NeuroImage 58 (2011) 818–828

Contents lists available at ScienceDirect

NeuroImage

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

AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America

Eric Westman [a,*], Andrew Simmons [b,c], J-Sebastian Muehlboeck [d], Patrizia Mecocci [f], Bruno Vellas [g], Magda Tsolaki [h], Iwona Kłoszewska [i], Hilkka Soininen [e], Michael W. Weiner [j], Simon Lovestone [b,c], Christian Spenger [d], Lars-Olof Wahlund [a]
for the AddNeuroMed consortium and the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
[b] King's College London, Institute of Psychiatry, London, UK
[c] NIHR Biomedical Research Centre for Mental Health, London, UK
[d] Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden
[e] Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
[f] Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy
[g] INSERM U 558, University of Toulouse, Toulouse, France
[h] Aristotle University of Thessaloniki, Thessaloniki, Greece
[i] Medical University of Lodz, Lodz, Poland
[j] VA Medical Center, University of California, San Francisco, USA

ARTICLE INFO

Article history:
Received 14 February 2011
Revised 21 June 2011
Accepted 23 June 2011
Available online 1 July 2011

Keywords:
AddNeuroMed
ADNI
MRI
OPLS
AD
MCI

ABSTRACT

The European Union AddNeuroMed program and the US-based Alzheimer Disease Neuroimaging Initiative (ADNI) are two large multi-center initiatives designed to collect and validate biomarker data for Alzheimer's disease (AD). Both initiatives use the same MRI data acquisition scheme. The current study aims to compare and combine magnetic resonance imaging (MRI) data from the two study cohorts using an automated image analysis pipeline and a multivariate data analysis approach. We hypothesized that the two cohorts would show similar patterns of atrophy, despite demographic differences and could therefore be combined. MRI scans were analyzed from a total of 1074 subjects (AD = 295, MCI = 444 and controls = 335) using Freesurfer, an automated segmentation scheme which generates regional volume and regional cortical thickness measures which were subsequently used for multivariate analysis (orthogonal partial least squares to latent structures (OPLS)). OPLS models were created for the individual cohorts and for the combined cohort to discriminate between AD patients and controls. The ADNI cohort was used as a replication dataset to validate the model created for the AddNeuroMed cohort and *vice versa*. The combined cohort model was used to predict conversion to AD at baseline of MCI subjects at 1 year follow-up. The AddNeuroMed, the ADNI and the combined cohort showed similar patterns of atrophy and the predictive power was similar (between 80 and 90%). The combined model also showed potential in predicting conversion from MCI to AD, resulting in 71% of the MCI converters (MCI-c) from both cohorts classified as AD-like and 60% of the stable MCI subjects (MCI-s) classified as control-like. This demonstrates that the methods used are robust and that large data sets can be combined if MRI imaging protocols are carefully aligned.

Introduction

Alzheimer's disease is the most common form of neurodegenerative disorder and is characterized by gradual loss of cognitive functions such as episodic memory. The disease is related to pathological amyloid depositions and hyperphosphorylation of structural proteins in the brain which lead to progressive loss of function, metabolic alterations and structural changes in the brain.

Translational biomarkers in Alzheimer's disease based on non-invasive *in vivo* methods are highly desirable to monitor disease progression and treatment. Magnetic Resonance Imaging (MRI),

* Corresponding author at: Karolinska Universitetssjukhuset, Novum, Plan 5, 141 86 Stockholm, Sweden. Fax: +46 8 517 761 11.
E-mail address: eric.westman@ki.se (E. Westman).

[1] Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Authorship_List.pdf).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.06.065

Positron Emission Tomography (PET), and CSF measures all allow different aspects of AD pathology to be studied. Magnetic resonance imaging (volumetric and cortical thickness measures of atrophy) is a non-invasive technique, applicable *in vivo* in both humans and experimental animals.

In recent years two large multi-center studies have been carried out for the evaluation and establishment of markers for early diagnosis, monitoring disease progression and the detection of pharmaceutical treatment effects — the European based AddNeuroMed program and the US-based Alzheimer's disease Neuroimaging Initiative (ADNI). Both studies have collected magnetic resonance data from multiple sites across Europe and North America respectively. The AddNeuroMed MRI data acquisition was designed to be compatible with ADNI (Jack et al., 2008) so that comparisons between the two studies could be made.

An ideal MR marker would detect a fundamental feature of AD neuropathology, be diagnostically sensitive and specific (validated through neuropathology) and produce accurate and reproducible results (Kantarci, 2005). Several groups including our own have proposed the use of multivariate techniques for analyzing multiple regional measures from MRI to aid diagnosis of Alzheimer's disease and to predict future conversion from the prodromal stage of Alzheimer's disease (mild cognitive impairment (MCI)) to Alzheimer's disease itself (McEvoy et al., 2009; Westman et al., 2011). Results from different studies are sometimes hard to compare due to variations in the number of subjects studied, ethnic origin, age, gender, years of education and country/continent of recruitment etc. We believed that despite demographic differences such as these it would be possible to combine two large data sets if the MR-protocols were carefully aligned. We assumed that regardless of the demographic nature of the two cohorts (AddNeuroMed and ADNI), they would show similar patterns of atrophy and could therefore be combined.

The aims of our study were (1) To investigate if multivariate classification techniques trained on one of the cohorts would give similar results when applied to the other, (2) To assess whether patterns of atrophy would be similar for the two cohorts, when comparing AD and healthy controls (CTL), (3) To investigate if comparable classification sensitivities and specificities would be obtained for the two cohorts, and (4) To assess the accuracy of prediction of conversion from MCI to AD after 1 year using the combined AddNeuroMed/ADNI cohort.

Material and methods

Data

The datasets used here were obtained from two large multi center cohorts, namely AddNeuroMed and the Alzheimer's disease Neuroimaging Initiative (ADNI). AddNeuroMed, a part of InnoMed, (Innovative Medicines in Europe) is an Integrated Project funded by the European Union Sixth Framework program (Lovestone et al., 2007, 2009). AddNeuroMed aims to develop and validate novel surrogate markers of disease and treatment, based upon *in vitro* and *in vivo* models in animals and humans in Alzheimer's disease (AD). The neuroimaging part of AddNeuroMed uses magnetic resonance imaging (MRI) and magnetic resonance spectroscopy (MRS) to establish imaging markers for early diagnosis and detection of disease and efficacy of disease modifying therapy in man, as well as translational imaging biomarkers in animal models of AD. Human data was collected from six different sites across Europe; University of Kuopio, Finland, University of Perugia, Italy, Aristotle University of Thessaloniki, Greece, King's College London, United Kingdom, University of Lodz, Poland and University of Toulouse, France (Lovestone et al., 2009; Simmons et al., 2009, 2011).

The ADNI dataset was downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI, PI Michael M. Weiner). ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET) and other biological markers are useful in clinical trials of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. ADNI subjects aged 55 to 90 from over 50 sites across the U.S. and Canada participated in the research and more detailed information is available at www.adni-info.org.

Inclusion and diagnostic criteria

A total of 1074 subjects are included in the current study (AD = 295, MCI = 444 and CTL = 335). The demographics of the cohorts are given in Table 2. Although both studies were longitudinal in nature only baseline MRI data from the two cohorts were used in this study. Future analyses will focus on longitudinal analyses. For the AddNeuroMed cohort all AD and MCI subjects were recruited from the local memory clinics of one of the six participating sites while the control subjects were recruited from non-related members of the patient's families, caregiver's relatives and social centers for the elderly or GP surgeries. Informed consent was obtained where the research participant had capacity, and in those cases where dementia compromised capacity then assent from the patient and consent from a relative, according to local law and process, was obtained. This study was approved by ethical review boards in each participating country. The inclusion and exclusion criteria were as follows.

Alzheimer's disease (*Inclusion criteria*): 1) ADRDA/NINCDS and DSM-IV criteria for probable Alzheimer's disease. 2) Mini Mental State Examination score ranged from 12 to 28. 4) Age 65 years or above. *Exclusion criteria*: 1) Significant neurological or psychiatric illness other than Alzheimer's disease. 2) Significant unstable systematic illness or organ failure.

Mild Cognitive Impairment and Controls (*Inclusion criteria*): 1) Mini Mental State Examination score range between 24 and 30. 2) Geriatric Depression Scale score less than or equal to 5. 3) Age 65 years or above. 4) Medication stable. 5) Good general health. *Exclusion criteria*: 1) Meet the DSM-IV criteria for Dementia. 2) Significant neurological or psychiatric illness other than Alzheimer's disease. 3) Significant unstable systematic illness or organ failure.

The distinction between MCI and controls was based on two criteria: 1) subject scores 0 on Clinical Dementia Rating Scale = control. 2) Subject scores 0.5 on Clinical Dementia Rating scale = MCI. For the MCI subjects it was preferable that the subject and informant reported occurrence of memory problems. All AD subjects had a CDR score of 0.5 or above.

CDR, Mini-Mental State, and CERAD Cognitive Battery were assessed for each subject. The CERAD Cognitive Battery was replaced with the Alzheimer's Disease Assessment Scale (ADAS–Cog) for the AD subjects. This cognitive test battery is specially designed for AD trials (Rosen et al., 1984). Both the ADAS-Cog and the CERAD battery use the same 10-word recall task, the only difference is that the scoring is inverted. The mean number of words not recalled in the CERAD word list immediate recall task was calculated. The variable obtained was named ADAS1, corresponding to the first subtest of ADAS-Cog. This was performed to obtain comparable measures between groups.

For the ADNI cohort a detailed description of the inclusion criteria can be found on the ADNI webpage (http://www.adni-info.org/Scientists/AboutADNI.aspx#). Subjects were between 55 and 90 years of age. They had a study partner able to provide an independent evaluation of functioning, and spoke either English or Spanish. All subjects were willing and able to undergo all test procedures including neuroimaging and agreed to longitudinal follow up. Specific psychoactive medications were excluded.

Alzheimer's disease (*General inclusion/exclusion criteria*): 1) MMSE scores between 20 and 26. 2) CDR of 0.5 or 1.0. 3) met NINCDS/ADRDA criteria for probable AD.

Mild cognitive impairment (*General inclusion/exclusion criteria*): 1) subjects had MMSE scores between 24 and 30 (inclusive). 2) memory complaint, had objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II. 3) CDR of 0.5. 4) absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia.

Controls (*General inclusion/exclusion criteria*): 1) MMSE scores between 24 and 30 (inclusive), 2) CDR of zero, 3) They were non-depressed, non MCI, and non-demented.

MRI

Data acquisition for the AddNeuroMed study was designed to be compatible with the Alzheimer Disease Neuroimaging Initiative (ADNI). (Jack et al., 2008) The imaging protocol for both studies included a high resolution sagittal 3D T1-weighted MPRAGE volume (voxel size $1.1 \times 1.1 \times 1.2$ mm^3) and axial proton density/T2-weighted fast spin echo images. The MPRAGE volume was acquired using a custom pulse sequence specifically designed for the ADNI study to ensure compatibility across scanners (Jack et al., 2008). Full brain and skull coverage was required for both of the latter datasets and detailed quality control carried out on all MR images from both studies according to the AddNeuroMed quality control procedure (Simmons et al., 2009, 2011).

Regional volume segmentation and cortical thickness parcellation

We utilized the Freesurfer pipeline (version 4.5.0), which produces regional cortical thickness and volumetric measures. Cortical reconstruction and volumetric segmentation includes removal of non-brain tissue using a hybrid watershed/surface deformation procedure (Segonne et al., 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter volumetric structures (including hippocampus, amygdala, caudate, putamen, and ventricles) (Fischl et al., 2002, 2004a; Segonne et al., 2004) intensity normalization (Sled et al., 1998), tessellation of the gray matter white matter boundary, automated topology correction (Fischl et al., 2001; Segonne et al., 2007), and surface deformation following intensity gradients to optimally place the gray/white and gray/cerebrospinal fluid borders at the location where the greatest shift in intensity defines the transition to the other tissue class (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000). Once the cortical models are complete, registration to a spherical atlas takes place which utilizes individual cortical folding patterns to match cortical geometry across subjects (Fischl et al., 1999). This is followed by parcellation of the cerebral cortex into units based on gyral and sulcal structure (Desikan et al., 2006; Fischl et al., 2004b). The pipeline generated 68 cortical thickness measures (34 from each hemisphere) and 50 regional volumes. Volumes of white matter hypointensities, optic chiasm, right and left vessel, and left and right choroid plexus were excluded from further analysis. Cortical thickness and volumetric measures from the right and left side were averaged. In total 57 variables obtained from the pipeline were used as input variables for the OPLS classification, 34 regional cortical thickness measures and 23 regional volumes (Table 1). All volumetric measures from each subject were normalized by the subject's intracranial volume. This segmentation approach has been used for multivariate classification of Alzheimer's disease and healthy controls (Westman et al., 2010, 2011), neuropsychological-image analysis (Liu et al., 2009, 2010c), imaging-genetic analysis (Liu et al., 2010a, 2010b) and biomarker discovery (Thambisetty et al., 2010).

Multivariate data analysis

MRI measures were analyzed using orthogonal partial least squares to latent structures (OPLS) (Bylesjo et al., 2007; Johan Trygg, 2002; Rantalainen et al., 2006; Westman et al., 2010, 2011; Wiklund et al., 2008), a supervised multivariate data analysis method included in the software package SIMCA (Umetrics AB, Umea, Sweden). A very similar method, partial least squares to latent structures (PLS) has previously been used in several studies to analyze MR-data (Levine et al., 2008; McIntosh and Lobaugh, 2004; Oberg et al., 2007; Westman et al., 2007, 2009). OPLS and PLS give the same predictive accuracy, but the advantage of OPLS is that the model created to compare groups is rotated, which means that the information related to class separation is found in the first component of the model, the predictive component. The other orthogonal components in the model, if any, relate to variation in the data not connected to class separation. Focusing the information related to class separation on the first component makes data interpretation easier (Wiklund et al., 2008). OPLS also allows the generation of loading plots which illustrate the importance of the input variables in the classification (Fig. 2).

Table 1
Variable included in OPLS analysis.

Cortical thickness measures	Volumetric measures
Banks of superior temporal sulcus	Third ventricle
Caudal anterior cingulate	Fourth ventricle
Caudal middle frontal gyrus	Brainstem
Cuneus cortex	Corpus callosum anterior
Entorhinal cortex	Corpus callosum central
Fusiform gyrus	Corpus callosum midanterior
Inferior parietal cortex	Corpus callosum midposterior
Inferior temporal gyrus	Corpus callosum posterior
Isthmus of cingulate cortex	CSF
Lateral occipital cortex	Accumbens
Lateral orbitofrontal cortex	Amygdala
Lingual gyrus	Caudate
Medial orbitofrontal cortex	Cerebellum cortex
Middle temporal gyrus	Cerebellum white matter
Parahippocampal gyrus	Hippocampus
Paracentral sulcus	Inferior lateral ventricle
Frontal operculum	Putamen
Orbital operculum	Cerebral cortex
Triangular part of inferior frontal gyrus	Cerebral white matter
Pericalcarine cortex	Lateral ventricle
Postcentral gyrus	Pallidum
Posterior cingulate cortex	Thalamus proper
Precentral gyrus	Ventral DC
Precuneus cortex	
Rostral anterior cingulate cortex	
Rostral middle frontal gyrus	
Superior frontal gyrus	
Superior parietal gyrus	
Superior temporal gyrus	
Supramarginal gyrus	
Frontal pole	
Temporal pole	
Transverse temporal cortex	
Insular	
Banks of superior temporal sulcus	

57 variables in total included in OPLS analysis, 34 cortical thickness measures and 23 volumetric measures.

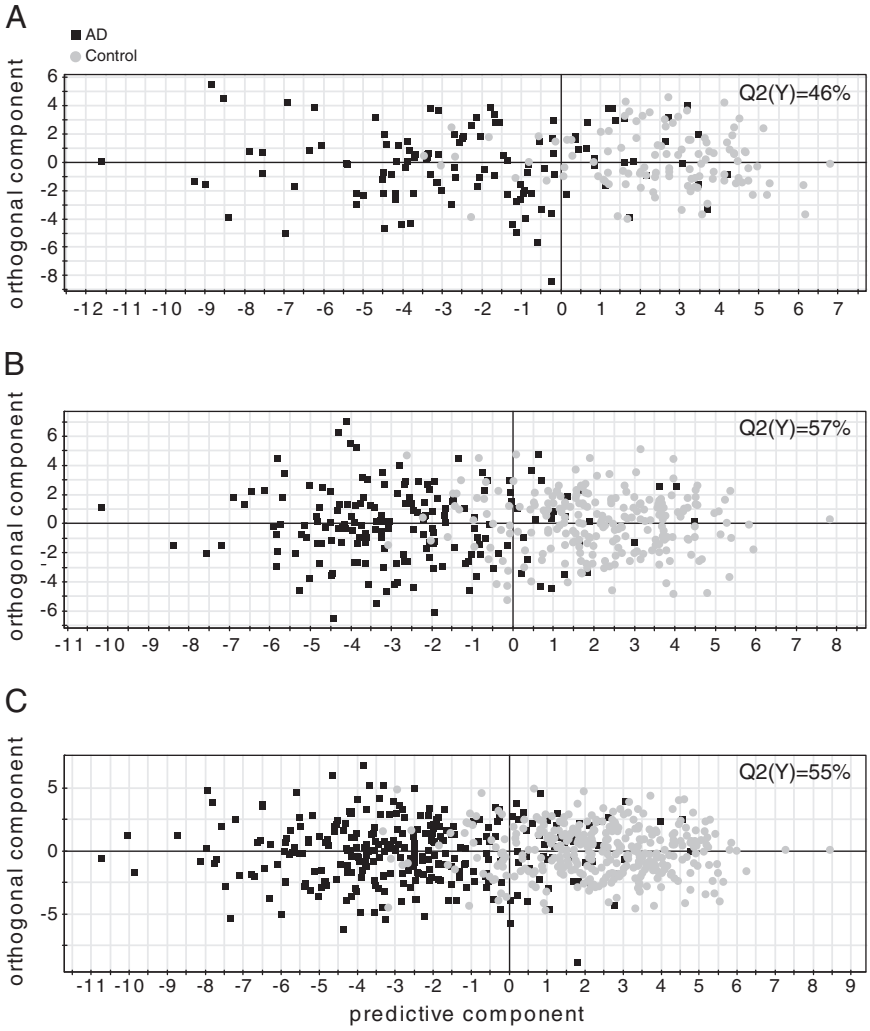

Fig. 1. OPLS cross validated score plots for (A) AddNeuroMed cohort (B) ADNI cohort (C) Combined cohort. The scatter plots visualize group separation and the predictability of the three different AD *vs.* CTL models. Each black circle represents an AD subject and each gray square a control subject. Control subjects to the left of zero and AD subjects to the right of zero are falsely predicted. $Q^2(Y)>0.05$ (a statistically significant model) and $Q^2(Y)>0.5$ (a good model).

Pre-processing was performed using mean centring and unit variance scaling. Mean centring improves the interpretability of the data, by subtracting the variable average from the data. By doing so the data set is repositioned around the origin. Large variance variables are more likely to be expressed in modeling than low variance variables. Consequently, unit variance scaling was selected to scale the data appropriately. This scaling method calculates the standard deviation of each variable. The inverse standard deviation is used as a scaling weight for each MR-measure.

The results from the OPLS analysis are visualized in a scatter plot (Fig. 1) by plotting the predictive component, which contains the information related to class separation. Components are vectors, which are linear combinations of partial vectors and are dominated by the input variables (x), in this case the regional volumes and regional cortical thickness measures given by Table 1. The first and second components are by definition orthogonal to each other and span the projection plane of the points. Each point in the scatter plot represents one individual subject. The predictive component receives a $Q^2(Y)$ value that describes its statistical significance for separating groups. $Q^2(Y)$ values>0.05 are regarded as statistically significant and models with $Q^2(Y)>0.5$ are regarded as good (Eriksson et al., 2006), where

$$Q^2(Y) = 1 - \mathrm{PRESS} / \mathrm{SSY} \quad (1)$$

where PRESS (predictive residual sum of squares) $= \Sigma(y_{\mathrm{actual}} - y_{\mathrm{predicted}})^2$ and SSY is the total variation of the Y matrix after scaling and mean centring (Eriksson et al., 2006). $Q^2(Y)$ is the fraction of the total variation of the Ys (expected class values) that can be predicted by a component according to cross validation (CV). Cross validation is a statistical method for validating a predictive model which involves building a number of parallel models. These models differ from each other by leaving out a part of the data set each time. The data omitted is then predicted by the respective model. In this study we used seven fold cross-validation, which means that 1/7th of the data is omitted for each cross-validation round. Data is omitted once and only once. To determine the number of components in a model, it is important to have a good balance between fit, $R^2(Y)$ and predictive ability, $Q^2(Y)$ (R^2 = the explained variation and Q^2 = the predicted variation). $R^2(Y)$ and $Q^2(Y)$ vary differently with model complexity. $R^2(Y)$ will rapidly increase towards unity (1) with model complexity, while $Q^2(Y)$ will not. Additional components are generated as long as $Q^2(Y)$ increases. If components were generated when $Q^2(Y)$ started to decrease then only noise would be modeled. $R^2(X)$ explains how much of the variation of the original variables are modeled, while $R^2_p(X)$ explains how much variation is modeled by the predictive component.

Variables were plotted according to their importance for the separation of groups (Fig. 2). The plot shows the MRI measures and their corresponding jack-knifed confidence intervals. Jack-

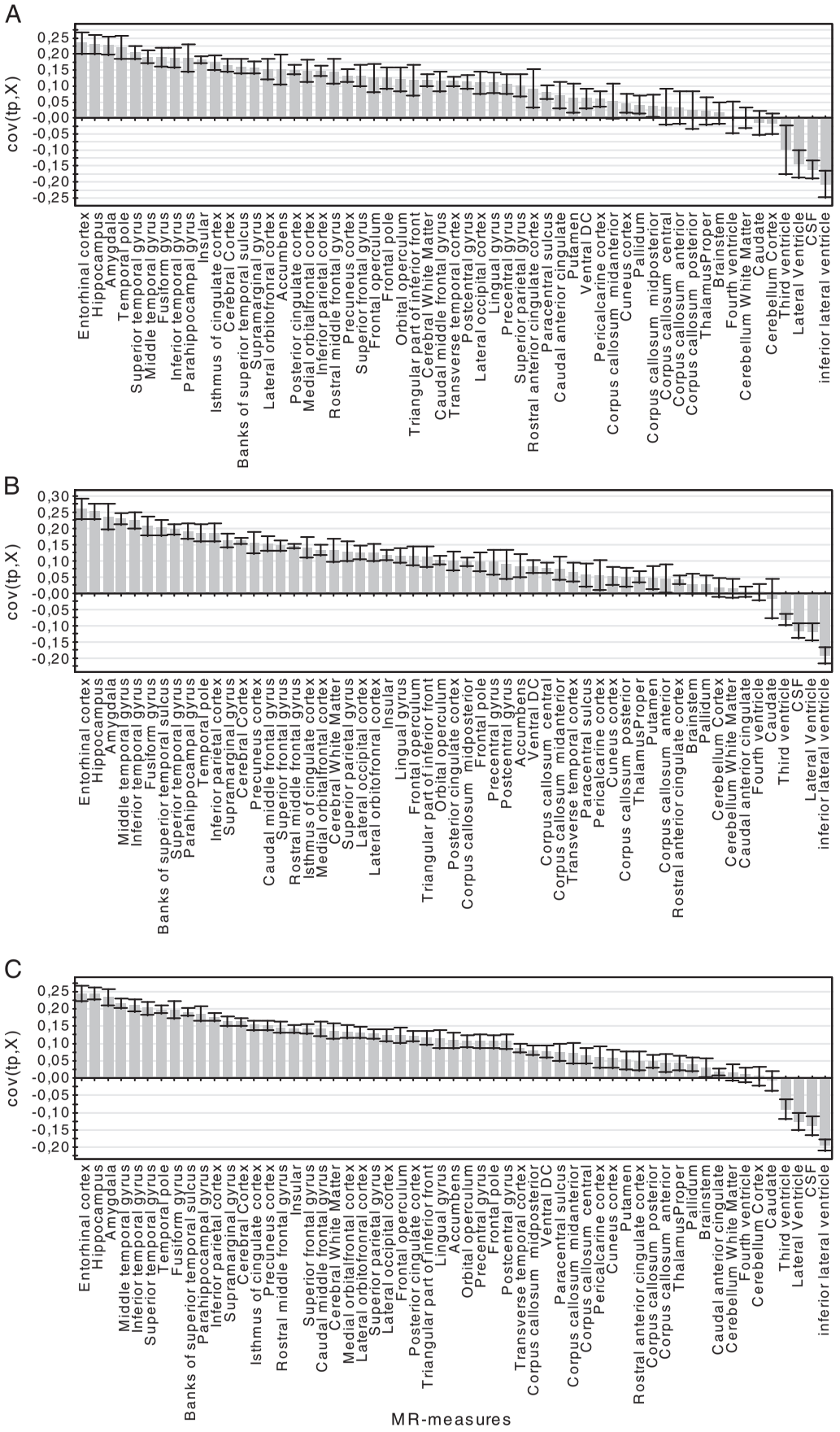

Fig. 2. MRI measures of importance for the separation between AD and CTL (A) AddNeuroMed cohort (B) ADNI cohort (C) Combined cohort. Measures above zero have a larger value in controls compared to AD and measures below zero have a lower value in controls compared to AD. A measure with a high covariance is more likely to have an impact on group separation than a measure with a low covariance. Measures with jack knifed confidence intervals that include zero have low reliability.

knifing is used to estimate the bias and standard error. Measures with confidence intervals that include zero have low reliability (Wiklund et al., 2008). Covariance is plotted on the y-axis, where

$$Cov(t, X_i) = t^T X_i / (N-1) \quad (2)$$

where t is the transpose of the score vector t in the OPLS model, i is the centered variable in the data matrix X and N is the number of variables (Wiklund et al., 2008). A measure with high covariance is more likely to have an impact on group separation than a variable with low covariance. MRI measures below zero in the scatter plot have lower values in controls compared to AD subjects, while MRI measures above zero are higher in controls compared to AD subjects in the model.

Altogether 57 variables were used for OPLS analysis. No feature selection was performed, meaning all measured variables were included in the analysis. Three OPLS models were created. The first model contained the AddNeuroMed cohort, the second model contained the ADNI cohort and the third model combined both the AddNeuroMed and the ADNI cohorts.

Models containing age were also created to test if there were any significant differences between the diagnostic groups in relation to the variable. We investigated whether age would increase the predictive power of the models using it as x-variable. To further validate the models created we used the AddNeuroMed cohort as a training dataset and the ADNI cohort as a test set to see how well the model could predict new and unseen data. We also used the ADNI cohort as a training dataset and the AddNeuroMed cohort as a test set.

Sensitivity and specificity were calculated from the cross-validated prediction values of the OPLS models. Finally, the positive and negative likelihood ratios (LR+ = sensitivity/(100 − specificity) and LR− = (100 − sensitivity)/specificity) were calculated. A positive likelihood ratio between 5 and 10 or a negative likelihood ratio between 0.1 and 0.2 increases the diagnostic value in a moderate way, while a value above 10 or below 0.1 significantly increases the diagnostic value of the test (Qizilbash et al., 2002).

Finally the combined AddNeuroMed/ADNI cohort was used as a training set to investigate how well it could predict conversion from MCI to AD after 1 year follow-up.

Results

Subject cohort

1074 subjects were included in this study: 295 AD patients, 444 MCI patients and 335 control subjects. Using age as an x-variable in the OPLS models did not have any effect on the predictive power of the models separating the groups when all image variables were included. Therefore, age was excluded from further analysis. All MRI volumetric measures were normalized by dividing by each subject's intracranial volume. As expected, performance on the MMSE, CDR and ADAS1 was poorest among AD patients and best among controls (Table 2). The MCI group had scores between the AD and the control groups (Table 2).

OPLS modeling and quality

Three OPLS models were created using (1) the AddNeuroMed cohort (2) the ADNI cohort (3) the combined AddNeuroMed and ADNI cohort for AD *vs.* CTL. The first model (AddNeuroMed cohort) resulted in one predictive and two orthogonal (1 + 2) components with the total explained variance $R^2(X) = 60\%$, the variance related to class separation $R^2_p(X) = 24\%$, the cross validated predictability $Q^2(Y) = 46\%$ and $R^2(Y) = 0.55$. The second model (ADNI cohort) resulted in one predictive and two orthogonal (1 + 2) components with $R^2(X) = 58\%$, $R^2_p(X) = 18\%$, cross validated predictability $Q^2(Y) = 57\%$ and $R^2(Y) = 0.63$. Finally, the combined AddNeuroMed/ADNI cohort yielded one predictive and three orthogonal (1 + 3) components with $R^2(X) = 60\%$, $R^2_p(X) = 20\%$, cross validated predictability $Q^2(Y) = 55\%$ and $R^2(Y) = 0.59$.

Cross validated scatter plots

Fig. 1 illustrates the separation between groups and the predictive power of the models $Q^2(Y)$. Fig. 1A shows the separation between AD and controls in the AddNeuroMed cohort. This model resulted in a sensitivity of 79.0% and a specificity of 90.0%. Fig. 1B show the separation between AD and controls in the ADNI cohort, which resulted in a sensitivity of 86.9% and specificity of 86.7%. Finally, Fig. 1C shows a scatter plot illustrating the separation between AD and controls using the combined AddNeuroMed/ADNI cohort resulting in a sensitivity of 83.4% and specificity of 87.8%. Table 3 shows the sensitivity, the specificity, the accuracy and the likelihood ratios for the different cohorts.

Variables of importance

Fig. 2 illustrates the MRI measures of importance when comparing AD *vs.* controls in the different cohorts. As can be observed the pattern of atrophy is very similar between the AddNeuroMed cohort (Fig. 2A) and the ADNI cohort (Fig. 2B). As expected the most important regions for the separation are medial temporal lobe structures with entorhinal cortex, hippocampus and amygdala being the most influential. Other temporal and limbic areas are also of importance for the separation between AD and controls in the two cohorts. The order of the variables is slightly different between the two cohorts but the general pattern of the structures is very similar. Examples of these structures are the parahippocampal gyrus, temporal pole, isthmus of cingulate cortex, inferior, superior and middle temporal gyrus. The amount of CSF is increased in the two Alzheimer cohorts relative to

Table 2
AddNeuroMed and ADNI subject characteristics.

	AddNeuroMed			ADNI			ADNI/AddNeuroMed		
	CTL	MCI	AD	CTL	MCI	AD	CTL	MCI	AD
Number	110	119	119	225	325	176	335	444	295
Female/Male	60/50	59/60	79/40	111/116	124/201	86/90	171/166	183/261	165/130
Age	72.9 ± 6.5	74.3 ± 5.7	75.6 ± 6.0	76.0 ± 5.0	74.5 ± 7.1	75.3 ± 7.5	75.0 ± 5.7	74.5 ± 6.8	75.4 ± 6.9
Education	10.8 ± 4.8	8.9 ± 4.3	8.0 ± 3.0	16.0 ± 3.0	15.6 ± 3.0	14.6 ± 3.2	14.3 ± 4.3	14.0 ± 4.6	12.0 ± 4.8
MMSE	29.1 ± 1.2	27.1 ± 1.7	20.9 ± 4.7	29.1 ± 1.0	27.0 ± 1.8	23.3 ± 2.0	29.1 ± 1.1	27.1 ± 1.7	22.3 ± 3.6
CDR	0	0.5	1.2 ± 0.5	0	0.5	0.7 ± 0.3	0	0.5	0.9 ± 0.4
ADAS1	3.5 ± 1.5	5.3 ± 1.2	6.6 ± 1.5	2.8 ± 1.1	4.5 ± 1.4	6.1 ± 1.4	3.1 ± 1.3	4.7 ± 1.4	6.3 ± 1.5

Data are represented as mean ± standard deviation. AD = Alzheimer's disease, MCI = Mild Cognitive Impairment, CTL = healthy control, Education in years, MMSE = Mini Mental State Examination, ADAS1 = Word list non-learning (mean), CDR = Clinical Dementia Rating.

Table 3
Sensitivity/specificity and likelihood ratio for the different cohort models.

	Sensitivity	Specificity	Accuracy	LR+	LR−
AddNeuroMed (cv)	79.0 (70.8–85.4)	90.0 (83.0–94.3)	84.3 (79.0–88.4)	7.9 (4.5–13.9)	0.23 (0.16–0.33)
ADNI (cv)	86.9 (81.1–91.1)	86.7 (81.6–90.5)	86.8 (83.0–94.3)	6.5 (4.7–9.1)	0.15 (0.10–0.22)
Combined (cv)	83.4 (78.7–87.2)	87.8 (86.3–96.3)	85.7 (82.8–88.2)	6.8 (5.1–9.1)	0.19 (0.15–0.25)
AddNeuroMed[a]	75.0 (66.3–81.7)	92.7 (84.1–95.0)	83.4 (78.0–87.7)	10.3 (5.2–20.2)	0.27 (0.20–0.37)
ADNI[b]	83.0 (76.7–87.8)	88.4 (83.6–92.0)	86.0 (82.3–89.1)	7.2 (5.0–10.4)	0.19 (0.14–0.27)

[a] AddNeuroMed data set as test set and ADNI data set as training set.
[b] ADNI data set as test set and AddNeuroMed data set as training set, confidence intervals within parenthesis, cv = cross-validation, LR+ = positive likelihood ratio and LR− = negative likelihood ratio.

controls as a result of atrophy. Regions of very little or no importance for the separation between groups are also similar between the two cohorts. Examples include the brainstem, cerebellum, pallidum and corpus callosum. Further, these regions are not expected to be affected in AD which strengthens the results obtained from the two cohorts. Fig. 1C shows the pattern of atrophy when the AddNeuroMed cohort and the ADNI cohort were combined. Once again, a very similar pattern of atrophy is observed.

Model validation

Cross-validation is a good way to show the robustness of a model. All the models discussed so far have used 7-fold cross-validation. However, the best way of validating a model is to use an external test set. Therefore, we used the ADNI model as a training set and predicted the AddNeuroMed cohort on to this model and *vice versa*. The results are similar to those obtained by cross-validation (Table 3). Using the AddNeuroMed cohort as a test set and the ADNI cohort as a training set resulted in a sensitivity of 75.0% and a specificity of 92.7%. Using the ADNI dataset as a test set and the AddNeuroMed dataset as a training set resulted in a sensitivity of 83.0% and a specificity of 88.4%. For further validation, we compared if subjects were classified differently between the different models (for example classified as AD in one model and control in another model). We compared the single cohort cross-validated models with the combined cross-validated cohort model and the single cohort models using the train/test set approach. The results show that the classification agreement for the different comparisons lies between 91 and 96% (Table 4).

MCI conversion prediction

As a final step we investigated if the combined AddNeuroMed/ADNI cohort could predict conversion from MCI to AD at baseline. All MCI subjects from the two cohorts who had 1 year follow-up data were classified as either belonging to the AD or the control group using the combined cohort model. The results are given by Table 5. This resulted in 71% of the MCI converters (MCI-c) from both the cohorts classified as AD-like and 60% of the MCI stable (MCI-s) classified as control-like at baseline. For the AddNeuroMed cohort, 64% of the MCI-c subjects who converted to AD at 1 year follow-up were classified as AD-like and 71% of the MCI-s predicted as control-like at baseline. For the ADNI cohort 74% of the MCI-c subjects who converted to AD at 1 year follow-up were classified as AD-like and 56% of the MCI-s predicted as control-like.

Discussion

This study was designed to investigate if multivariate classification techniques trained on one of two large cohorts (AddNeuroMed and ADNI) with identical MR acquisition protocols would be applicable to the other and give similar results. We wished to assess whether we would find similar patterns of atrophy in the two cohorts, when comparing AD and healthy controls. Further, we wanted to investigate if comparable sensitivities and specificities between the cohorts could be obtained. Finally, we wanted to assess the prediction of conversion from MCI to AD using the combined AddNeuroMed/ADNI cohort.

The multivariate method we have used in this study is orthogonal partial least square to latent structures (OPLS). This method has previously been used to distinguish AD patients from healthy controls and to predict conversion from MCI to AD using MRI data as input (Westman et al., 2011). OPLS has also been applied to MRI and MRS data, with the combination of the two measures showing improved discrimination between Alzheimer's disease and controls (Westman et al., 2010). Bylesjö et al. have shown that OPLS can be used to combine different types of omics data. They showed that the systematic variation from two analytical platforms could be combined and separated from the systematic variation specific to each analytical platform (Bylesjo et al., 2007). This illustrates the unique property of OPLS compared to other linear regression methods which is that it divides the systematic variation within the dataset into two parts, one correlated with Y and one uncorrelated with Y (structured noise), making data interpretation easier (Wiklund et al., 2008). At the same

Table 4
Comparison of subject classification between cohort models.

	ANM and ANM/ADNI	ADNI and ANM/ADNI	ANM and ANMonADNI	ADNI and ADNIonANM
Total n	229	401	229	401
Same classification	217	385	209	372
% Same classification	95%	96%	91%	93%
Different classification	12	16	20	29
% Different classification	5%	4%	9%	7%

ANM = AddNeuroMed cohort model, ANM/ADNI = the combined AddNeuroMed and ADNI cohort model, ANMonADNI = AddNeuroMed cohort test set and ADNI cohort training set, ADNIonANM = ADNI cohort test set and AddNeuroMed cohort training set, Total n = AD and CTL subjects, Same classification = number of subjects predicted in the same way, % same classification = percentage of subjects predicted in the same way, Different classification = number of subjects predicted differently, % different classification = percentage of subjects predicted differently.

Table 5
MCI prediction subject characteristics.

	Number	AD-like	CTL-like
Combined cohort MCI converters	84	60 (71%)	24 (29%)
Combined cohort MCI stable	353	141(40%)	212 (60%)
AddNeuroMed MCI converters	22	14 (64%)	8 (36%)
AddNeuroMed MCI stable	97	28 (29%)	69 (71%)
ADNI MCI converters	62	46 (74%)	16 (26%)
ADNI MCI stable	256	113 (44%)	143 (56%)
ADNI MCI to CTL	7	1 (14%)	6 (86%)

AD = Alzheimer's disease, MCI = Mild Cognitive Impairment, CTL = healthy control. MCI subjects were predicted on to the AddNeuroMed/ADNI model containing AD and CTL subjects.

time, OPLS maximizes the covariance between the dependent and the independent variables. This can be of particular importance when data from two studies are compared. Other advantages of OPLS are that it does not require homogeneous classes and the number of variables do not need to be smaller than the number of samples, which is also the case for SVM. All types of matrices can be used as input for multivariate analysis using OPLS, long and lean, short and fat, almost square (Eriksson et al., 2006). Further, the OPLS loadings maps illustrated by Fig. 2 show which input variables are important for the classification, and allow the biological plausibility of the classification to be investigated. Although SVM weights can show which variables are important for separating groups for linear SVMs, this is not the case for non-linear SVMs or LDA.

Several other studies have utilized other multivariate techniques such as principal component analysis (PCA), partial least squares to latent structures (PLS) and support vector machines (SVM) to analyze MR-data (Boguszewicz et al., 2010; Fan et al., 2005, 2008; Kloppel et al., 2008a, 2008b; Levine et al., 2008; McIntosh and Lobaugh, 2004; Plant et al., 2009; Vemuri et al., 2008; Westman et al., 2007, 2009). As input for OPLS analysis we only use automated regional MRI measures (regional cortical thickness and regional volumes). These may have particular advantages when it comes to widespread uptake in either clinical or research practice. Manual measures of different brain regions are time consuming and operator dependent and therefore not always practical in a clinical settings. However automated tools must be precise, accurate, fast and must be validated and tested on large cohorts.

Model predictability

In this study three different models were created to investigate the ability of OPLS to distinguish between AD and healthy elderly controls. The models were created from the AddNeuroMed cohort, the ADNI cohort and the combined AddNeuroMed/ADNI cohort. Further, the model created from the AddNeuroMed cohort was validated using the ADNI cohort as a test set and likewise the model created from the ADNI cohort was validated by using the AddNeuroMed cohort as a test set. The sensitivities and specificities as well as the likelihood ratios of the different models can be observed in Table 3. Both the negative and positive likelihood ratios are similar for each of the models, although the sensitivity and specificity for the different models vary a little. We believe that the models created from the different cohorts are robust since we have used both full cross validation and external test sets for model validation. An optimistic bias in classification accuracy can be obtained if cross-validation is not used (Schulerud and Albregtsen, 2004; Simon Spycher, 2004). Naturally, the use of an external test set is preferable, but this is not always possible if only small data sets are available.

Previous studies have attempted to distinguish between subjects with AD and healthy controls. Some have used hippocampal or entorhinal cortex measures for classification with a high degree of accuracy (80%–90%) (Fox et al., 1996; Jack et al., 1992, 1997; Juottonen et al., 1999; Killiany et al., 1993; Laakso et al., 1996, 1998; Lehericy et al., 1994; Seab et al., 1988). Accuracies of up to 100% have been achieved when discriminating between AD and controls, but these studies included more severely impaired AD subject (Callen et al., 2001; Juottonen et al., 1999; Lerch et al., 2008), had very small sample sizes (Lerch et al., 2008), or did not apply full cross-validation (Callen et al., 2001; Fan et al., 2008; Juottonen et al., 1999; Killiany et al., 2002).

We have previously obtained a sensitivity of 93% and specificity of 86% using the same method with the same input variables on a separate cohort (Westman et al., 2010). However, the sample size in this latter study (30 AD and 36 controls) was much smaller, which is probably the reason for the slightly higher sensitivity and specificity. The AddNeuroMed cohort has previously been analyzed with the OPLS method, with different input variables (automatically measured regional volumes in combination with manually measured hippocampus volumes) (Westman et al., 2011). Similar specificity was achieved when full cross-validation was applied but the sensitivity was higher in the previous study than the present study (90% compared to 79%). A possible explanation for this is the effect of the manual measures of hippocampus included in the previous study which may demonstrate the need for future improvements to automated segmentation techniques. The ADNI cohort has also previously been studied using a similar multivariate technique (Linear discriminant analysis (LDA)) (McEvoy et al., 2009). Included in the study of McEvoy et al. were quantitative structural neuroimaging measures of regional MRI volumes and regional cortical thicknesses to distinguish between Alzheimer's disease and healthy controls (McEvoy et al., 2009). McEvoy and collaborators found a sensitivity of 83% and a specificity of 93% with LDA, compared to the present study where we found a sensitivity of 86.9% and specificity of 86.7% (OPLS analysis with very similar input variables for the ADNI cohort). Methods such as LDA require both strictly homogeneous classes and that the number of variables is much smaller than the number of samples (Eriksson et al., 2006). These criteria do not need to be fulfilled when using PLS and OPLS. The current study included a larger number of subjects than the study of McEvoy et al. which may be the reason for the difference in sensitivity and specificity. We took the view that we should include all subjects from both cohorts who passed our image quality control measures (Simmons et al., 2009, 2011). Cuingnet et al. have also investigated the ADNI cohort using ten different methods for automated classification of patients with AD from structural MRI. They received sensitivities up to 81% and specificities up to 95% with their best method. We believe that our method is robust and accurate in distinguishing between and AD and controls and is comparable to the results of other groups.

Combining the two separate cohorts we found a sensitivity of 83.4% and a specificity of 87.8%, further demonstrating the robustness of the analysis approach used in this study. Very similar patterns of atrophy were obtained for the two cohorts and when they were combined a similar proportion of subjects were correctly predicted as for the two separate cohorts. The agreement in classification between the ADNI cohort and the combined cohort (subjects predicted in the same way) was 96%. A similar agreement was also found for the comparison of the AddNeuroMed and the combined cohort (95%). These results again demonstrate the robustness of the model.

The major difference in demographics between the two cohorts is the number of years of education. The education level of the ADNI cohort is around six years higher for all three diagnostic groups (AD, MCI and healthy controls) than the AddNeuroMed cohort. This difference seems to be due to a particularly high level of education in the ADNI cohort and perhaps also because years of education were lower in the post war years during the reconstruction of Europe when the AddNeuroMed cohort was growing up. It is possible that the ADNI cohort might hence be rather selective in this regard and therefore not as representative of the broader population as it could be. Although we can only speculate as to the reason for the small variation between the two cohorts, the differing education levels of the cohorts may be one reason and additionally the ADNI cohort may be more homogeneous than the AddNeuroMed cohort. Within the AddNeuroMed cohort, data was collected from six different countries. These reasons and others may explain why the sensitivity and the specificity are lower for the AddNeuroMed cohort compared to the ADNI cohort. There are other factors which are different between the two cohorts, such as the age of the controls. The average age of the controls are lower for the AddNeuroMed cohort than the ADNI cohort, but when the controls are averaged together in the combined cohort, it matches the age of the other diagnostic groups better than they do in the respective cohort separately.

MRI measures of importance

Neurofibrillary tangles (NFTs) are one of the pathological hallmarks of AD and are associated with neuronal loss and brain volume reduction. According to Braak and Braak neurofibrillary tangles spread in a defined and specific pattern composed of 6 different stages (Braak and Braak, 1991). During the first two Braak stages pathology is confined to the entorhinal cortex/transentorhinal cortex with minimal involvement of the hippocampus. In the third and fourth stages the disease spreads to the hippocampus and the medial temporal limbic areas and in the final two stages the pathology extends to the isocortical association areas. The ADNI and the AddNeuroMed cohorts show very similar pattern of atrophy as illustrated by Fig. 2. The atrophy pattern in the two cohorts can be compared to the staging of NFTs described by Braak and Braak. The most important measure for the separation between the two subject groups (AD and CTL) is the entorhinal cortex. The second and third most important regions are the hippocampus and amygdala. Other medial temporal limbic areas are also of importance, such as the parahippocampal gyrus and posterior cingulate. As can be observed from the loadings plot (Fig. 2) the atrophy also spreads to the lateral temporal, parietal and frontal lobes according to the Braak scheme. Karas et al. have shown a similar pattern of atrophy in a VBM study where atrophy started in the medial temporal lobe and then spread to lateral temporal, parietal and frontal lobe (Karas et al., 2003). The staging of AD pathology proposed by Braak et al. is also supported by many other VBM studies (Zakzanis et al., 2003). There are several previous studies which have used FreeSurfer to analyze data from the ADNI cohort (Lehmann et al., 2010; McEvoy et al., 2009; Murphy et al., 2010). Using feature selection McEvoy et al. found that the most important regions for discriminating between AD and controls were: hippocampus, entorhinal cortex, middle temporal gyrus and banks of superior temporal sulcus among others. Each of the regions that they found to be important was also found to be key factors in our study. However, we did not use any feature selection since we did not find it to improve the performance of our models, though feature selection does increase computational time (Cuingnet et al., 2011).

MCI prediction from the combined cohort

The final aim of this study was to predict conversion from MCI to AD using the baseline scans and a 1 year clinical follow-up. The combined AddNeuroMed/ADNI cohort (controls and AD subjects) was used as a training set and all the MCI subjects from the two cohorts which had 1 year clinical follow-up as the test set. MCI subjects were either classified as AD like or as control like with the results summarized in Table 5. 71% of the MCI converters (MCI-c) from both the cohorts were classified as AD-like at baseline and 60% of the MCI stable (MCI-s) subjects classified as control-like.

Considering the two cohorts separately, MCI-c subjects from the ADNI cohort were more accurately predicted to be AD-like, while MCI-s subjects were more accurately predicted to be control-like in the AddNeuroMed cohort. The MCI subjects are a heterogeneous group of subjects; some subjects remain stable over long periods of time, some convert to AD or another neurological disorder and a smaller number revert to a cognitively normal status. Due to the heterogeneity of this group, clinical inclusion and exclusion criteria are of great importance. The clinical criteria are slightly different between the ADNI and the AddNeuroMed cohorts which is the most likely reason for the differences observed in the accuracy of prediction of conversion from MCI to AD. It is also likely that the AddNeuroMed MCI population is more heterogeneous in general than the ADNI cohort and the criteria for recruitment within ADNI is a little more specific for recruiting subjects with amnestic MCI. Further a 1 year follow-up period is a relatively short length of time, which may explain why many of the MCI-s subjects are classified as AD like. It is very likely that many of the subjects in this group will convert to Alzheimer's disease at a later time, but morphological changes in the brain are already present now. However, this does not explain why the likelihood is higher for MCI-s in the ADNI cohort to be more AD-like than for the AddNeuroMed cohort. This may be because the AddNeuroMed cohort is more heterogeneous, and does not contain as many amnestic MCI subjects as the ADNI cohort, resulting in fewer subjects converting to AD. Another explanation could be the higher educational level of the ADNI subjects. In this group there are highly educated people who will develop dementia. It has been shown that atrophic changes may be advanced in more educated people while cognition is still preserved (brain reserve/cognitive reserve) (Foubert-Samier et al., 2010; Ngandu et al., 2007).

In a previous study Ewers et al. (2010) combined different markers (MRI, CSF and neuropsychological tests) to predict conversion to AD from MCI up to 3.3 years after baseline testing (AD n = 81, control n = 101, MCI-c = 58 and MCI-s = 72). Ewers et al. also trained their classification model on AD and control subjects resulting in prediction accuracies between 60 and 70% for MCI-c and MCI-s. We found a similar accuracy (65.5%) using imaging measures alone for a much larger cohort. In another previous study, Davatzikos et al. used a MRI classification technique which they termed SPARE-AD (Spatial patterns of Abnormalities for recognition of early AD) created from an AD (n = 54) *vs.* control (n = 63) model (all subjects from the ADNI cohort). These patterns were used to predict conversion in MCI subjects (MCI-c = 69 and MCI-s = 170) between 6 and 36 months after baseline (Davatzikos et al., 2010). Davatzikos et al. found a prediction accuracy of 55.8% (sensitivity = 94.7%, specificity = 37.8%) compared to our combination of subjects from both AddNeuroMed and ADNI where we found a prediction accuracy 65.5% (sensitivity = 71%, specificity = 60%). The prediction accuracy is higher for our study than Davatzikos et al., while we also found less difference between sensitivity and specificity than they did. The two studies used different analytical techniques which could be the reason for the difference in sensitivity and specificity. It is also possible that these differences are due to the size of the training set (117 subjects compared to 630 subjects). Smaller datasets have previously been found to give high prediction accuracies for MCI-c with lower specificities. Our model is more balanced in terms of sensitivity and specificity results and our larger dataset is more likely to accurately reflect the general population.

We have previously analyzed the AddNeuroMed MCI-c group with the same method (OPLS) but with different input variables (automatically measured regional volumes in combination with manually measured hippocampus volumes) (Westman et al., 2011). The prediction result from this earlier study were good with 74% of the MCI-c classified as more AD-like compared to the 64% found in the present study. A possible explanation for this could again be the nature of the training set. The training set in that study included only the AddNeuroMed cohort which is likely to more closely match the patterns of atrophy observed in the AddNeuroMed MCI-c group than that of the combined AddNeuroMed and ADNI cohort in this study. It has been proposed that the pattern of atrophy can progress in other ways than that described by Braak and Braak, where the hippocampus and entorhinal cortex are not as affected (Shiino et al., 2006). Different types of atrophy patterns were observed but all the subjects had the same clinical diagnosis. Larger cohorts are more likely to contain subjects with different types of atrophy which could explain the decrease in prediction accuracy received in the present study. Moreover if a new completely unseen data set from neither the AddNeuroMed nor the ADNI cohort was assessed, then it is likely that the combined cohort would give better prediction accuracy than a smaller cohort, since the combined cohort is more likely to resemble the overall population.

Conclusion

To conclude, we have used OPLS, a powerful multivariate technique to compare and combine structural MRI data collected from two large multi-center studies (AddNeuroMed and ADNI).

Multivariate data analysis in combination with quantitative structural MRI can accurately classify elderly people from patients with Alzheimer's disease. The atrophy pattern of the two cohorts is very similar and corresponds well to the spread of NFTs described by Braak and co-workers. Further, the prediction accuracy of the different models created was also very similar, independent of the different cohorts and method used for validation. This made it possible to combine the two cohorts to create a large and robust model which could be used to predict MCI conversion into AD. These methods also show promise in predicting future progression to AD from the transitional MCI state. There is scope to further investigate the MCI group, for example by dividing MCI subjects into subgroups, due to the heterogeneity within the group. Furthermore a 1 year-follow up time is relatively short time span and it is most likely that the group labeled MCI stable contains many subjects who will convert to AD at a later stage. We will explore this as part of a future study. The current study demonstrates that the methods used are robust and that large data sets can be combined if MRI imaging protocols are carefully aligned. Combining large data sets is warranted since they are likely to more closely match the general population and give a more complete picture of AD in the general population. This will hopefully lead to a better understanding of the etiology of the disease and aid diagnosis at an earlier stage of the disease in the future.

Acknowledgments

This study was supported by InnoMed, (Innovative Medicines in Europe) an Integrated Project funded by the European Union of the Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5, Life Sciences, Genomics and Biotechnology for Health.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904).

ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

Also, thanks to the Foundation Gamla Tjänarinnor, the Swedish Alzheimer's Association and Swedish Brain Power, Health Research Council of Academy of Finland and Stockholm Medical Image Laboratory and Education (SMILE). AS and SL were supported by funds from NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Kings College London.

References

Boguszewicz, L., Blamek, S., Sokol, M., 2010. Pattern recognition methods in (1)H MRS monitoring in vivo of normal appearing cerebellar tissue after treatment of posterior fossa tumors. Acta Neurochir. Suppl. 106, 171–175.

Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. Acta Neuropathol. 82, 239–259.

Bylesjo, M., Eriksson, D., Kusano, M., Moritz, T., Trygg, J., 2007. Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. Plant J. 52, 1181–1191.

Callen, D.J., Black, S.E., Gao, F., Caldwell, C.B., Szalai, J.P., 2001. Beyond the hippocampus: MRI volumetry confirms widespread limbic atrophy in AD. Neurology 57, 1669–1674.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., Chupin, M., Benali, H., Colliot, O., 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. Neuroimage 56 (2), 766–781.

Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5, 162–176.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179–194.

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q., 2010. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol. Aging 2010 Jun 29 (Electronic publication ahead of print).

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wiksröm, C., Wold, S., 2006. Multi- and Megavariate Data Analysis (Part I -Basics and Principals and Applications), 2nd ed. Umetrics AB, Umeå.

Ewers, M., Walsh, C., Trojanowski, J.Q., Shaw, L.M., Petersen, R.C., Jack Jr., C.R., Feldman, H.H., Bokde, A.L.W., Alexander, G.E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H., 2010. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol. Aging 2010 Dec 13 (Electronic publication ahead of print).

Fan, Y., Shen, D., Davatzikos, C., 2005. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Med. Image Comput. Comput. Assist. Interv. 1–8.

Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., 2008. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 39, 1731–1743.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. U.S.A. 97, 11050–11055.

Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.

Fischl, B., Liu, A., Dale, A.M., 2001. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans. Med. Imaging 20, 70–80.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Fischl, B., Salat, D.H., van der Kouwe, A.J., Makris, N., Segonne, F., Quinn, B.T., Dale, A.M., 2004a. Sequence-independent segmentation of magnetic resonance images. Neuroimage 23 (Suppl. 1), S69–S84.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004b. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.

Foubert-Samier, A., Catheline, G., Amieva, H., Dilharreguy, B., Helmer, C., Allard, M., Dartigues, J.-F., 2010. Education, occupation, leisure activities, and brain reserve: a population-based study. Neurobiol. Aging 2010 Nov 11 (Electronic publication ahead of print).

Fox, N.C., Warrington, E.K., Freeborough, P.A., Hartikainen, P., Kennedy, A.M., Stevens, J.M., Rossor, M.N., 1996. Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. Brain 119 (Pt 6), 2001–2007.

Jack Jr., C.R., Petersen, R.C., O'Brien, P.C., Tangalos, E.G., 1992. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42, 183–188.

Jack Jr., C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., Kokmen, E., 1997. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology 49, 786–794.

Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L., Whitwell J., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Johan Trygg, S.W., 2002. Orthogonal projections to latent structures (O-PLS). J. Chemometrics 16, 119–128.

Juottonen, K., Laakso, M.P., Partanen, K., Soininen, H., 1999. Comparative MR analysis of the entorhinal cortex and hippocampus in diagnosing Alzheimer disease. AJNR Am. J. Neuroradiol. 20, 139–144.

Kantarci, K., 2005. Magnetic resonance markers for early diagnosis and progression of Alzheimer's disease. Expert Rev. Neurother. 5, 663–670.
Karas, G.B., Burton, E.J., Rombouts, S.A.R.B., van Schijndel, R.A., O'Brien, J.T., Scheltens, P.h., McKeith, I.G., Williams, D., Ballard, C., Barkhof, F., 2003. A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. Neuroimage 18, 895–907.
Killiany, R.J., Moss, M.B., Albert, M.S., Sandor, T., Tieman, J., Jolesz, F., 1993. Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Arch. Neurol. 50, 949–954.
Killiany, R.J., Hyman, B.T., Gomez-Isla, T., Moss, M.B., Kikinis, R., Jolesz, F., Tanzi, R., Jones, K., Albert, M.S., 2002. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. Neurology 58, 1188–1196.
Kloppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S., 2008a. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. Brain 131, 2969–2974.
Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S.J., 2008b. Automatic classification of MR scans in Alzheimer's disease. Brain 131, 681–689.
Laakso, M.P., Partanen, K., Riekkinen, P., Lehtovirta, M., Helkala, E.L., Hallikainen, M., Hanninen, T., Vainio, P., Soininen, H., 1996. Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia: an MRI study. Neurology 46, 678–681.
Laakso, M.P., Soininen, H., Partanen, K., Lehtovirta, M., Hallikainen, M., Hanninen, T., Helkala, E.L., Vainio, P., Riekkinen Sr., P.J., 1998. MRI of the hippocampus in Alzheimer's disease: sensitivity, specificity, and analysis of the incorrectly classified subjects. Neurobiol. Aging 19, 23–31.
Lehericy, S., Baulac, M., Chiras, J., Pierot, L., Martin, N., Pillon, B., Deweer, B., Dubois, B., Marsault, C., 1994. Amygdalohippocampal MR volume measurements in the early stages of Alzheimer disease. AJNR Am. J. Neuroradiol. 15, 929–937.
Lehmann, M., Douiri, A., Kim, L.G., Modat, M., Chan, D., Ourselin, S., Barnes, J., Fox, N.C., 2010. Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. Neuroimage 49, 2264–2274.
Lerch, J.P., Pruessner, J., Zijdenbos, A.P., Collins, D.L., Teipel, S.J., Hampel, H., Evans, A.C., 2008. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol. Aging 29, 23–30.
Levine, B., Kovacevic, N., Nica, E.I., Cheung, G., Gao, F., Schwartz, M.L., Black, S.E., 2008. The Toronto traumatic brain injury study: injury severity and quantified MRI. Neurology 70, 771–778.
Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L.O., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2011. Combination analysis of neuropsychological tests and structural MRI measures in differentiating AD, MCI and control groups — the AddNeuroMed study. Neurobiol. Aging Jul;32(7):1198-1206. Epub 2009 Aug 14.
Liu, Y., Paajanen, T., Westman, E., Wahlund, L.O., Simmons, A., Tunnard, C., Sobow, T., Proitsi, P., Powell, J., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010a. Effect of APOE epsilon4 allele on cortical thicknesses and volumes: the AddNeuroMed study. J. Alzheimers Dis. 21, 947–966.
Liu, Y., Paajanen, T., Westman, E., Zhang, Y., Wahlund, L.O., Simmons, A., Tunnard, C., Sobow, T., Proitsi, P., Powell, J., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010b. APOE epsilon2 allele is associated with larger regional cortical thicknesses and volumes. Dement. Geriatr. Cogn. Disord. 30, 229–237.
Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L.O., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010c. Analysis of regional MRI volumes and thicknesses as predictors of conversion from mild cognitive impairment to Alzheimer's disease. Neurobiol. Aging Aug;31(8):1375-85. Epub 2010 May 5.
Lovestone, S., Francis, P., Strandgaard, K., 2007. Biomarkers for disease modification trials — the innovative medicines initiative and AddNeuroMed. J. Nutr. Health Aging 11, 359–361.
Lovestone, S., Francis, P., Kloszewska, I., Mecocci, P., Simmons, A., Soininen, H., Spenger, C., Tsolaki, M., Vellas, B., Wahlund, L.-O., Ward, M., 2009. AddNeuroMed; the European Collaboration for the Discovery of Novel Biomarkers for Alzheimer's Disease. Ann. N. Y. Acad. Sci. 1180, 36–46.
McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler Jr., J.D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M., 2009. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 2009 Apr;251(1):195-205. Epub 2009 Feb 6.
McIntosh, A.R., Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23, S250–S263.
Murphy, E.A., Holland, D., Donohue, M., McEvoy, L.K., Hagler Jr, D.J., Dale, A.M., Brewer, J.B., 2010. Six-month atrophy in MTL structures is associated with subsequent memory decline in elderly controls. Neuroimage 53, 1310–1317.
Ngandu, T., von Strauss, E., Helkala, E.L., Winblad, B., Nissinen, A., Tuomilehto, J., Soininen, H., Kivipelto, M., 2007. Education and dementia: what lies behind the association? Neurology 69, 1442–1450.
Oberg, J., Spenger, C., Wang, F.H., Andersson, A., Westman, E., Skoglund, P., Sunnemark, D., Norinder, U., Klason, T., Wahlund, L.O., Lindberg, M., 2008. Age related changes in brain metabolites observed by (1)H MRS in APP/PS1 mice. Neurobiol. Aging Sep;29(9):1423-33. Epub 2007 Apr 16.
Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2009. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 50, 162–174.
Qizilbash, S., Chui, Tariot, Brodaty, Kaye, Erkinjuntti, 2002. Evindenced-based Dementia Pratice. Blackwell Publishing, Oxford, pp. 20–23.
Rantalainen, M., Cloarec, O., Beckonert, O., Wilson, I.D., Jackson, D., Tonge, R., Rowlinson, R., Rayner, S., Nickson, J., Wilkinson, R.W., Mills, J.D., Trygg, J., Nicholson, J.K., Holmes, E., 2006. Statistically integrated metabonomic-proteomic studies on a human prostate cancer xenograft model in mice. J. Proteome Res. 5, 2642–2655.
Rosen, W.G., Mohs, R.C., Davis, K.L., 1984. A new rating scale for Alzheimer's disease. Am. J. Psychiatry 141, 1356–1364.
Schulerud, H., Albregtsen, F., 2004. Many are called, but few are chosen. Feature selection and error estimation in high dimensional spaces. Comput. Methods Programs Biomed. 73, 91–99.
Seab, J.P., Jagust, W.J., Wong, S.T., Roos, M.S., Reed, B.R., Budinger, T.F., 1988. Quantitative NMR measurements of hippocampal atrophy in Alzheimer's disease. Magn. Reson. Med. 8, 200–208.
Segonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060–1075.
Segonne, F., Pacheco, J., Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans. Med. Imaging 26, 518–529.
Shiino, A., Watanabe, T., Maeda, K., Kotani, E., Akiguchi, I., Matsuda, M., 2006. Four subgroups of Alzheimer's disease based on patterns of atrophy using VBM and a unique pattern for early onset disease. Neuroimage 33, 17–26.
Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Wahlund, L.-O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., 2009. MRI measures of Alzheimer's disease and the AddNeuroMed study. Ann. N. Y. Acad. Sci. 1180, 47–55.
Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Wahlund, L.-O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., for the AddNeuroMed consortium, 2011. The AddNeuroMed framework for multi-centre MRI assessment of longitudinal changes in Alzheimer's disease: experience from the first 24 months. Int. J. Geriatr. Psychiatry 26 (1), 75–82.
Simon Spycher, M.N.J.G., 2004. Comparison of different classification methods applied to a mode of toxic action data set. QSAR Comb. Sci. 23, 779–791.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.
Thambisetty, M., Simmons, A., Velayudhan, L., Hye, A., Campbell, J., Zhang, Y., Wahlund, L.O., Westman, E., Kinsey, A., Guntert, A., Proitsi, P., Powell, J., Causevic, M., Killick, R., Lunnon, K., Lynham, S., Broadstock, M., Choudhry, F., Howlett, D.R., Williams, R.J., Sharp, S.I., Mitchelmore, C., Tunnard, C., Leung, R., Foy, C., O'Brien, D., Breen, G., Furney, S.J., Ward, M., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Hodges, A., Murphy, D.G., Parkins, S., Richardson, J.C., Resnick, S.M., Ferrucci, L., Wong, D.F., Zhou, Y., Muehlboeck, S., Evans, A., Francis, P.T., Spenger, C., Lovestone, S., 2010. Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. Arch. Gen. Psychiatry 67, 739–748.
Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack Jr., C.R., 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 39, 1186–1197.
Westman, E., Spenger, C., Wahlund, L.-O., Lavebratt, C., 2007. Carbamazepine treatment recovered low N-acetylaspartate + N-acetylaspartylglutamate (tNAA) levels in the megencephaly mouse BALB/cByJ-Kv1.1mceph/mceph. Neurobiol. Dis. 26, 221–228.
Westman, E., Spenger, C., Oberg, J., Reyer, H., Pahnke, J., Wahlund, L.O., 2009. In vivo 1H-magnetic resonance spectroscopy can detect metabolic changes in APP/PS1 mice after donepezil treatment. BMC Neurosci. 10, 33.
Westman, E., Wahlund, L.-O., Foy, C., Poppe, M., Cooper, A., Murphy, D., Spenger, C., Lovestone, S., Simmons, A., 2010. 1H-MRS a valuable complement to MRI in the early diagnosis of Alzheimer's Disease. J. Alzheimers Dis. 22 (1), 171–181.
Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L., consortium, f.t.A., 2011. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. Neuroimage 54 (2), 1178–1187.
Wiklund, S., Johansson, E., Sjostrom, L., Mellerowicz, E.J., Edlund, U., Shockcor, J.P., Gottfries, J., Moritz, T., Trygg, J., 2008. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Anal. Chem. 80, 115–122.
Zakzanis, K.K., Graham, S.J., Campbell, Z., 2003. A meta-analysis of structural and functional brain imaging in dementia of the Alzheimer's type: a neuroimaging profile. Neuropsychol. Rev. 13, 1–18.