A Brief History of Voxel-Based Grey Matter Analysis in Alzheimer's Disease

Article *in* Journal of Alzheimer's disease: JAD · September 2013

DOI: 10.3233/JAD-130362 · Source: PubMed

7 authors, including:

Lara Z Diaz-de-Grenu
Cambridge University Hospitals NHS Founda…
6 PUBLICATIONS **49** CITATIONS
SEE PROFILE

Julio Acosta-Cabronero
University College London
62 PUBLICATIONS **1,460** CITATIONS
SEE PROFILE

Seyed A Sajjadi
University of California, Irvine
18 PUBLICATIONS **218** CITATIONS
SEE PROFILE

Guy B Williams
University of Cambridge
111 PUBLICATIONS **4,264** CITATIONS
SEE PROFILE

Journal of Alzheimer's Disease xx (20xx) x–xx
DOI 10.3233/JAD-130362
IOS Press

A Brief History of Voxel-Based Grey Matter Analysis in Alzheimer's Disease

Lara Z. Diaz-de-Grenu[a], Julio Acosta-Cabronero[a,e], Yao Feng Victor Chong[b], Joao M.S. Pereira[a,c], Seyed A. Sajjadi[a], Guy B. Williams[d] and Peter J. Nestor[e,*]
[a]*Herchel Smith Building for Brain and Mind Sciences, Department of Clinical Neurosciences, University of Cambridge School of Clinical Medicine, Cambridge, UK*
[b]*School of Clinical Medicine, University of Cambridge, Cambridge, UK*
[c]*Laboratory of Biostatistics and Medical Informatics, IBILI - Faculty of Medicine, University of Coimbra, Portugal*
[d]*Wolfson Brain Imaging Centre, Department of Clinical Neurosciences, University of Cambridge School of Clinical Medicine, Cambridge, UK*
[e]*German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany*

Handling Associate Editor: Murray Grossman

Accepted 31 July 2013

Abstract. Voxel-based morphometry (VBM) and cortical thickness measurement are common techniques to identify regional atrophy in neurodegenerative diseases such as Alzheimer's disease (AD). Because studies employing these methods draw conclusions regarding patterns of regional cortical degeneration, it is important to be aware of their possible limitations. To evaluate the effect of different VBM versions, we performed voxel-based analyses through successive versions—from SPM99 to SPM8—as well as FSL-VBM on $n = 20$ AD patients and $n = 20$ controls. Reproducibility was assessed in an independent sample, again of $n = 20$ per group, from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Further, we tested the hypothesis that VBM can sensitively detect hippocampal atrophy, but is relatively insensitive to changes in the cortical ribbon, by contrasting VBM with FreeSurfer cortical thickness measurements. The results with both datasets confirmed that VBM preferentially identifies grey matter lesions in the mesial temporal lobe but is largely insensitive to isocortical atrophy. In contrast, FreeSurfer identified thinning of cortical ribbon association cortex more significant in post- rather than pre-Rolandic areas and with relative preservation of primary sensory-motor regions—in other words precisely as would be expected in AD. The results highlight a bias that VBM has toward detecting mesial temporal lobe atrophy. This finding has important implications for interpretation of clinical and cognitive studies in AD.

Keywords: Alzheimer's disease, atrophy, computer-assisted, diagnosis, image interpretation, magnetic resonance imaging

INTRODUCTION

Tools to detect structural change in neurodegenerative diseases (e.g., voxel-based morphometry (VBM) or cortical thickness measurement) have been evolving for more than a decade. These techniques have the potential to offer mechanistic insights into conditions such as Alzheimer's disease (AD) by identifying regions of selective vulnerability. The tools themselves, however, may have technical limitations that need to be taken into account when interpreting results. Furthermore, although discrepancies between studies are often considered in terms of differing dementia severity and power, little attention has been paid to the possible impact that successive versions of these tools might have had.

VBM is a neuroimaging analysis tool included in the Statistical Parametric Mapping (SPM) package

*Correspondence to: Peter Nestor, German Center for Neurodegenerative Diseases (DZNE), Leipziger Strasse 44, 39210 Magdeburg, Germany. Tel.: +49 391 6724513; Fax: +49 391 6724528; E-mail: peter.nestor@dzne.de.

(http://www.fil.ion.ucl.ac.uk/spm/) that can identify group-level structural changes in patients compared to controls [1]. Since it is fully automated, this technique is less time consuming than region of interest (ROI) analysis, and, because it obviates the need to pre-specify these ROIs, it is often described as an "unbiased" technique. It has been shown, however, that registration accuracy varies across brain regions and diseases [2], therefore introducing its own form of bias by potentially reducing sensitivity in areas of greater misregistration. VBM has been implemented in several versions of the SPM package—SPM99, SPM2, SPM5, and SPM8—with presumed improvements to the algorithm in successive versions though the basic concept remains the same. The FMRIB Software Library (FSL) (http://www.fmrib.ox.ac.uk/fsl/fslvbm) also contains a version of VBM whose principles are very similar to those of SPM; in fact, the former are based on the optimized SPM-VBM protocol developed by [3]. The algorithms used for segmentation, registration, etc. in FSL, are, however, different from those in SPM and therefore might lead to divergent results.

An alternative approach to automated atrophy detection in the cerebral cortex is cortical thickness measurement. FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) is an image analysis suite that enables such measurements through cortical reconstruction and volumetric segmentation. The principal concept behind this tool is that volumetric T1-weighted images are processed not only at a volumetric level but also by creating surfaces of the brain—grey matter/white matter and grey matter/pial boundaries—to produce statistical maps allowing for measurement of cortical thickness.

There are other technical differences in VBM and FreeSurfer that could have an impact on the performance of each method. For instance, inter-subject registration in SPM-VBM versions up to and including SPM5 involved matching subjects' scans to a common template; this process only used around 1000 parameters and therefore was unable to model very detailed deformation. SPM8's use of Diffeomorphic Anatomical Registration using Exponentiated Lie Algebra (DARTEL) has tried to compensate for this inter-subject registration problem by computing a "flow field" to generate the scan deformations [4]. FreeSurfer, in contrast, uses a spherical surface-based coordinate system, which allows for a high accuracy in localizing brain structures [5]. Another strength of FreeSurfer subject warping, at least in theory, is that it uses the white matter surface as its target, therefore grey matter changes should not influence registration accuracy.

The first aim of this study was to provide an objective comparison of voxel-based analyses through successive SPM versions (from SPM99 to SPM8) as well as FSL-VBM to assess the validity of directly contrasting results of studies that employed these different methods or whether such differences need to be considered as potential confounds. The "history" mentioned in the title of this article refers to the types of VBM methods that have been published in past studies on clinical AD (Table 1), which includes [6–15]; the study does not concern itself with technical details that might improve its sensitivity in the future. The second aim of the study was to test a specific hypothesis: that the VBM technique, as published in clinical studies, is preferentially biased toward the detection of hippocampal atrophy when compared to the cortical ribbon. The reasoning behind this being that the hippocampus is a large grey matter structure in terms of thickness and, most importantly, has a highly similar topography across subjects; the cortical ribbon, however, is comparatively thin and has a gyral pattern unique to each individual. This hypothesis was tested by contrasting the VBM results with those obtained from cortical thickness measurements using FreeSurfer. All analyses involved the same dataset of $n = 20$ AD patients and $n = 20$ controls. Additionally, in order to test the reproducibility of the results, the same analyses were performed with data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

Undertaking head-to-head comparisons of this type in a clinical sample raises the perennial problem of a lack of gold standard if results are discordant. While this is a valid concern, it cannot be used as an excuse to avoid such studies; if one relies solely on validation in non-clinical samples such as simulations, one risks the problem that the simulation might inadequately represent the real world. In this study, the gold standard was prior knowledge of the degeneration pattern in AD. It is well known that hippocampal atrophy is a feature of AD. Meanwhile, numerous cerebral metabolic studies [16–21] have highlighted that the most involved areas of degeneration in the cortical ribbon are posterior cingulate and precuneus, extending on the lateral surface to parietal and temporal association areas. Therefore, if an atrophy detection technique revealed this same pattern, it would be highly likely to represent a true positive rather than a false positive. Hypothetically, quantitative neuropathological analysis of cortical volume would be the ultimate gold standard but, unsurprising given the labor involved, no studies of this type that cover the entire brain exist;

Table 1
Short literature review of types of VBM methods published in past studies on clinical AD

Paper	No. AD	No. Control	MMSE	Version -Method	Smoothing (mm)	Correction	Atrophic areas
[6]	19	16	22.4 ± 2.5	SPM99-Standard VBM	12	$p < 0.1$ Corrected for multiple comparisons	Amygdala, anterior hippocampus, insula, entorhinal cortex areas and posterior cingulate cortex
[7]	29	26	21 ± 4	SPM99-Standard VBM	12	$p < 0.05$ corrected for multiple comparisons	Mainly in hippocampus, amygdala and uncus. Smaller areas in fusiform gyrus, precuneus, insula and anterior and posterior cingulate
[8]	61	33	17.9 ± 4	SPM5	12	FDR $p < 0.01$	Temporal, frontal, parietal and basal ganglia
						FWE $p < 0.01$	Left inferior and superior temporal gyrus, right inferior frontal gyrus, parahippocampal gyrus, hippocampus, amygdala and left caudate
[9]	30	50	18.5 ± 4.6	DARTEL	6	FDR $p < 0.05$	Parietal and temporal regions
[10]	26	92	20.7 ± 3.1	SPM99-Standard VBM	12	$p < 0.001$ uncorrected	Hippocampus, entorhinal cortex, parahippocampal cortex
[11]	55 MCI	32	24 ± 2	SPM99-Standard VBM	12	$p < 0.001$ uncorrected	Amygdala and head and tail of hippocampus
[12]	7	7	21.6 ± 4	SPM99-Standard VBM	8	$p < 0.0005$ uncorrected	Hippocampus, insula, caudate nucleus
[13]	51	43	21.8	SPM2-Optimised VBM	12	$p < 0.05$ FWE	Neocortical association areas, lateral temporal neocortex, hippocampus
[14]	50	80	19.6 ± 4.0	DARTEL	8	FWE $p < 0.05$	Medial temporal structures: amygdala, hippocampus and hippocampal gyri
[15]	18	15	24.3 ± 2.6	SPM2-Optimised VBM	14.6	FDR $p < 0.05$	Hippocampus, temporal neocortex and posterior associative areas

MCI, mild cognitive impairment.

even if such studies were available, they would be biased to end stage disease and therefore possibly not reflective of the atrophy pattern in early disease.

METHODS

Subjects

The study comprised $n = 20$ patients diagnosed with mild probable Alzheimer's disease according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Diseases Association criteria [22] and $n = 20$ healthy elderly controls (Table 2). The cohort size was chosen to be of a similar order of magnitude to those of past clinical studies (Table 1). All controls were screened to ensure that none suffered from neurological or major psychiatric illness, and their performance was normal according to a global cognitive measure (the Addenbrooke's cognitive examination-revised or ACE-R [23]). All participants gave written informed consent to participate in the research and the study was approved by the regional Research Ethics Committee.

Image acquisition

Magnetic resonance imaging (MRI) scans were acquired on a Siemens Trio 3T system (Siemens Medical Systems, Erlangen, Germany) equipped with gradient coils capable of 45 mT/m and slew rate of 200 T/m/s, and a 12-channel phased-array total imaging matrix head-coil (Siemens Medical Systems, Erlangen, Germany). The 3D T1-weighted acquisition was an MPRAGE pulse sequence with the following scan parameters: TR/TE/inversion time (TI)/flip angle = 2300 ms/2.86 ms/900 ms/9°, 144 slices, 192×192 matrix dimensions and $1.25 \times 1.25 \times 1.25\,mm^3$ voxel size. Receiver bandwidth and echo spacing were 240 Hz/pixel and 6.7 ms, respectively.

Table 2
Demographics and general cognitive abilities of the AD and control groups

Demographics	AD	Controls	*p*
Gender, M:F	13 : 7	8 : 12	NS
Age, y	63.3 (4.8)	68.9 (6.3)	NS
Tests			
MMSE /30	23.6 (3.3)	28.9 (1.1)	<0.0001
ACE-R /100	71.4 (10.7)	94.1 (3.4)	<0.0001

Values are given as mean (standard deviation). NS, not significant; MMSE, Mini-Mental State Examination; ACE-R, Addenbrooke's cognitive examination-revised.

Prior to acquisition, the field of view was aligned in stereotactic space: the axial plane was aligned with the anterior commissure–posterior commissure line on the midline sagittal slice, and the inter-hemispheric fissure was set parallel to the sagittal plane and perpendicular to the coronal plane. In order to maximize acquisition consistency across subjects, the scanning bed was adjusted to match the center of the thalamus in the mid-sagittal plane with the scanner isocenter.

Voxel-based morphometry

Two population contrasts (AD versus controls) were used to compare VBM performance with the following SPM versions: SPM99, standard and optimized; SPM2, standard and optimized; SPM5; and SPM8 with DARTEL. In each analysis, the selected algorithm aimed to mirror previous studies with their respective methodology. The studies listed in Table 1 show that previous studies use a variety of discordant parameters for their analyses with regard to smoothing kernels and statistical thresholds, etc. Replicating each method precisely though would have meant that the first aim of the study—to assess whether different versions of VBM have different sensitivities—would not be possible because discrepancies might have reflected these different parameters. Therefore a middle-ground methodology with respect to such parameters was chosen.

Standard VBM (SPM99 and SPM2) involved spatial normalization to the SPM99 and SPM2 T1 templates respectively, segmentation of the spatially normalized structural images into grey matter, white matter, and cerebrospinal fluid, and modulation of tissue segments. One of the drawbacks of this technique is that spurious volumetric differences might be detected due to spatial normalization errors. It is known that VBM may lead to ambiguous results due to systematic miswarping of a particular region across groups [24]. In order to address this known problem, an optimized version of VBM (for both SPM99 and SPM2) was introduced using an iterative approach to segmentation and normalization [3]. Optimized VBM consists of a native segmentation step using prior tissue probabilities; this is followed by spatial normalization of the resulting grey matter segment, which leads to an optimal transformation that is then applied to the original structural image. The standardized image is subsequently segmented and modulated using the Jacobian determinant of the transformation field; in SPM2, this procedure is included, whereas in SPM99, manual intervention is required. SPM5 introduced "Unified segmentation"

[25], which optimized image registration, tissue classification, and radio-frequency bias correction in the same step. Default settings were employed to process whole-head (default SPM5) T1-weighted MRI data. DARTEL [4] implements a new registration approach: a "flow field able to generate both forward and backward deformations" [4]. Images were first segmented with the unified segmentation method and were then imported to DARTEL, where all grey and white matter segments were simultaneously registered to a custom-average template. Deformation fields defining the transformation of the average template to each individual image were computed, and a new custom template was generated by averaging the inverse transformation of all images. This procedure was repeated 18 times prior to modulation.

In all pipelines described above, modulated grey matter segments were smoothed with an 8-mm full-width-at-half-maximum (FWHM) isotropic Gaussian kernel. Finally, group comparisons (via two-sample unpaired *T*-tests) were performed with a) a false-discovery rate (FDR) statistical threshold of $q < 0.05$ and b) an uncorrected threshold of $p < 0.05$.

The software release used for FSL-VBM was FSL v4.1.2. First, structural images were skull-stripped using Brain Extraction Tool (BET v2.1) [26] with default settings; this step was followed by tissue-type segmentation, which was carried out using FMRIB's Automated Segmentation Tool (FAST4) [27]. Grey-matter partial volume images were then affine-registered to the MNI152 standard template (Montreal Neurological Institute, McGill University, Canada) using FMRIB's Linear Image Registration Tool (FLIRT v5.5) [28, 29]. The resulting images were averaged to create a study-specific template, to which native grey matter segments were non-linearly registered using FMRIB's Nonlinear Registration Tool (FNIRT) [30] and modulated (by the Jacobian determinant of the warp field to correct for local expansion/contraction). Modulated segments were then smoothed with an isotropic Gaussian kernel (sigma = 3.5 mm, equivalent to FWHM = 8 mm). Finally, a permutation-based (5,000 runs) voxel-wise approach was applied to perform non-parametric statistical comparisons between groups. As with SPM, two-population contrasts were examined at an uncorrected threshold of $p < 0.05$ and with FDR correction of $q < 0.05$ to account for the effect of multiple comparisons (http://www.fmrib.ox.ac.uk/fsl/randomise/fdr.html).

FreeSurfer cortical thickness analysis

Detailed information on the processing steps for FreeSurfer (FS v5.1.0, http://surfer.nmr.mgh.harvard.edu) have been extensively described in previous publications [31–38]. The FreeSurfer pipeline includes a skull-stripping step prior to pial

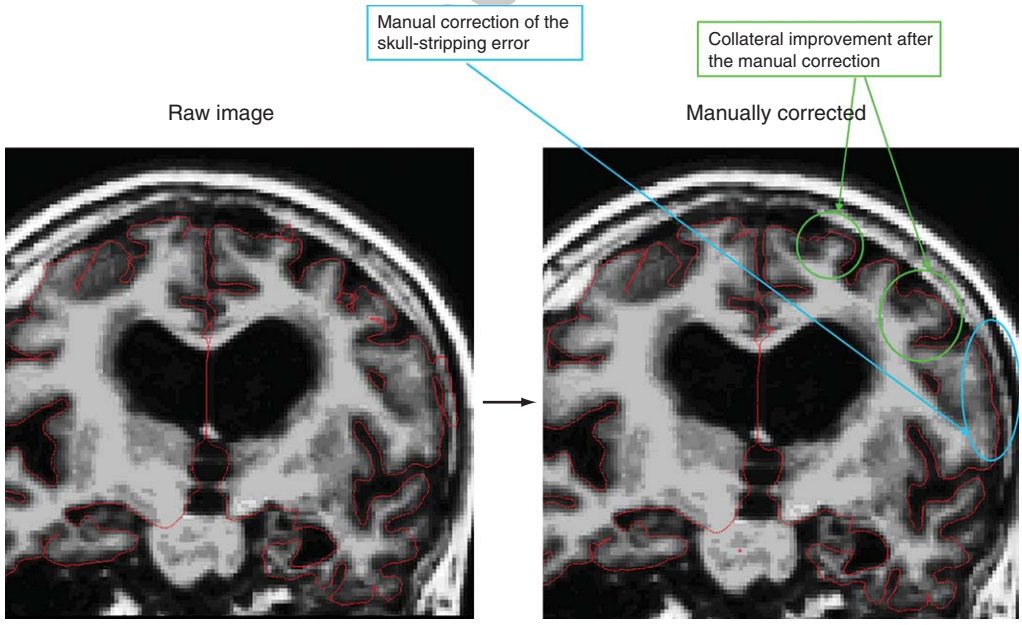

Fig. 1. FreeSurfer's pial surface delineation before (left) and after (right) manual correction of skull-stripping errors.

surface and grey-white matter interface reconstruction. Cortical thickness is calculated from the intensity and continuity information of the generated surfaces. Three different analyses were performed. First, the standard FS stream was used with default settings. Next, in order to test whether correcting errors in pial surface delineation would improve sensitivity, two additional processing pipelines were attempted. Pre-processing errors were reported in previous publications [9, 39], and were clearly evident in the current study. Around half of the pial surface errors produced by the skull-stripping step in FS arose on the ventral surface of the brain, one third on the hemispheric convexities, and one sixth on the medial surface. The most commonly affected structures were the inferior temporal and lateral occipitotemporal gyri, precentral gyrus, and medial occipitotemporal gyrus. The most common error found across all subjects regardless of group was failure to exclude the tentorium cerebelli and parts of the cerebellum. Moreover, a few errors that could not be corrected by voxel deletion, such as failure to include the parahippocampal gyrus in its entirety, were consistently found.

In an attempt to correct these errors, an alternative approach involved substituting FS pre-processing step *autorecon1*, with a different pre-processing pipeline described elsewhere [2] that also included radio-frequency bias correction [40]. Subsequent FS steps—*autorecon2* and *autorecon3*—were used with default settings. The third approach was full manual editing of pial delineations in an attempt to correct for skull-stripping failure. Errors were edited using FreeSurfer's built-in scan visualization and editing tool, Tkmedit. Manual improvements in the delineation of the pial surface are illustrated in Fig 1. Cortical thickness measurements of all three processing pipelines were visualized on the inflated surface of the average of all subjects' brain reconstruction.

The optimal smoothing kernel for FS is a matter of debate but it is known to be influenced by the number of subjects involved in the analysis: the smaller the cohort, the larger the required smoothing kernel [41]. The default is a 10-mm FWHM Gaussian kernel, but more recent studies have shown that larger kernels might render more plausible results [41, 42]. Consequently, a set of six different surface-based Gaussian blurring kernels were assessed: a) 10-mm, b) 15-mm, c) 20-mm, d) 25-mm, e) 30-mm, and f) 35-mm FWHM. Multiple comparison correction was performed to control the false-discovery rate at a 0.05 significance level.

Replication in an independent dataset

ADNI data was used to test the reproducibility of the results following the same analysis methods (see Supplementary Material). The ADNI database (http://adni.loni.ucla.edu/) is a longitudinal multi-site study to assess the progression of amnesic mild cognitive impairment and probable AD through serial MRI, positron emission tomography (PET), other biological markers, and clinical tests. ADNI is supported by the National Institutes of Health, private pharmaceutical companies, and non-profit organizations, with approximately 50 medical center and university sites across the United States and Canada. Up to date information can be found at (http://www.adni-info.org/).

RESULTS

FDR corrected results of VBM

VBM results (summarized in Figs. 2 and 3) are presented at FDR, $q < 0.05$. In SPM99 (standard and optimized), differences between patients and controls did not survive this statistical threshold. Grey matter atrophy was found in medial temporal lobe structures (hippocampi, amygdalae, and parahippocampal gyri) for all remaining VBM analyses (Fig. 2) with statistically stronger results for DARTEL and FSL. The DARTEL results were more restricted to the hippocampus, whereas in other VBM approaches—including FSL—they extended to adjacent mesial temporal areas.

SPM2 (standard and optimized) identified cortical atrophy, specifically in the posterior cingulate/precuneus (see Fig. 3) and insula. All other VBM versions—including FSL—failed to identify any lesions beyond the mesial temporal area. SPM2 was also the only algorithm that detected other additional regions such as the thalamus and basal ganglia.

FS results

Atrophy was detected in the mesial temporal lobe and in isocortical areas at an FDR corrected threshold of $q < 0.05$ (Fig. 4). Thinning of the isocortex involved posterior cingulate/precuneus on the mesial surface and posterior parietal and temporal lobes on the lateral surface. The different smoothing levels had a clear impact on the FS results (Fig. 4). While using 10- and 15-mm FWHM smoothing filters showed small, patchy areas of abnormality, 20-, 25-, 30-, and 35-mm kernels presented cortical thinning as a progressively more confluent change. At an uncorrected threshold

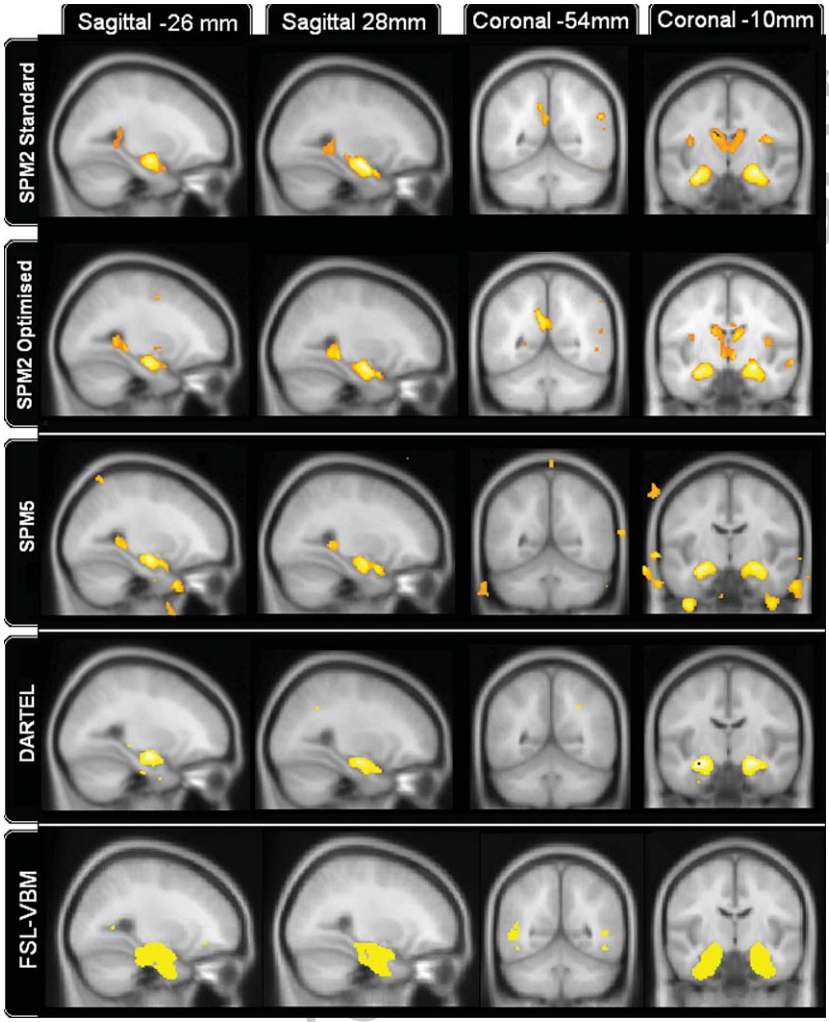

Fig. 2. Statistical maps of atrophy in AD illustrating VBM-version dependence across SPM2 Standard, SPM2 Optimized, SPM5, DARTEL, and FSL-VBM. The sections of the brain presented here focus on illustrating the detection of hippocampal degeneration by the different VBM methods. Results are thresholded at FDR corrected $q < 0.05$.

of $p < 0.05$ (Supplementary Fig. 1), lesion distribution was similar to the corrected results (Fig. 4) but without the emergence of new areas to any appreciable degree. The impact of increasing the smoothing kernel on VBM is shown, for contrast, in Supplementary Fig. 2. The effect of error correction in the pial surface is shown in Supplementary Fig. 3. Overall, no significant difference to the results was seen from these manipulations.

Uncorrected results of VBM

As would be expected, uncorrected VBM results ($p < 0.05$) yielded more extensive statistical differences (Fig. 5). Additional areas appeared: in SPM2 (standard and optimized), precuneus, thalamus, third and lateral ventricles, lateral temporo-parietal lobes, and more extensive mesial temporal changes were found; SPM5 detected the precuneus, the superior parietal lobe, the ventricles, the putamen, the insula, and extensive regions surrounding the outside of the brain boundary; DARTEL found more extensive temporal lobe abnormalities as well as posterior cingulate/precuneus and superior parietal lobe; FSL-VBM found some changes in posterior lateral temporal and occipital lobes, in the posterior cingulate cortex and in the frontal lobe. As SPM99 failed to identify any significant changes with FDR correction, it was not investigated further with an uncorrected analysis.

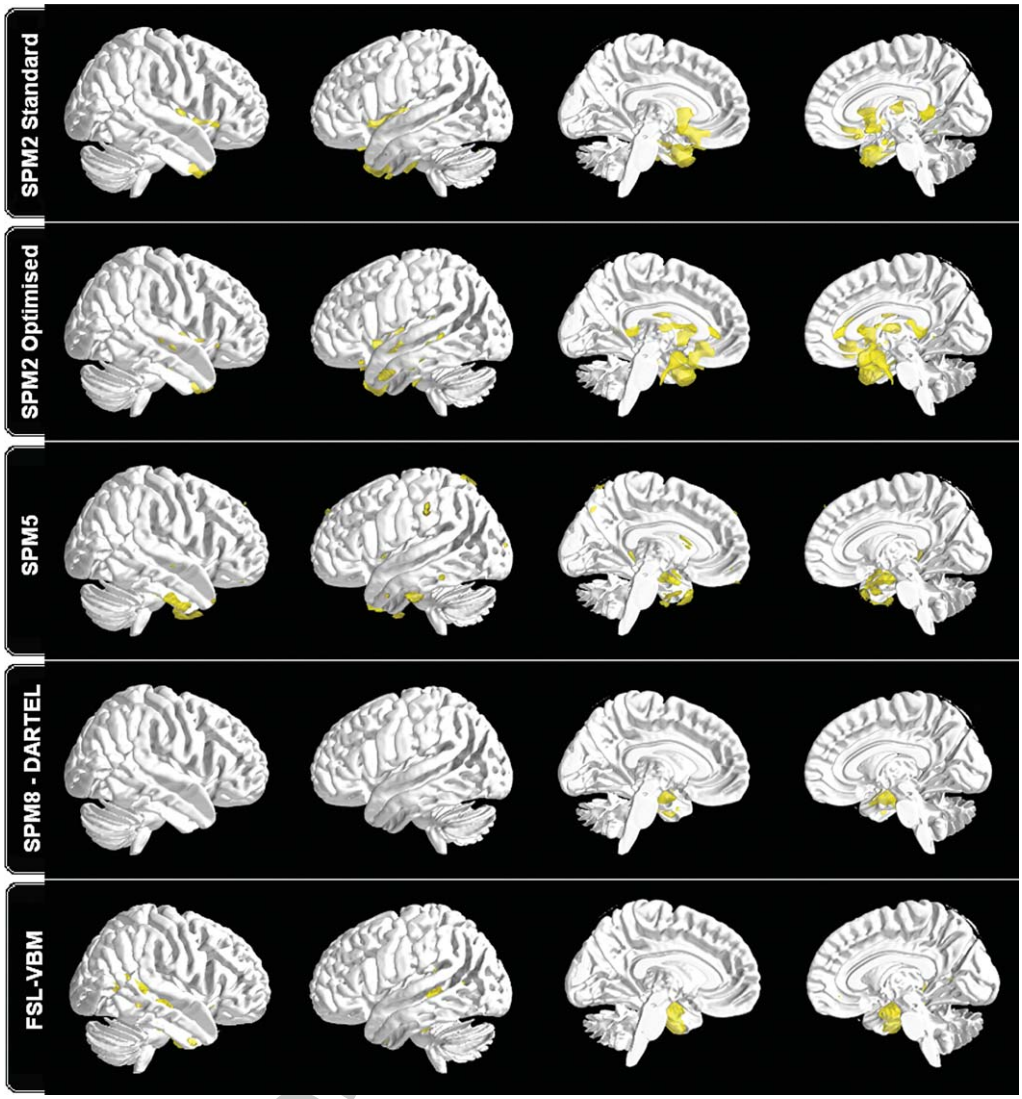

Fig. 3. Statistical maps of atrophy in AD illustrating VBM-version dependence across SPM2 Standard, SPM2 Optimized, SPM5, DARTEL, and FSL-VBM. The sections of the brain presented here focus on illustrating the detection of isocortical degeneration by the different VBM methods. Results are thresholded at FDR corrected $q < 0.05$.

Replication with ADNI dataset

Results for the replication study using the ADNI cohort are presented in the Supplementary Material.

DISCUSSION

The results of this study highlight that with different voxel-based algorithms there are commonalities but also some striking differences in the pattern of atrophy detection in AD. The results obtained with the local cohort showed that all VBM methods were good at detecting hippocampal atrophy, with the exception of SPM99. The hypothesis that VBM methods are insensitive to isocortical atrophy as confirmed in that none of these analyses revealed anything other than minor changes in the cortical ribbon—some nothing at all—whereas FS revealed a pattern that was predicted from FDG-PET [16, 17, 20, 21]. Similarly, the ADNI data also demonstrated that all VBM methods, with the exception of SPM99, could identify hippocampal atrophy at a corrected statistical threshold (Supplementary Fig. 4). ADNI data from SPM2 onwards, however, also identified cortical ribbon atrophy—albeit with a

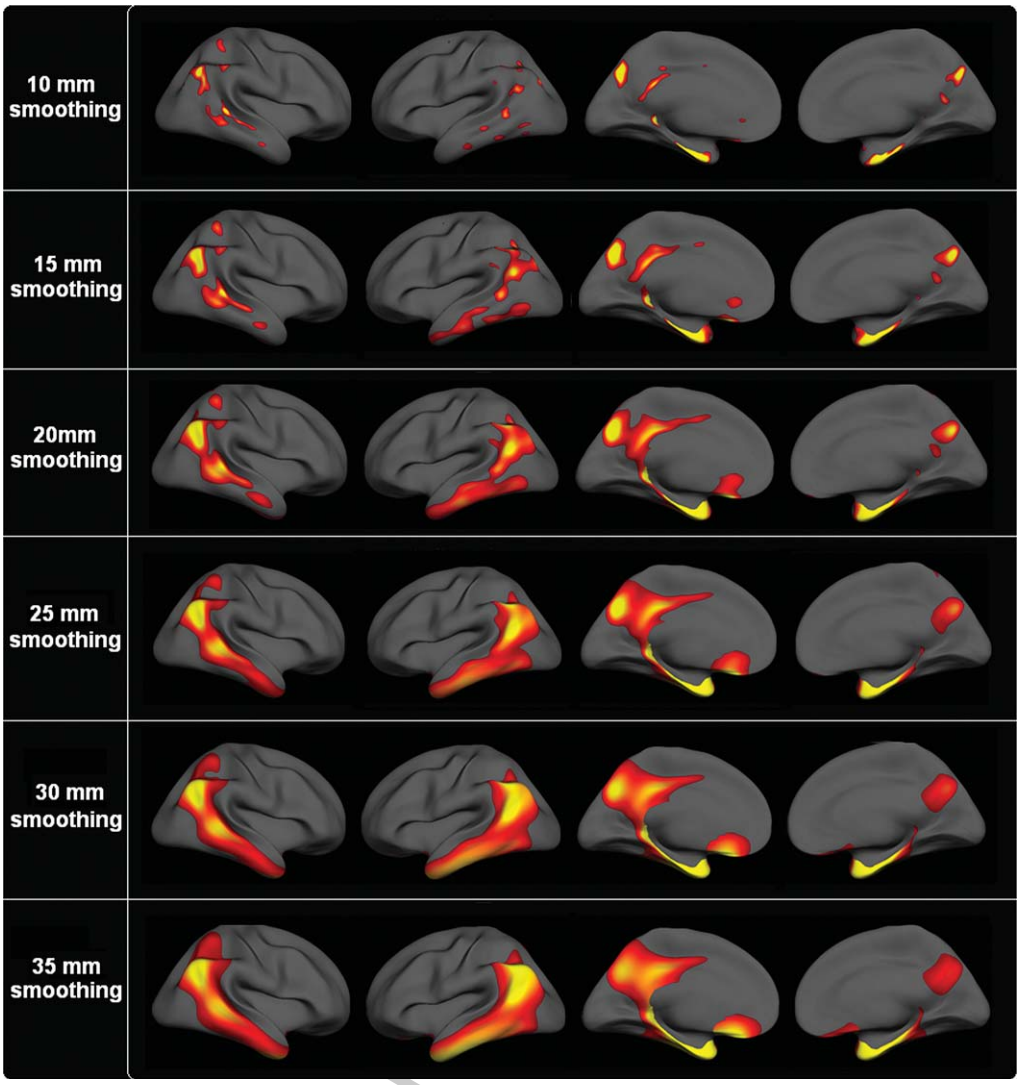

Fig. 4. Distribution of statistically significant differences in cortical thickness between subjects with AD and age-matched controls applying six different widths of Gaussian kernel (10-, 15-, 20-, 25-, 30-, and 35-mm FWHM). Results are shown at an FDR corrected threshold ($q < 0.05$). Red (low)/yellow (high) indicates statistical significance, i.e., regions where cortex is thinner in AD than in controls.

weaker statistical power to that obtained for the hippocampus (Supplementary Fig. 6). With the ADNI data, FS again revealed the same pattern of isocortical plus hippocampal atrophy as with the local cohort. The ADNI FS analysis was again concordant with that expected from prior knowledge in that posterior association cortex was most affected, dorso-lateral prefrontal cortex was affected but to a lesser degree, and there was preservation of primary motor, somato-sensory, auditory, and visual cortices (Supplementary Fig. 5).

Going through the VBM results in detail, these indicate that there was a major step up in sensitivity from SPM99 to later versions for detecting mesial temporal lobe atrophy. The first clinical VBM studies used SPM99, so in light of the present results, their findings probably should be considered as not directly comparable to studies using later VBM methods. With the local cohort, the results indicate that from SPM2 onwards, VBM has been very good at detecting mesial temporal lobe atrophy, though curiously, subsequent versions have become progressively less sensitive at detecting changes in the cortical ribbon. By the time one reaches what is, arguably, the current state-of-the-art—DARTEL—the results appear to be extremely 'cleaned up' in that the statistical effects in AD with FDR correction are restricted to the hippocampus. The

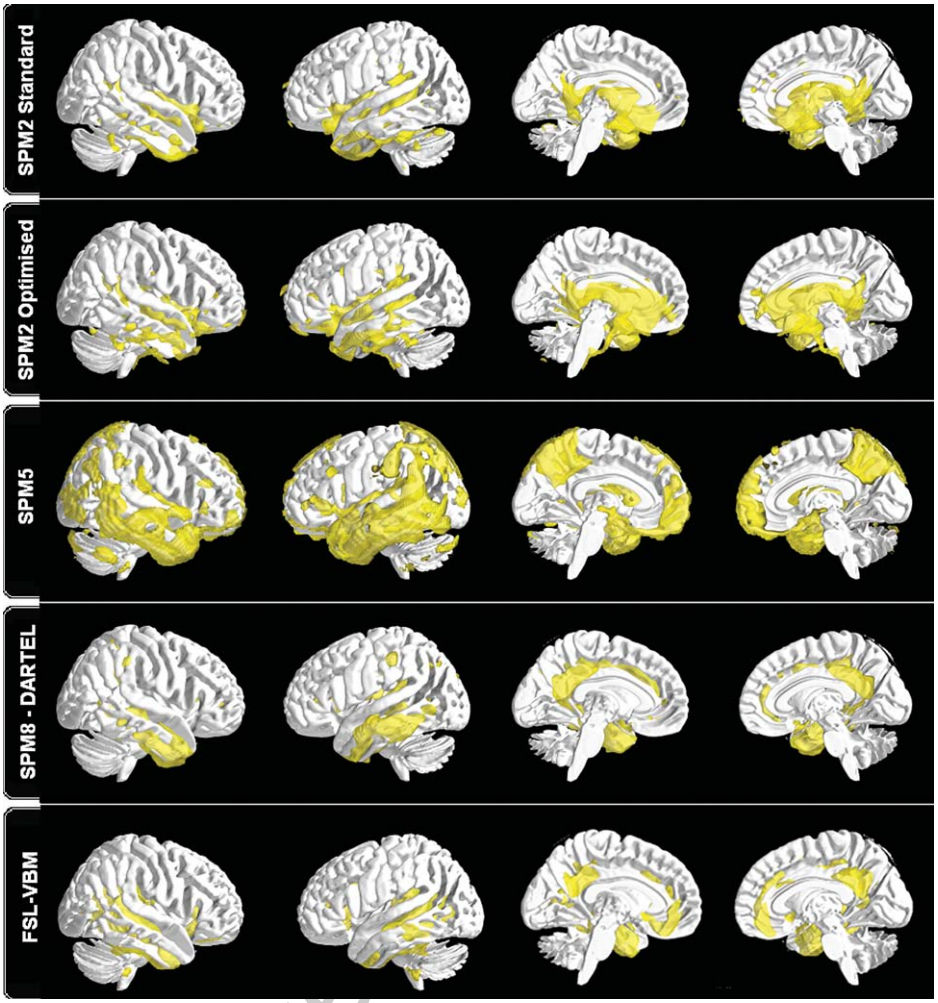

Fig. 5. Statistical maps of atrophy in AD illustrating VBM-version dependence across SPM2 Standard, SPM2 Optimized, SPM5, DARTEL, and FSL-VBM. The sections of the brain presented here focus on illustrating the detection of isocortical degeneration by the different VBM methods. Results are thresholded at an uncorrected level of $p < 0.05$.

VBM analyses were also scrutinized at an uncorrected threshold of 0.05. Clearly, such statistical lenience would be unacceptable in a clinical study, but was done here to explore whether a plausible pattern of cortical atrophy would emerge that did not meet the FDR threshold. This was true to the extent that from SPM2 onwards, posterior cingulate abnormalities were identified. In terms of a lateral temporo-parietal pattern, SPM2 came the closest, but generally speaking all versions produced a lot of additional noise at an uncorrected threshold. The ADNI dataset diverged somewhat from the local analysis in that more isocortical changes were identified. Examining the various VBM versions, however, yielded an even more disturbing observation than the false-negatives in the local dataset analysis: the pattern of cortical atrophy (Supplementary Fig. 6) varied enormously between different versions. SPM2 found a relatively focal pattern of rostral temporal and orbito-frontal change that looked more reminiscent of frontotemporal dementia than AD; SPM5 found predominantly lateral temporo-parieto-occipital changes; DARTEL again identified more intense but the most localized change, though this time also involving the lateral as well as mesial temporal lobe; FSL-VBM found caudal, without rostral, temporal lobe and dorsal frontal lobe changes. It is important to stress that these changes were all found using FDR correction for multiple comparisons, yet in spite of this, results were markedly inconsistent across VBM platforms.

The discrepancies between local and ADNI data with VBM also merit some discussion. Obviously, there is always some heterogeneity between datasets that can explain some inconsistency. Additional factors that may possibly contribute are that the ADNI cases were significantly older; ADNI data was acquired at a lower field strength; and the ADNI data in this study used a non-isotropic resolution with the same in-plane resolution as the local data but a marginally narrower slice thickness (1.2 mm versus 1.25 mm).

FreeSurfer, in contrast, produced the expected pattern of posterior association cortex thinning with both datasets—local (Fig. 4) and ADNI (Supplementary Fig. 5)—only with more marked findings using the latter. This result was entirely consistent with the pattern of cortical metabolic abnormalities seen with FDG-PET [16, 17, 20, 21]; it was also highly consistent with past FreeSurfer studies in AD [9, 43]. In addition, it must be noted that this cortical pattern can be seen with VBM, provided a study is highly powered, e.g., Whitwell et al. [44] who contrasted $n = 72$ AD patients (mean MMSE = 21/30) with $n = 72$ controls. The point remains, though, that within any given cohort, VBM will preferentially identify mesial temporal lobe changes over changes in the isocortex.

A few technical observations are worthy of discussion regarding the FreeSurfer analysis. The first was that, consistent with previous observations [39, 42], FreeSurfer commonly produced errors in delineating the cortical ribbon, particularly the pial surface. Somewhat surprisingly, however, correcting these errors, either by an automated skull-stripping approach or by a full manual correction, did not lead to any great improvement in sensitivity. Furthermore, this was in spite of the intriguing observation that manual correction of errors even led to improvements in delineating the cortical ribbon, remote from the sites of manual editing (Fig. 1). Presumably, this lack of improvement is because such errors are somewhat unique in terms of spatial location, rather than systematically affecting a particular area in each subject and, therefore, do not exert an effect on the overall results. The second technical detail of interest was that the default smoothing level in the FreeSurfer package (10-mm FWHM) yielded only patchy, biologically implausible results for AD—observed with both datasets. In order to visualize the confluent posterior association cortex lesion that is expected in AD, very large kernels are necessary, at least for studies of comparable power to the present. Although, the aim of the present study was to examine the types of VBM methods used in past clinical studies, out of interest we examined the effect of very large kernels on VBM as well (Supplementary Fig. 2). For the most part, this only blurred existing blobs using standard kernels outside the brain.

The ramifications of the present study are not limited to the technical aspect of showing that two different styles of automated image analysis have differential sensitivity for atrophy detection in AD; rather that they have potentially important implications for theories regarding the temporal evolution of neurodegeneration in this disease. Whereas early mesial temporal lobe atrophy is a well-established finding in AD, metabolic studies with FDG-PET identify posterior cingulate cortex hypometabolism as the earliest and most significant abnormality [18–20]. A diaschisis hypothesis—often also referred to as a "disconnection" hypothesis—has been proposed to reconcile the findings of structural and metabolic data [19, 45–51]. It proposes that cortical hypometabolism is primarily a consequence of synapse loss from neurons originating in the mesial temporal lobe. Evidence for the diaschisis hypothesis is essentially negative in the sense that it is predicated on a failure to detect posterior cingulate atrophy of a commensurate level to the metabolic lesion in the posterior cingulate. The technique used to interrogate atrophy patterns in such studies was, however, VBM, and, as shown in the present study, this method is either grossly insensitive to cortical ribbon atrophy, implausible or inconsistent between methods while, at the same time, being exquisitely sensitive to mesial temporal lobe atrophy; it is, in other words, not as unbiased as its automated methodology would imply. Whereas posterior cingulate atrophy detection with VBM has been rather inconsistent, other methods such as manual volumetry [52] or serial fluid registration [53, 54] have found this area to be site of very early and focal volume loss.

We propose that VBM is, in fact, biased toward mesial temporal lobe atrophy detection. As discussed in the introduction, the hippocampus is a relatively bulky grey matter structure that is similar in gross morphology in all humans—two features that probably make it ideal for detecting change in degenerative disease with a VBM-style approach. In contrast, the cortical ribbon is thin, and the precise pattern of gyral convolutions at any given point is relatively idiosyncratic. These factors, combined with the previous observation that the degree of registration error varies across different brain regions and different degenerative disease groups [2] are likely to conspire to mean that VBM is relatively insensitive to atrophy of the cortical ribbon.

Finally, the present results have important implications for both clinical and cognitive neuroscience in AD. Consider a hypothetical study where researchers were testing the hypothesis that a particular phenomenon occurred in AD at a time when neurodegeneration was restricted to the mesial temporal lobe (and therefore was a consequence of such degeneration). Testing such a hypothesis with, for instance, a DARTEL VBM analysis of the current dataset would perfectly, yet spuriously, have been confirmed. Similarly, a DARTEL analysis of the ADNI data that hypothesized a phenomenon was underpinned by preservation of the pre-frontal cortex and degeneration of the temporal lobe would have suffered the same fate.

ACKNOWLEDGMENTS

This work was funded by a Medical Research Council grant to PJN. LZD-d-G was supported by a Pinsent-Darwin scholarship and JA-C by a fellowship from Alzheimer's Research UK (ARUK). The study was also supported by the National Institute of Health Research, Cambridge Biomedical Research Centre.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=1893).

SUPPLEMENTARY MATERIAL

Supplementary figures and tables are available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-130362.

REFERENCES

[1] Ashburner J, Friston KJ (2000) Voxel-based morphometry–the methods. *Neuroimage* **11**, 805-821.

[2] Pereira JM, Xiong L, Acosta-Cabronero J, Pengas G, Williams GB, Nestor PJ (2010) Registration accuracy for VBM studies varies according to region and degenerative disease grouping. *Neuroimage* **49**, 2205-2215.

[3] Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* **14**, 21-36.

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

[5] Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* **8**, 272-284.

[6] Baron JC, Chetelat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F (2001) *In vivo* mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage* **14**, 298-309.

[7] Frisoni GB, Testa C, Zorzan A, Sabattoli F, Beltramello A, Soininen H, Laakso MP (2002) Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. *J Neurol Neurosurg Psychiatry* **73**, 657-664.

[8] Kim S, Youn YC, Hsiung GY, Ha SY, Park KY, Shin HW, Kim DK, Kim SS, Kee BS (2010) Voxel-based morphometric study of brain volume changes in patients with Alzheimer's disease assessed according to the Clinical Dementia Rating score. *J Clin Neurosci* **18**, 916-921.

[9] Lehmann M, Douiri A, Kim LG, Modat M, Chan D, Ourselin S, Barnes J, Fox NC (2009) Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *Neuroimage* **49**, 2264-2274.

[10] Ohnishi T, Matsuda H, Tabira T, Asada T, Uno M (2001) Changes in brain morphology in Alzheimer disease and normal aging: Is Alzheimer disease an exaggerated aging process? *AJNR Am J Neuroradiol* **22**, 1680-1685.

[11] Pennanen C, Testa C, Laakso MP, Hallikainen M, Helkala EL, Hanninen T, Kivipelto M, Kononen M, Nissinen A, Tervo S, Vanhanen M, Vanninen R, Frisoni GB, Soininen H (2005) A voxel based morphometry study on mild cognitive impairment. *J Neurol Neurosurg Psychiatry* **76**, 11-14.

[12] Rombouts SA, Barkhof F, Witter MP, Scheltens P (2000) Unbiased whole-brain analysis of gray matter loss in Alzheimer's disease. *Neurosci Lett* **285**, 231-233.

[13] Senjem ML, Gunter JL, Shiung MM, Petersen RC, Jack CR Jr (2005) Comparison of different methodological implementations of voxel-based morphometry in neurodegenerative disease. *Neuroimage* **26**, 600-608.

[14] Shiino A, Akiguchi I, Watanabe T, Shirakashi Y, Nozaki K, Tooyama I, Inubushi T (2012) Morphometric characterization of Binswanger's disease: Comparison with Alzheimer's disease. *Eur J Radiol* **81**, 2375-2379.

[15] Villain N, Desgranges B, Viader F, de la Sayette V, Mezenge F, Landeau B, Baron JC, Eustache F, Chetelat G (2008) Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in Alzheimer's disease. *J Neurosci* **28**, 6174-6181.

[16] Chase TN, Foster NL, Fedio P, Brooks R, Mansi L, Di Chiro G. (1984) Regional cortical dysfunction in Alzheimer's disease as determined by positron emission tomography. *Ann Neurol* **15** (Suppl), S170-S174.

[17] Duara R, Grady C, Haxby J, Sundaram M, Cutler NR, Heston L, Moore A, Schlageter N, Larson S, Rapoport SI (1986) Positron emission tomography in Alzheimer's disease. *Neurology* **36**, 879-887.

[18] Minoshima S, Foster NL, Kuhl DE (1994) Posterior cingulate cortex in Alzheimer's disease. *Lancet* **344**, 895.

[19] Minoshima S, Giordani B, Berent S, Frey KA, Foster NL, Kuhl DE (1997) Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neurol* **42**, 85-94.

[20] Nestor PJ, Fryer TD, Smielewski P, Hodges JR (2003) Limbic hypometabolism in Alzheimer's disease and mild cognitive impairment. *Ann Neurol* **54**, 343-351.

[21] Reiman EM, Caselli RJ, Yun LS, Chen K, Bandy D, Minoshima S, Thibodeau SN, 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.

[22] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (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.

[23] Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR (2006) The Addenbrooke's Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* **21**, 1078-1085.
[24] Bookstein FL (2001) "Voxel-based morphometry" should not be used with imperfectly registered images. *Neuroimage* **14**, 1454-1462.
[25] Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* **26**, 839-851.
[26] Smith SM (2002) Fast robust automated brain extraction. *Hum Brain Mapp* **17**, 143-155.
[27] Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* **20**, 45-57.
[28] Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825-841.
[29] Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. *Med Image Anal* **5**, 143-156.
[30] Andersson JLR, Jenkinson M, Smith S (2007) TR07JA2: Non-linear registration, aka Spatial normalisation. *FMRIB Group Analysis Technical Report*, http://www.fmrib.ox.ac.uk/analysis/techrep/tr07ja2/tr07ja2.pdf.
[31] Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194.
[32] Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A* **97**, 11050-11055.
[33] Fischl B, Salat DH, 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 AM (2002) Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-355.
[34] Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT, Dale AM (2004) Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**(Suppl 1), S69-S84.
[35] Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195-207.
[36] Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004) Automatically parcellating the human cerebral cortex. *Cereb Cortex* **14**, 11-22.
[37] 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.
[38] Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B (2004) A hybrid approach to the skull stripping problem in MRI. *Neuroimage* **22**, 1060-1075.
[39] Kempton MJ, Underwood TS, Brunton S, Stylios F, Schmechtig A, Ettinger U, Smith MS, Lovestone S, Crum WR, Frangou S, Williams SC, Simmons A (2011) A comprehensive testing protocol for MRI neuroanatomical segmentation techniques: Evaluation of a novel lateral ventricle segmentation method. *Neuroimage* **58**, 1051-1059.
[40] Acosta-Cabronero J, Williams GB, Pereira JM, Pengas G, Nestor PJ (2008) The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. *Neuroimage* **39**, 1654-1665.
[41] Lerch JP, Evans AC (2005) Cortical thickness analysis examined through power analysis and a population simulation. *Neuroimage* **24**, 163-173.
[42] Lehmann M, Crutch SJ, Ridgway GR, Ridha BH, Barnes J, Warrington EK, Rossor MN, Fox NC (2011) Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. *Neurobiol Aging* **32**, 1466-1476.
[43] Du AT, Schuff N, Kramer JH, Rosen HJ, Gorno-Tempini ML, Rankin K, Miller BL, Weiner MW (2007) Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. *Brain* **130**, 1159-1166.
[44] Whitwell JL, Weigand SD, Shiung MM, Boeve BF, Ferman TJ, Smith GE, Knopman DS, Petersen RC, Benarroch EE, Josephs KA, Jack CR Jr (2007) Focal atrophy in dementia with Lewy bodies on MRI: A distinct pattern from Alzheimer's disease. *Brain* **130**, 708-719.
[45] Chetelat G, Desgranges B, de la Sayette V, Viader F, Eustache F, Baron JC (2003) Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? *Neurology* **60**, 1374-1377.
[46] Chetelat G, Landeau B, Eustache F, Mezenge F, Viader F, de la Sayette V, Desgranges B, Baron JC (2005) Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *Neuroimage* **27**, 934-946.
[47] Jobst KA, Smith AD, Barker CS, Wear A, King EM, Smith A, Anslow PA, Molyneux AJ, Shepstone BJ, Soper N, et al. (1992) Association of atrophy of the medial temporal lobe with reduced blood flow in the posterior parietotemporal cortex in patients with a clinical and pathological diagnosis of Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **55**, 190-194.
[48] Meguro K, Blaizot X, Kondoh Y, Le Mestric C, Baron JC, Chavoix C (1999) Neocortical and hippocampal glucose hypometabolism following neurotoxic lesions of the entorhinal and perirhinal cortices in the non-human primate as shown by PET. Implications for Alzheimer's disease. *Brain* **122**(Pt 8), 1519-1531.
[49] Meguro K, LeMestric C, Landeau B, Desgranges B, Eustache F, Baron JC (2001) Relations between hypometabolism in the posterior association neocortex and hippocampal atrophy in Alzheimer's disease: A PET/MRI correlative study. *J Neurol Neurosurg Psychiatry* **71**, 315-321.
[50] Nestor PJ, Fryer TD, Ikeda M, Hodges JR (2003) Retrosplenial cortex (BA 29/30) hypometabolism in mild cognitive impairment (prodromal Alzheimer's disease). *Eur J Neurosci* **18**, 2663-2667.
[51] Small GW, Ercoli LM, Silverman DH, Huang SC, Komo S, Bookheimer SY, Lavretsky H, Miller K, Siddarth P, Rasgon NL, Mazziotta JC, Saxena S, Wu HM, Mega MS, Cummings JL, Saunders AM, Pericak-Vance MA, Roses AD, Barrio JR, Phelps ME (2000) Cerebral metabolic and cognitive decline in persons at genetic risk for Alzheimer's disease. *Proc Natl Acad Sci U S A* **97**, 6037-6042.
[52] Pengas G, Hodges JR, Watson P, Nestor PJ (2010) Focal posterior cingulate atrophy in incipient Alzheimer's disease. *Neurobiol Aging* **31**, 25-33.
[53] Buckner RL, Snyder AZ, Shannon BJ, LaRossa G, Sachs R, Fotenos AF, Sheline YI, Klunk WE, Mathis CA, Morris

JC, Mintun MA (2005) Molecular, structural, and functional characterization of Alzheimer's disease: Evidence for a relationship between default activity, amyloid, and memory. *J Neurosci* **25**, 7709-7717.

[54] Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC (2002) Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. *Proc Natl Acad Sci U S A* **99**, 4703-4707.