relatively subtle changes in brain volume while rejecting false signals and noise). Whereas outliers above the ray in the x - y coordinate system represent *tissue gain* (Fox et al., 1996 have referred to these outliers as “tissue volume gain or relocation”; Smith et al., 2002 have referred to them as “tissue growth”), outliers below the ray represent *tissue loss*.

The IPCA automatically characterizes the ray, identifies significant outliers, and computes between-scan changes in brain volume. While the ray may be defined as the major PCA axis—the eigenvector associated with the largest eigenvalue for the 2×2 covariance matrix of multiple (x_i, y_i) pairs—it is important to determine this ray free from the potentially confounding (i.e., asymmetric) effect of outliers. By applying the PCA to the (x_i, y_i) points in an iterative manner, the IPCA successively removes outliers in determining the major axis direction (see Appendix A for details). Once the ray is determined, outliers can be identified as those that are a sufficiently large projected distance away from the ray. Utilizing a Gaussian probability density function, the projected distance between each (x, y) pair and the ray may be expressed in terms of standard deviations and p values. For instance, an (x, y) pair 3.291 standard deviation away from the ray would correspond to $P = 0.0005$. Below, we will express the outlier threshold in terms of P values. As described in Results, an “outlier threshold” of $P =$

0.0005 optimized the trade-off between sensitivity and specificity in detecting significant outliers from the coregistered human MRIs, thus permitting the detection of relatively subtle changes in brain volume while rejecting false signals and noise. Following a common convention (Fox et al., 1996), we express brain volume change as $I_{\%} = ((N_a - N_g) / N) 100\%$, where N_a and N_g are the number of voxels reflecting significant atrophy or gain, respectively, using a specified outlier threshold, and N is the total number of brain voxels. In the assessment of brain atrophy, the volume change index accounts for any tissue volume gain due to either image-misregistration (which is minimal as the coregistration accuracy is up to within a fraction of a millimeter (Friston et al., 1995) or localized displacement of brain tissue (Fox et al., 1996).

We have incorporated SPM99 into the IPCA software package to automatically coregister (Friston et al., 1995) the follow-up MRI to the baseline image (with or without three additional scaling procedures to account for a between-scan drift in voxel geometry (i.e., a between-image change in voxel dimensions that may occur due to gradient calibration drift or variable local field distortions; Freeborough et al., 1996), correct them for image inhomogeneities (Ashburner and Friston, 1997), and provide baseline measurements of whole-brain or intracranial volume (to compute changes in brain volume as a percentage of baseline measurements). The entire procedure permits researchers to blindly review the images and provide online commenting at the beginning and end of each step as needed.

Evaluating the IPCA method

Simulated and real MRI data pairs were used to (i) validate the IPCA method’s underlying assumption of a linear relationship in paired voxel intensities, characterize a between-MRI shift in voxel intensity, demonstrate the need for the PCA to be performed iteratively, and confirm its ability to measure changes in brain volume even in the presence of nonzero voxel-intensity shift, which could confound the results obtained by the DS method; (ii) identify a “voxel-intensity outlier threshold” that optimizes the trade-off between sensitivity and specificity in the detection of small volume changes, and demonstrate its ability to detect changes as small as 0.04% of baseline brain volume while rejecting false signals and noise; (iii) compare the IPCA to the DS method in its ability to characterize and contrast rates of whole-brain atrophy in patients with probable AD and age-matched normal control subjects; and (iv) characterize the effects of different combinations of SPM99 image-preprocessing settings on IPCA-computed rates of whole-brain atrophy in these subjects.

(a) To demonstrate a linear relationship in the coregistered images’ paired voxel intensities, all of the (x_i, y_i) pairs were plotted for each of the 23 normal volunteers studied in Arizona, the four simulated MRI data pairs, and the eight patients with probable AD and age-matched eight normal control subjects studied in London; and the PCA and linear regression analysis were each used to characterize the relationship among the paired voxel intensities. The paired voxel-intensity linearity assumption was examined using correlation coefficients, P values obtained in the regression analysis, and visual inspection. The voxel-intensity shift was defined by the intercepts. Since the DS method assumes that there is no between-scan shift in voxel intensity, we examined the voxel intensity shift in the 13 subjects with no changes in brain volume (i.e., their baseline and follow-up images were

acquired 30-min apart). To consider the potentially confounding effects of a voxel-intensity shift on DS measurements of between-scan differences in regional brain volume, we mapped DS-computed changes in brain volume onto the baseline images in subjects whose voxel intensity shift exceeded 20% of the maximal image signal intensity.

(b) To identify an outlier threshold that optimizes the ability to detect subtle changes in brain volume while rejecting false signals and noise, Monte Carlo simulations were performed with and without simulated atrophy in the follow-up image using MRI pairs from the simulated brain phantom and from the 13 normal volunteers who had an interval of 30 min between their baseline and follow-up scans, as described in the next two paragraphs. The optimal outlier threshold was defined as the minimum projected distance between a candidate outlier and the ray (in terms of P value) needed to detect the smallest change in simulated brain atrophy while most successfully rejecting false signals and noise.

To evaluate the IPCA's ability to reject false signals and noise in the assessment of between-scan differences in brain volume, the IPCA was applied to the coregistered images from the simulated brain phantom and the 13 normal volunteers using different outlier thresholds. Since the coregistered images should not have any true between-scan differences in brain volume, any changes in brain volume detected by the IPCA (the index $I_{\%}$ defined above) would reflect false signals or noise, which is expressed as $I_{\%false}$.

To evaluate the IPCA's detection sensitivity to small between-scan differences in brain volume, known amounts of atrophy were artificially introduced in the follow-up MRI and the IPCA was applied to the coregistered images from the simulated brain phantom and 13 normal volunteers using different outlier thresholds and different levels of noise in the simulated region of atrophy (ROA). ROAs were simulated in the coregistered follow-up image using the following procedures: (i) Gray matter was identified in each follow-up image. (The gray matter binary index map downloaded from the brain web was used to identify gray matter in each of the simulated MRIs; SPM99 brain tissue segmentation routine and a gray matter probability of at least 0.80 were used to identify gray matter in each of the human MRIs.) (ii) ROAs were alternatively defined as the collection of voxels representing the outer and inner gray matter boundaries or as a more circumscribed, manually defined brain region (e.g., in the parietal lobe). (iii) The mean and standard deviation (SD) of CSF voxel intensity was estimated from all of the follow-up MRI voxels in CSF space. (The CSF binary index map downloaded from the brain web was used to identify CSF in each of the simulated MRIs; the CSF probability map generated by SPM segmentation routine was used to identify CSF in each of the human MRIs.) (iv) Original voxel intensities within the ROA in the coregistered follow-up MRI were replaced with Gaussian random numbers with the CSF mean intensity and different percentages of the CSF SD (thus simulating different levels of uncertainty (i.e., noise) in ROA intensity values). (v) IPCA-detected ROAs and the simulated ROAs are superimposed onto baseline MRI space, thus providing information about the IPCA's ability to detect known atrophy, which is expressed as $I_{\%true}$ numerically.

(c) To compare the IPCA to the DS method in the ability to characterize and contrast rates of whole-brain atrophy associated with AD from those associated with normal aging, the IPCA (using $P = 0.0005$ as the optimal outlier detection threshold, determined as described above) and the DS method were each used to measure the annual rate of whole-brain atrophy in the eight patients with

probable AD and the eight age-matched normal control subjects who had been studied in London. For this analysis (and for the simulation study, described above), we used our default combination of SPM99 image-preprocessing settings: (i) the follow-up MRI was coregistered to the baseline image using a nine-parameter affine transformation (three parameters each for rotation, linear translation, and scaling, respectively); (ii) images were corrected for inhomogeneities with moderate levels of regularization of the modulation field smoothness; and (iii) brain volume in the baseline image was automatically defined as the brain mask extracted by SPM99, followed by the global erosion and deletion of isolated regions outside of the cortical surface.

(d) Finally, we investigated the IPCA's ability to characterize abnormally high rates of whole-brain atrophy in the patients with probable AD using 18 combinations of image preprocessing settings. The IPCA was repeatedly used to compute whole-brain atrophy in the eight patients with probable AD and the eight age-matched normal control subjects using every combination of three different image-alignment techniques, three levels of inhomogeneity correction, and two options for characterizing baseline whole-brain or intracranial volume. (i) The three image-alignment techniques included the procedure originally used in conjunction with the DS method (Freeborough et al., 1996; Fox and Freeborough, 1997); the alignment algorithm in SPM99 (Friston et al., 1995) that includes three rotation and three linear translation parameters; and an alternative alignment algorithm in SPM99 that also includes three scaling parameters to correct images for any potential between-session drift in voxel dimensions. (ii) The three levels of inhomogeneity correction (using multiplicative modulation of the voxel intensity by a scalar field), all of which are available in SPM99 included no inhomogeneity correction, correction with moderate levels of regularization of the modulation field smoothness, and correction with heavy levels of regularization of the modulation field smoothness (Ashburner and Friston, 1997). (iii) The two options for characterizing baseline whole-brain or intracranial volume included the partly manual technique (Fox et al., 1996) originally used to define baseline intracranial volume in conjunction with the DS method and the automated SPM99 whole-brain extraction algorithm, which we modified slightly to automatically delete extracerebral regions as noted above. Thus, we evaluated the effects of 18 combinations of IPCA preprocessing settings on IPCA-computed rates of whole-brain atrophy, including the combination originally used by Fox et al. (i.e., their alignment algorithm with a correction for voxel drift, no correction for inhomogeneities in MRI voxel intensity, and their partly manual method for the characterization of intracranial volume in the baseline image) and our combination of default settings, noted above. The general linear model was used to examine the effects of the different combinations of SPM99 image preprocessing settings on the ability of the IPCA to distinguish whole-brain atrophy rates associated with probable AD from those associated with normal aging.

Results

(a) This study (i) validated IPCA's underlying assumption of a linear relationship in paired voxel intensities, (ii) identified the presence of the between-scan shift in voxel intensity in the human brain MRIs, and (iii) demonstrated the need to perform the PCA iteratively. Graphically plotting paired voxel intensities

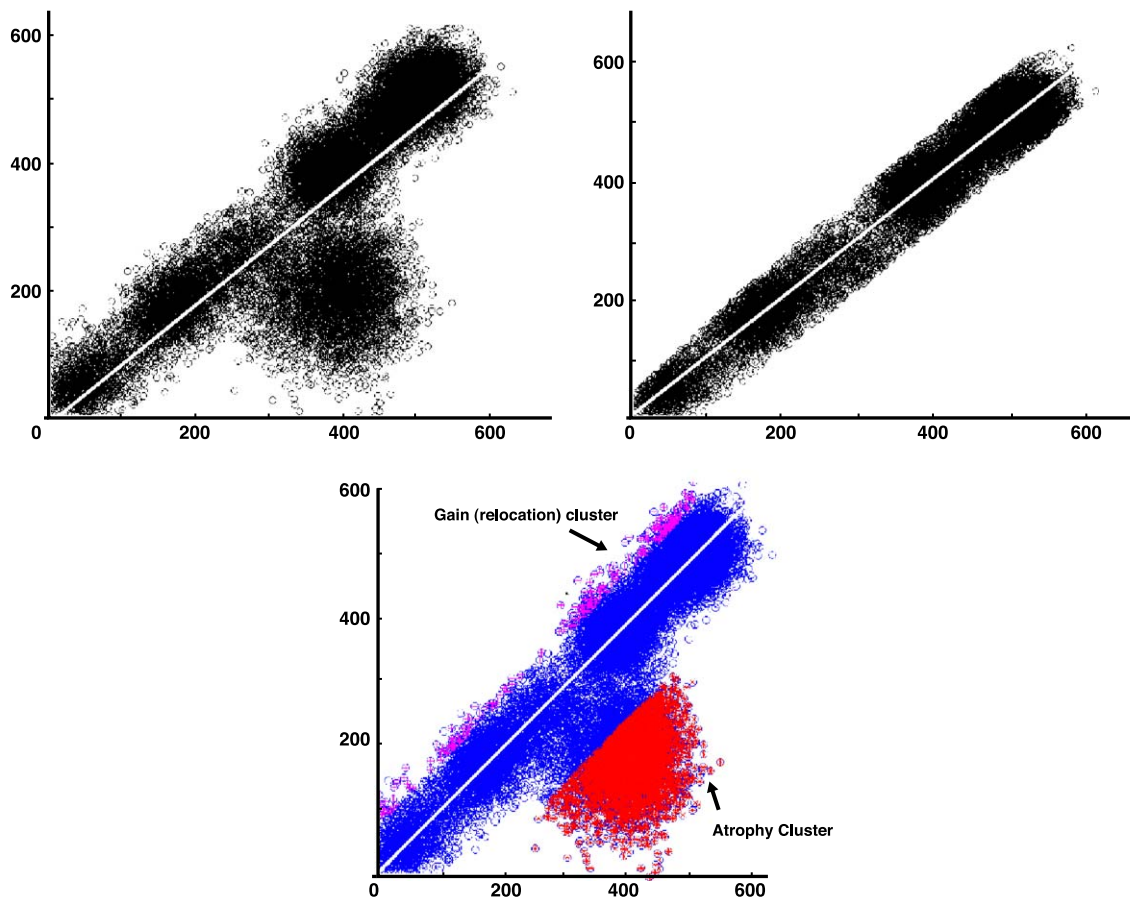


Fig. 1. Graphical representation of the IPCA, atrophy detection procedure. In the upper left panel, the ray was characterized in a single iteration, using the data from all voxel pairs over the brain volume, which included the potentially confounding effects of outliers related to brain atrophy. In the upper right panel, the ray was characterized after three iterations, which eliminated the outliers and characterized the ray independent of outlier effects. The lower panel illustrates the identification of (x_i, y_i) outliers by the IPCA in a simulated MRI set that included a region-of-atrophy (ROA) in the coregistered follow-up image. Voxels identified as volume loss are shown in red while those identified as volume gain are shown in pink.

from the baseline and follow-up images for each of the simulated data sets and all of the human MRIs, there was a highly linear relationship in paired voxel intensities (all correlation coefficients >0.98 , $P < 10^{-7}$). (Note that the direction of the regression line is different from that of the PCA major axis. Whereas the regression

line is estimated using distances perpendicular to the x -axis, the PCA estimates the ray using distances perpendicular to the ray itself.)

Although voxel-intensity shifts were not apparent in the data from the four sets of simulated MRI data, voxel-intensity shifts were apparent in the data from all of the human subjects. Most of the subjects had intercepts in the range of 1–8% of their maximal voxel intensity. Indeed, three of the healthy subjects who had no true changes in brain volume (their baseline and follow-up images were acquired 30-min apart) had intercepts greater than 20% of their maximal voxel intensity. The possibility of large between-image voxel-intensity shifts indicates one advantage of the IPCA. Whereas the DS method assumes that there is no voxel-intensity shift between the baseline and follow-up images, the IPCA computes changes in brain volume independent of the potentially confounding effects of a shift in voxel intensity. In fact, the DS method incorrectly detected extensive increases and decreases in regional brain volume in two of the three subjects with greatest voxel-intensity shifts (i.e., it led to numerous false signals), while the IPCA did not.

Fig. 1 illustrates the need to perform the PCA (i.e., characterize the ray) iteratively, using data from a simulated MRI set that included an ROA in the coregistered follow-up image. In the upper left panel, the ray was characterized in a single iteration, using the

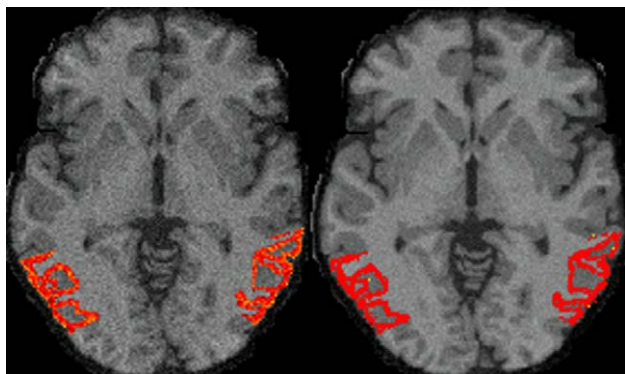


Fig. 2. IPCA-detected changes in brain volume mapped onto the baseline MRI. This figure illustrates the close correspondence between the anatomical distribution of simulated atrophy (left image, in red) and IPCA-detected atrophy (right image, in red).

Table 1
Effects of different outlier detection thresholds on the IPCA's ability to detect brain atrophy in a typical human subject^a

Outlier threshold	$I_{\%false}$	$I_{\%true}$		
		SD \times 0%	SD \times 50%	SD \times 100%
0.0005	0.074	5.21	4.97	4.41
0.001	0.145	5.30	5.10	4.66
0.005	0.31	5.35	5.18	4.78

^a The IPCA was evaluated using coregistered MRIs that were acquired 30-min apart in a normal volunteer, a simulated region-of-atrophy (ROA, 5.5% of baseline brain volume) in the follow-up scan, and three different levels of noise in the ROA (i.e., replacing ROA voxel intensities with Gaussian random numbers corresponding to the CSF mean intensity and different percentages of the CSF SD).

data from all voxel pairs over the brain volume, which included the potentially confounding effects of outliers related to brain atrophy; in the upper right panel, the ray was characterized after three iterations, which eliminated the outliers and characterized the ray independent of outlier effects.

(b) This study (i) identified an outlier threshold with which IPCA could detect known changes in brain volume as small as 0.04% of the whole brain volume while distinguishing this detected change from false signals and noise and (ii) demonstrated the potential value of empirically determining the optimal outlier threshold for an individual scanner, while providing a strategy for doing so. Table 1 shows the changes in brain volume detected using three outlier thresholds including $P = 0.0005$ when the IPCA was applied to a baseline and follow-up MRIs acquired 30-min apart in a normal volunteer with and without the introduction of a simulated ROA (5.5% of baseline brain volume) in the follow-up scan. In each of the 13 volunteers who had baseline and follow-up scans only 30-min apart and whose data were analyzed with and without the introduction of a simulated ROA of the same size, the outlier threshold of $P = 0.0005$ yielded a $I_{\%false}$ value (i.e., a falsely detected change in brain volume) less than 0.1% of whole brain volume (range -0.0294 to 0.074) and an $I_{\%true}$ (i.e., a correctly detected change in brain volume within the ROA itself) value at least 80% of the actual volume change within the ROA. The three columns collectively labeled $I_{\%true}$ list the IPCA-detected atrophy in the simulated ROA using three levels of uncertainty, such that the original voxel intensity in each ROA voxel was replaced with a Gaussian random number with 0%, 50% and 100% of the SD around the mean signal intensity in CSF space. As expected, the detected atrophy (with zero gain) coincided with the introduced atrophy and $I_{\%true}$ decreased as the CSF

voxel uncertainty level increased. While the results were similar for the simulated data, the outlier detection threshold for detecting $I_{\%false}$ below 0.1% and $I_{\%true}$ above 80% for the simulated data set was $P = 0.001$, different from the optimal outlier detection threshold of $P = 0.0005$ for our human data. While we recommend the default threshold of $P = 0.0005$, we have incorporated the ROA simulation procedure into the IPCA software package to permit researchers to confirm the suitability of this threshold on their own MRI system.

As illustrated in Table 2, the optimal outlier threshold determined above using our human MRI was further validated after introducing ROAs of different sizes, ranging from 0.04% to 5.5% of whole-brain volume, into the follow-up image. Each detected atrophy value in the table reflects the mean (\pm SD) after applying IPCA to an ROA of the same size and shape but with 50 different realizations in which Gaussian random numbers were introduced to simulate the conversion from gray matter to CSF. Although the IPCA overestimated the volume change for the smallest ROA (i.e., estimating a 0.125% volume change for the 0.04% ROA), the detected volume change was significantly greater than the $I_{\%false}$ value of 0.074% ($n = 50$, one-tailed t test: $P < 10^{-5}$); indeed, the detected volume change exceeded this $I_{\%false}$ value in all 50 simulations. The results were similar for the simulated MRI data ($P = 2.3 \times 10^{-5}$, mean detected atrophy 0.096 > false atrophy 0.058). Moreover, the IPCA-detected atrophy value was closely correlated with the corresponding simulated ROA size (100% SD: $R = 0.996$, $P = 1.3 \times 10^{-4}$ for the real data, and $R = 0.9976$, $P = 7.2 \times 10^{-5}$ for the simulated MRI; 50% SD: $R = 0.999$, $P = 3.6 \times 10^{-5}$ for the real data and $R = 0.9999$, $P = 3 \times 10^{-8}$ for the simulated data).

The lower panel of Fig. 1 graphically illustrates the identification of (x_i, y_i) outliers by the IPCA for a simulated MRI set that included an ROA in the coregistered follow-up image. Voxels identified as volume loss are shown in red while those identified as volume gain are shown in pink. The IPCA permits us to map the location of volume changes onto the baseline MRI. Fig. 2 shows the close correspondence between the anatomical distribution of voxels associated with simulated atrophy (left image) and those associated with IPCA-detected atrophy (right image).

(c) The IPCA method was comparable to the DS method in its ability to characterize and contrast rates of whole-brain atrophy in patients with probable AD and age-matched normal control subjects. Fig. 3 shows whole-brain atrophy rates in the eight patients with probable AD and eight normal controls computed using the IPCA and DS methods, respectively, using our default IPCA image-preprocessing settings and our default outlier detection threshold. The IPCA-detected atrophy rates of $2.17 \pm 0.52\%$ per

Table 2
Relationships between simulated and IPCA-detected atrophy sizes in a typical human subject^a

	Simulated ROA size (percent of whole brain volume)				
	0.04	0.45	0.9	2.0	5.5
ROA defined by	Detected atrophy (percent of whole brain volume)				
SD \times 100%	0.125 \pm 0.0254	0.346 \pm 0.073	0.856 \pm 0.162	1.347 \pm 0.275	4.411 \pm 0.657
SD \times 50%	0.141 \pm 0.0203	0.322 \pm 0.058	0.899 \pm 0.089	1.584 \pm 0.183	4.656 \pm 0.448

^a The IPCA was evaluated using coregistered MRIs that were acquired 30-min apart in a normal volunteer, simulated regions-of-atrophy (ROA) of different sizes, ranging from 0.04% to 5.5% of whole-brain volume, into the follow-up image, two different levels of noise (expressed as SD percent) in the ROA. Each detected atrophy percentage in the table reflects the mean (\pm SD) of IPCA-detected atrophy after applying it to an ROA of the same size and shape but with 50 different simulations in which Gaussian random numbers were used to characterize the CSF voxel intensities.

year in the patients vs. $0.41 \pm 0.22\%$ per year in the normal controls ($P = 7.8 \times 10^{-4}$ using the nonparametric Wilcoxon–Mann–Whitney test). In comparison, the DS method detected atrophy rates of $3.51 \pm 1.31\%$ per year in the patients vs. $0.48 \pm 0.29\%$ per year in the normal controls ($P = 7.8 \times 10^{-4}$). [Using Fox’s original image-preprocessing settings, the IPCA detected atrophy rates of $2.86 \pm 1.05\%$ per year in the patients vs. $0.49 \pm 0.36\%$ per year in the normal controls, while the DS method detected atrophy rates of $2.95 \pm 1.68\%$ per year in the patients vs. $0.22 \pm 0.16\%$ per year in the normal controls (again, $P = 7.8 \times 10^{-4}$ for both the IPCA and DS methods).] As illustrated in Fig. 3, there was no overlap between the detected annual rates of atrophy in the patient and normal control groups using either method. Furthermore, as shown in Fig. 4, high/low whole-brain atrophy rates detected by the IPCA method in the patients and normal controls were associated with the high/low rates detected using the DS method [Pearson correlation $R = 0.92$, $P = 2.1 \times 10^{-7}$ (nonparametric Spearman rank correlation = 0.94 , $P < 0.0005$).] Using the ratio of the lowest detected atrophy rate in the patient group to the highest rate in the controls as an indicator of the separation between subject groups, the ratio using our default image-preprocessing settings was 1.8 for both the IPCA and DS method.

(d) We were unable to find that the combination of SPM99 image-preprocessing settings had any significant effects on the ability of the IPCA to distinguish whole-brain atrophy rates in the patients with probable AD from those in the normal control subjects. Indeed, the IPCA detected higher rates of whole-brain atrophy in the patients than in the normal controls using every one of the 18 combinations of image-preprocessing settings ($P = 7.8 \times 10^{-4}$ for each comparison based on the Wilcoxon–Mann–Whitney test). Relatedly, we did not find significant difference among the 18 combinations of image-preprocessing settings in the ability of the IPCA to distinguish atrophy rates in the two subject groups

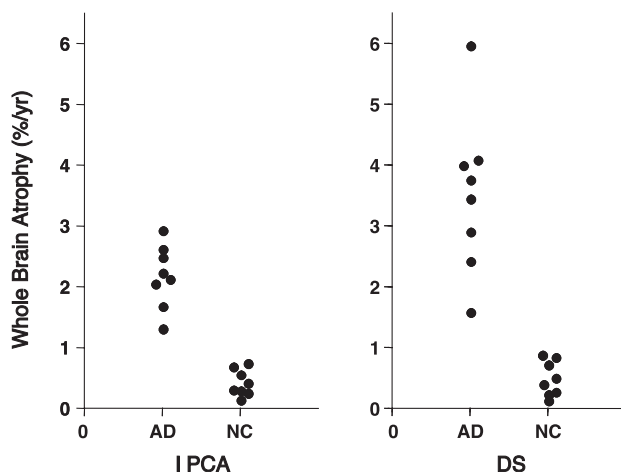


Fig. 3. Whole-brain atrophy rates (percent volume loss per year, expressed as a percent of baseline brain volume per year) in patients with probable AD and age-matched normal control subjects computed using the IPCA (left panel) and DS method (right panel) in conjunction with our default image preprocessing settings. The two methods were comparable in their ability to distinguish whole-brain atrophy rates in the patients from those in the controls ($P = 7.8 \times 10^{-4}$ for each method using the nonparametric Wilcoxon–Mann–Whitney test).

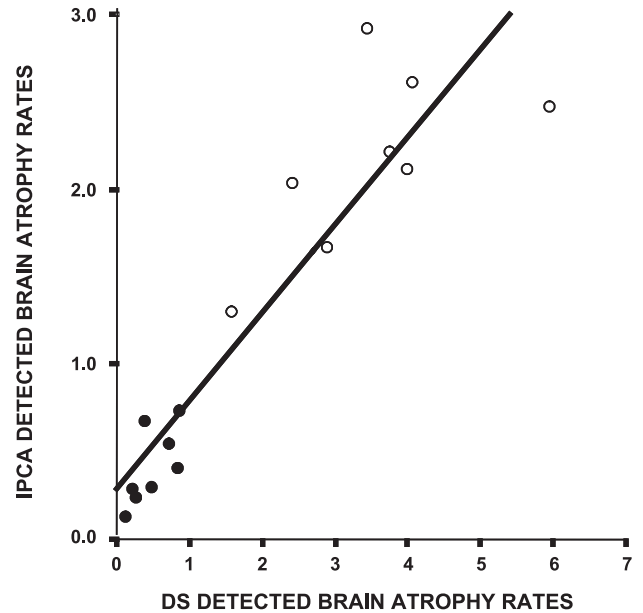


Fig. 4. Correlation between whole-brain atrophy rates computed in the patients with probable AD (open circles) and normal control subjects (closed circles) using the DS and IPCA methods (Pearson correlation $R = 0.92$, $P = 2.1 \times 10^{-7}$, nonparametric Spearman rank correlation = 0.94 , $P < 0.0005$).

(General linear model, F test, $P > 0.9$); and there was no overlap between the IPCA-detected annual rates of atrophy in the patient and normal control groups using any of these setting combinations. Again, using the ratio of the lowest detected atrophy rate in the patient group to the highest rate in the controls as an indicator of the separation between subject groups, this ratio was greater than 1.5 for all of combinations of image-preprocessing settings except one: the combination that included the nine-parameter coregistration, no inhomogeneity correction, and brain volume defined by the modified SPM brain extraction algorithm (ratio = 1.02). For each of the 18 combinations of image-preprocessing settings, whole-brain atrophy rates detected by the IPCA method in the patients and normal controls were closely correlated with the rates detected using the DS method: Pearson correlation coefficients ranged between 0.84 and 0.96 and P values ranged between 2.59×10^{-5} and 6.3×10^{-9} , respectively; Spearman rank correlation coefficients ranged between 0.85 and 0.98 (all $P < 0.0005$); and all of these correlations survived Bonferroni correction at $P < 0.05$ for the 18 comparisons. Relatedly, there was no significant difference among the 18 combinations of image-preprocessing settings in their correlation coefficients (general linear model, $P \geq 0.25$). While we were unable to detect any significant difference among the image preprocessing settings used to detect whole-brain atrophy rates using the IPCA, we currently recommend the use of default settings to promote comparability of findings among different laboratories.

Discussion

This article introduces an automated method for the computation of changes in brain volume from sequential magnetic resonance images (MRIs) using an iterative principal component

analysis (IPCA) and demonstrates the method's ability to characterize abnormally high rates of whole-brain atrophy in patients with Alzheimer's disease (AD). Analyses of simulated and real MRI data support the underlying assumption of a linear relationship in paired voxel intensities, identify an outlier distance threshold that optimizes the trade-off between sensitivity and specificity in the detection of small volume changes, and demonstrate an ability to detect changes as small as 0.04% of brain volume without confounding effects of between-scan shifts in voxel intensity. The IPCA was comparable to the established but partly manual digital subtraction (DS) method in characterizing and contrasting annual rates of whole-brain atrophy in patients with probable AD and normal control subjects: whole-brain atrophy rates detected using the two methods were correlated—and they were comparable in their ability to distinguish AD from normal aging.

Although the annual rates of atrophy computed using the DS method tend to be higher than those computed using the IPCA method, the two methods were comparable in their power to distinguish patients with probable AD from normal controls. Individual rates of whole-brain atrophy depend in part on the thresholds used by each method. Thus, brain volume changes computed using the DS method depend in part on the chosen I_1 and I_2 values, while those computed using the IPCA depend in part on the chosen outlier detection threshold. The default IPCA outlier detection threshold of $P = 0.0005$ was not selected because of its ability to reproduce the rates found using the DS method. Instead, this threshold was determined using an independent set of real data to optimize the IPCA's ability to distinguish subtle changes in brain volume from false signals and noise. While the DS and IPCA methods appear to have comparable power to track the progression of AD and evaluate the disease-modifying effects of candidate treatments over 1 year, additional comparisons are needed to consider the possibility that the IPCA might offer superior power to track AD and assess candidate treatments over shorter intervals.

While we recommend the default threshold of $P = 0.0005$ to facilitate the comparability of measurements among different laboratories, we have incorporated the ROA simulation procedure into the IPCA software package to permit researchers to confirm the suitability of this threshold on their own MRI system. This simulation procedure could also be used to characterize the comparability of measurements when hardware upgrades are introduced in longitudinal studies (a significant challenge to the computation of changes in brain volume from sequential MRIs) and to characterize the comparability of measurements in multi-center studies.

As previously noted, the IPCA depends on the existence of a linear relationship in the coregistered images' paired voxel intensities. This study demonstrated a highly linear relationship in each of the simulated and real image pairs, including those acquired 2 years apart and those in which ROAs 5.5% of whole-brain volume were introduced into the follow-up image. It also demonstrated an ability to detect known ROAs 0.04–5.5% of whole-brain volume, reflecting the range one could expect in association with AD and aging. Additional studies would be needed to confirm the existence of a linear relationship in paired voxel intensities and detect changes in brain volume following the introduction of a hardware upgrade and in cases associated with a more extreme change in brain volume (e.g., following a lobectomy).

Unlike the DS method, the IPCA computes changes in brain volume independent of the potentially confounding effects of a between-scan shift in voxel intensity. There was a between-scan

difference in voxel intensity in all of our human data, which cannot be fully addressed by the normalization procedure used in the DS method. Indeed, three of the subjects with the highest voxel-intensity shifts had been studied only 30-min apart. In two of these individuals, the DS method falsely detected volume increases and decreases in many areas of the brain, whereas the IPCA did not. While many of these DS-detected increases and decreases may be off-setting in the computation of whole-brain atrophy (an assertion that is supported by the comparability of DS and IPCA-generated measurements and findings in the study of probable AD patients and normal controls), the IPCA can address this potentially confounding effect in the computation of changes in either regional or whole-brain volume. [We have also observed a severe between-scan voxel-intensity shift in an ongoing mouse brain MRI study, which uses T_2 -weighted MRIs. Whereas the DS method falsely detected increases and decreases in mouse brain volume, our preliminary findings suggest that the IPCA approach may also have the potential for tracking changes in small animal MRI studies despite differences in MR field strength and sequences (Chen et al., 2002, abstract).] Simulation studies in our laboratory indicate that the DS method is more likely to detect false increases and decreases in regional brain volume when there is a large between-scan shift in global voxel intensity and/or a large range in voxel intensities in the baseline or follow-up image.

We have incorporated SPM99 into the IPCA software package to automatically coregister the follow-up MRI to the baseline image (with or without three additional scaling procedures) to account for a between-scan drift in voxel geometry, correct them for image inhomogeneities (Ashburner and Friston, 1997), and provide baseline measurements of whole brain or intracranial volume (to compute changes in brain volume as a percentage of baseline measurements). While we were unable to detect any significant difference among the image preprocessing settings used to detect whole-brain atrophy rates using the IPCA (a fact that might need to be reconfirmed with further tests in larger sample size), we recommend the use of our default settings to promote comparability of findings among different laboratories. This is also in recognition of the slightly different degree of nonoverlapping between the two groups when various preprocessing settings are used. We are currently incorporating SPM2 into the IPCA software package and will determine whether its preprocessing algorithms are associated with further improvement in the ability of the IPCA to characterize changes in brain volume.

Several other methods have been proposed to provide information about brain atrophy (e.g., a reduction in brain volume or gray matter concentration) from the cross-sectional or longitudinal analysis of brain MRIs. These methods include Structural Image Evaluation, using Normalization, of Atrophy (SIENA; Smith et al., 2002), Regional Analysis of Volumes Examined in Normalized Space (RAVENS; Davatzikos et al., 2001), Voxel-Based Morphometry (VBM; Ashburner and Friston, 2000; Good et al., 2001), and Dynamic Brain Mapping (Thompson et al., 2003). VBM, RAVENS, and the Dynamic Brain Mapping approaches are procedures that especially detect volume change voxel-wise in a standardized brain template space or the cortical surface. Complementary to the voxel-based approaches, IPCA, DS, and SIENA measure the brain volume changes globally using each subject's baseline as his/her own control, and need no spatial normalization. Like the IPCA method, SIENA (Smith et al., 2002) can compute changes in brain volume from sequential MRIs automatically, implicitly accounting for the potentially confounding effect of a between-scan shift in

voxel intensity. In contrast to SIENA, IPCA is relatively simple and straightforward. Further investigation is needed to directly compare the IPCA method with SIENA in the study of patients with probable AD and age-matched normal control subjects.

The IPCA method can be used to characterize and contrast rates of whole-brain atrophy associated with AD and normal aging from sequential MRIs. This image-analysis strategy is automated, capitalizes on widely available SPM99 software, may be relatively easy to implement at different centers, and may facilitate the comparison of measurements and findings among different users. Thus, this automated algorithm offers promise in tracking disease progression in persons afflicted by or at risk for AD and in evaluating the disease-modifying efficacy of putative AD treatments. It also offers promise in characterizing the longitudinal changes in brain volume associated normal aging, brain development, and the disorders to which they may be related.

Acknowledgments

This work was funded in part by the Arizona Alzheimer's Research Center, the National Institute of Mental Health (grant MH57899 to EMR), and the National Institute on Aging (Alzheimer's Disease Core Center grant AG19610 to EMR). The authors wish to thank Jennifer Frost, Alisa Domb, Sandy Goodwin, Les Mullen, Pat Raso, Anita Prouty, Christine Burns, and Connie Boker for their support and assistance, and they thank Dr. Lawrence Mayer and Dr. Richard Gerkin for their statistical advice.

Appendix A

Let d_1 be the atrophy threshold, such that atrophy occurs in location (x_i, y_i) if its distance to the major axis is greater than d_1 and it is below the major axis. The gain threshold d_2 is defined similarly. When the following procedure is implemented, none of the points are considered outliers:

(1) PCA Major Axis Determination: Until the major PCA axis converges or the maximal number of iterations is reached:

Do

Apply PCA:

Excluding outliers, apply PCA to the rest of (x_i, y_i) points.

Detect and Label Outliers

Project (x_i, y_i) points to the minor (second) axis, perpendicular to the major axis.

Identify the uppermost and lowermost 10% of points in terms of projection distances from median.

(We found that any value between 5% and 10% generated the same PCA axis).

Label these points as outliers.

End Do

(2) Atrophy and/or Gain Voxel Detection:

Detect atrophy and/or gain voxels:

Put back the points removed in the IPCA procedure.

Project all (x_i, y_i) points to the minor component,

Label points as outliers using d_1 and d_2 values.

Label outliers above the major axis to represent tissue gain (relocation).

Label outliers below the major axis to represent tissue loss.

Calculate the percentage of net brain volume change as the ratio of difference between tissue loss and gain to the whole brain volume.

References

- Ashburner, J., Friston, K., 1997. Multimodal image coregistration and partitioning—a unified framework. *NeuroImage* 6, 209–217.
- Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—the methods. *NeuroImage* 11, 805–821.
- Chen, K., Reiman, E., He, T., Galons, J., Hauss-Wegrzyniak, B., Stevenson, J., Valla, J., Trouard, T., Wenk, G., Alexander, G., 2002. Improved iterative principal component analysis for detecting whole brain atrophy from sequential mouse MRI's. Annual meeting of the Society of neuroscience, Orlando, FL. Abstract.
- Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C., 1998. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging* 17, 463–468.
- Davatzikos, C., Genc, A., Xu, D., Resnick, S.M., 2001. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *NeuroImage* 14, 1361–1369.
- 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., Freeborough, P.A., 1997. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. *J. Magn. Reson. Imaging* 7, 1069–1075.
- Fox, N.C., Freeborough, P.A., Rossor, M.N., 1996. Visualisation and quantification of rates of atrophy in Alzheimer's disease. *Lancet* 348, 94–97.
- Fox, N.C., Warrington, E.K., Rossor, M.N., 1999. Serial magnetic resonance imaging of cerebral atrophy in preclinical Alzheimer's disease. *Lancet* 353, 2125.
- Fox, N.C., Cousens, S., Seahill, 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., Scathill, 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.
- Freeborough, P.A., Fox, N.C., 1997. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans. Med. Imaging* 16, 623–629.
- Freeborough, P.A., Woods, R.P., Fox, N.C., 1996. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *J. Comput. Assist. Tomogr.* 20, 1012–1022.
- Friston, K.J., Ashburner, J., Poline, J.B., Frith, C.D., Heather, J.D., Frackowiak, R.S.J., 1995. Spatial registration and normalisation of images. *Hum. Brain Mapp.* 2, 165–189.
- 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.
- Kwan, R.K., Evans, A.C., Pike, G.B., 1999. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans. Med. Imaging* 18, 1085–1097.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan,

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