



## Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion<sup>☆</sup>

Eric Westman<sup>a,\*</sup>, J-Sebastian Muehlboeck<sup>a,1</sup>, Andrew Simmons<sup>a,b,c,1</sup>

<sup>a</sup> King's College London, Institute of Psychiatry, London, UK

<sup>b</sup> NIHR Biomedical Research Centre for Mental Health, London, UK

<sup>c</sup> Centre for Neurodegeneration Research, King's College London, London, UK

### ARTICLE INFO

#### Article history:

Accepted 28 April 2012

Available online 3 May 2012

#### Keywords:

CSF  
MRI  
OPLS  
AD  
MCI conversion

### ABSTRACT

The suggested revision of the NINCDS–ADRDA criterion for the diagnosis of Alzheimer's disease (AD) includes at least one abnormal biomarker among magnetic resonance imaging (MRI), positron emission tomography (PET) and cerebrospinal fluid (CSF). We aimed to investigate if the combination of baseline MRI and CSF could enhance the classification of AD compared to using either alone and predict mild cognitive impairment (MCI) conversion at multiple future time points. 369 subjects from the Alzheimer's disease Neuroimaging Initiative (ADNI) were included in the study (AD = 96, MCI = 162 and CTL = 111). Freesurfer was used to generate regional subcortical volumes and cortical thickness measures. A total of 60 variables were used for orthogonal partial least squares to latent structures (OPLS) multivariate analysis (57 MRI measures and 3 CSF measures: A $\beta$ <sub>42</sub>, t-tau and p-tau). Combining MRI and CSF gave the best results for distinguishing AD vs. CTL. We found an accuracy of 91.8% for the combined model at baseline compared to 81.6% for CSF measures and 87.0% for MRI measures alone. The combined model also gave the best accuracy when distinguishing between MCI vs. CTL (77.6%) at baseline. MCI subjects who converted to AD by 12 and 18 month follow-up were accurately predicted at baseline using an AD vs. CTL model (82.9% and 86.4% respectively), with lower prediction accuracies for those MCI subjects converting by 24 and 36 month follow up (75.4% and 68.0% respectively). The overall prediction accuracies for converters and non-converters ranged from 58.6% to 66.4% at different time points. Combining MRI and CSF measures in a multivariate model at baseline gave better accuracy for discriminating between AD and CTL, between MCI and CTL and for predicting future conversion from MCI to AD, than using either MRI or CSF separately.

© 2012 Elsevier Inc. All rights reserved.

### Introduction

Alzheimer's disease (AD) is one of the most common forms of neurodegenerative disorders characterized by a gradual loss of cognitive functions such as episodic memory. The disease is related to pathological amyloid depositions and hyperphosphorylation of structural proteins which leads to progressive loss of function, metabolic alterations and structural changes in the brain.

A revision of the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's disease and related Disorders Association (ADRDA) criterion (McKhann et al., 1984) for the diagnosis of AD has been suggested. The new research criterion is still centered on a clinical core of early and significant episodic memory impairment, but also includes at least one abnormal biomarker among magnetic resonance imaging (MRI), positron emission tomography (PET) and cerebrospinal fluid (CSF) (Dubois et al., 2007). The new suggested diagnostic criterion also utilizes biomarkers (McKhann et al., 2011). Since evidence for use of these biomarkers is growing, it strengthens their role in the diagnosis of AD. These biomarkers reflect different yet connected aspects of the disease, with MRI measuring early structural changes in the medial temporal lobe, particularly entorhinal cortex and hippocampus, fluorodeoxyglucose (FDG)-PET measuring glucose metabolism, amyloid PET measuring the build-up of amyloid in tissue and CSF biomarkers reflecting changes in levels of A $\beta$ , tau proteins and ratios of the two. The utilization of the three biomarkers varies from center to center and depends on factors including local availability, cost and historical patterns of usage.

<sup>☆</sup> Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database ([www.loni.ucla.edu/ADNI](http://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](http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Authorship_List.pdf)).

\* Corresponding author at: Department of Neuroimaging, Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, UK. Fax: +46 8 517 761 11.

E-mail address: [eric.westman@ki.se](mailto:eric.westman@ki.se) (E. Westman).

<sup>1</sup> For the Alzheimer's Disease Neuroimaging Initiative.

An effective combination of different biomarkers may prove to be more useful than using single biomarkers and could be a potent biomarker in itself for disease diagnosis and prediction of progression from MCI to AD. Therefore, we decided to investigate the impact of combining MRI and CSF measures for the classification of AD and to predict future conversion from the prodromal stage of the disease, mild cognitive impairment (MCI).

The combination of MRI and CSF (Ewers et al., in press; Kohannim et al., 2010; Nettiksimmons et al., 2010) or MRI, CSF and FDG-PET (Kohannim et al., 2010; Walhovd et al., 2010; Zhang et al., 2011) has previously been investigated, but few studies have previously utilized baseline MRI and CSF measures for classification of individual subjects (Ewers et al., in press; Kohannim et al., 2010; Walhovd et al., 2010). No previous studies have systematically and extensively examined the prediction of MCI conversion at multiple future time points.

We have utilized a novel technique (OPLS) with one of the largest sample sizes to date from the ADNI cohort to combine the two measures for individual classification. We aimed to compare the ability of the combination of MRI and CSF, MRI alone and CSF alone, to 1) distinguish subjects with AD from healthy controls at baseline, 2) distinguish subjects with MCI from healthy controls at baseline and 3) use baseline MRI and CSF data to predict conversion from MCI to AD, using the follow-up diagnosis at multiple future time points.

## Material and methods

### Data

Data was downloaded from the Alzheimer's disease Neuroimaging Initiative (ADNI) database ([www.loni.ucla.edu/ADNI](http://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 MRI, PET and other biological markers are useful in clinical trials of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. 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](http://www.adni-info.org).

### Inclusion and diagnostic criteria

A total of 369 subjects were included in the current study (AD = 96, MCI = 162 and CTL = 111). The demographics of the cohort are given in Table 1. Data was acquired from 55 different sites (AD and MCI patients from 47 different sites and CTL from 51 different sites), with one to six subjects from each group acquired at each site. We included all subjects

who had successful MRI and CSF measures at baseline which passed the quality control steps outlined below. We use the term baseline to refer to when the first MRI scans were performed on the ADNI cohort and a clinical examination carried out to determine the subject group (i.e. AD, MCI and CTL). Other investigators have selected smaller samples but we feel that it is important to evaluate novel methods on the full range of available data to avoid bias. Of the 162 MCI subjects, 81 converted to AD (MCIc) at 12 month, 18 month, 24 month or 36 month follow-up. Subjects who did not convert at each time point are referred to as MCI stable (MCIs) here.

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) Mini mental state examination (MMSE) scores between 20 and 26, 2) Clinical dementia rating scale (CDR) of 0.5 or 1.0, 3) Met NINCDS/ADRDA criteria for probable AD, 4) Geriatric Depression Scale <6, and 5) Subjects were excluded if they had any other significant neurologic disease other than Alzheimer's disease.

**Mild cognitive impairment (General inclusion/exclusion criteria):** 1) Subjects had MMSE scores between 24 and 30 (inclusive), 2) Memory complaint, with objective memory loss measured by education adjusted scores on the 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, 5) Geriatric Depression Scale <6, and 6) Subjects were excluded if they had any other significant neurologic disease other than Alzheimer's disease.

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

### MRI and CSF

Both MRI and CSF data was downloaded from the ADNI website ([www.loni.ucla.edu/ADNI](http://www.loni.ucla.edu/ADNI)). The description of the data acquisition of the ADNI study can be found at [www.loni.ucla.edu/ADNI/research/Cores/index.shtml](http://www.loni.ucla.edu/ADNI/research/Cores/index.shtml). Briefly, data from 1.5 T scanners was used with data collected from a variety of MR-systems with protocols optimized for each type of scanner. The MRI protocol included a high resolution sagittal 3D T1-weighted MPRAGE volume (voxel size  $1.1 \times 1.1 \times 1.2 \text{ mm}^3$ ) acquired using a custom pulse sequence specifically designed for the ADNI study to ensure compatibility across scanners. Baseline CSF  $A\beta_{42}$ , t-tau and p-tau measures were used as features in this study. CSF was collected in the morning after an overnight fast using a 20- or 24-gauge spinal needle, frozen within 1 h of collection, and transported on dry ice to the ADNI Biomarker Core laboratory at the University of Pennsylvania

**Table 1**  
Subject characteristics.

	AD (n = 96)	MCI (n = 162)	CTL (n = 111)	MCIs (n = 81)	MCIc (n = 81)	p
Female/male	41/55	61/101	55/56	29/52	32/49	–
Age	74.4 ± 7.8	74.1 ± 7.2	75.6 ± 5.2	73.9 ± 7.2	74.3 ± 7.2	–
Education	15.1 ± 3.2	15.9 ± 2.9	15.7 ± 2.8	16.1 ± 2.9	15.8 ± 2.9	–
MMSE	23.5 ± 1.8 <sup>ab</sup>	26.9 ± 1.8 <sup>b</sup>	29.1 ± 0.9	27.2 ± 1.7	26.6 ± 1.8	<0.001
CDR	0.7 ± 0.2 <sup>a,b</sup>	0.5 <sup>b</sup>	0	0.5	0.5	<0.001
ADAS1	6.2 ± 1.4 <sup>ab</sup>	4.6 ± 1.4 <sup>b</sup>	2.8 ± 1.2	4.2 ± 1.5	5.0 ± 1.2	<0.001

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. Chi-square was used for gender comparison.

Two-way student *t*-test with Bonferroni correction was used for age and education and neuropsychological tests comparisons.

<sup>a</sup> Indicates significance compared to MCI group.

<sup>b</sup> Indicates significance compared to control group.

Medical Center. The complete descriptions of the collection and transportation protocols are provided in the ADNI procedural manual at [www.adni-info.org](http://www.adni-info.org). Full brain and skull coverage was required for the MRI datasets and detailed quality control carried out on all MR images according to previously published quality control criteria (Simmons et al., 2009, 2011).

#### Regional subcortical volume segmentation and cortical thickness parcellation

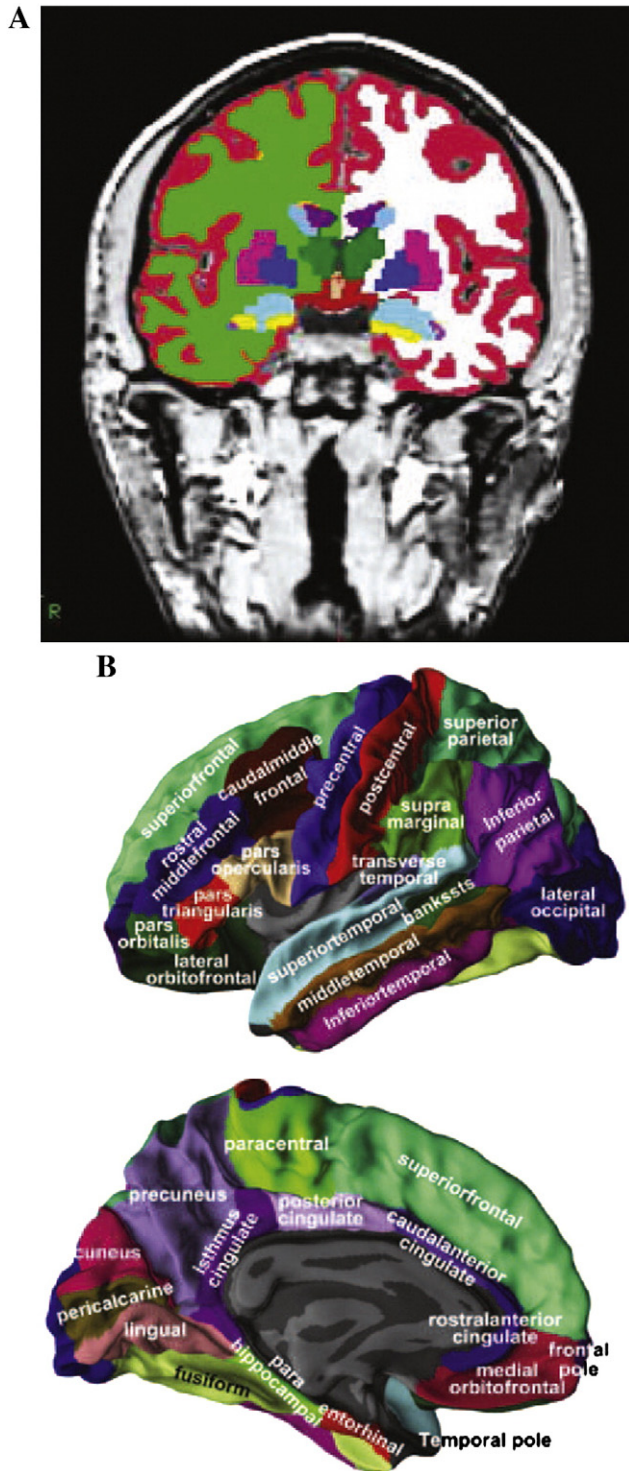
We applied the Freesurfer pipeline (version 4.5.0) to the MRI images to produce regional cortical thickness and subcortical volumetric measures (Fig. 1). Cortical reconstruction and subcortical volumetric segmentation include removal of non-brain tissue using a hybrid watershed/surface deformation procedure (Segonne et al., 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter volumetric structures (including hippocampus, amygdala, caudate, putamen, ventricles) (Fischl et al., 2002; Fischl et al., 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 subcortical volumes. Volumes of white matter hypointensities, optic chiasm, right and left vessel, and left and right choroid plexus were excluded from further analysis. White matter hypointensities were excluded since most subjects were characterized by zero values. Cortical thickness and subcortical volumetric measures from the right and left side were averaged, since this makes the data interpretation easier and the prediction accuracy does not significantly change if measures are averaged. In Westman et al. (2011a), we studied the whole ADNI cohort, rather than the sub-cohort who had both MRI and CSF measures at baseline and found for a AD vs. CTL model at baseline the averaged/not averaged sensitivity and specificity were 86.9%/86.3% and 86.7%/88.4%, respectively. Others have also applied this approach (Fjell et al., 2009; Walhovd et al., 2011). In total 57 MRI measures were used as input variables for the OPLS classification, 34 regional cortical thickness measures and 23 regional subcortical volumes (Table 2). All subcortical volumetric measures from each subject were normalized by the subject's intracranial volume, which is estimated based on an affine transform in Freesurfer.

Cortical thickness measures were not normalized and were used in their raw form. This segmentation approach has been used for neuropsychological-image analysis (Liu et al., 2010c; Liu et al., 2011), imaging-genetic analysis (Liu et al., 2010a; Liu et al., 2010b) and biomarker discovery (Thambisetty et al., 2010).

#### Multivariate data analysis

Before multivariate analysis, data was pre-processed using the software package SIMCA-P+ (Umetrics AB, Umea, Sweden), including 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.

After pre-processing data was analyzed using orthogonal partial least squares to latent structures (OPLS) (Bylesjo et al., 2007; Johan Trygg, 2002; Rantalainen et al., 2006; Westman et al., 2011c; Wiklund et al., 2008), a supervised multivariate data analysis method also included in the software package SIMCA (Umetrics AB, Umea, Sweden).



**Fig. 1.** Representations of ROIs included as candidate input variables in the multivariate OPLS model. (A) Regional subcortical volumes. (B) Regional cortical thickness measures.



**Table 2**  
Variable included in the OPLS analysis.

Cortical thickness measures	Subcortical volumetric measures	CSF
Banks of superior temporal sulcus	Third ventricle	A $\beta_{42}$
Caudal anterior cingulate	Fourth ventricle	t-tau
Caudal middle frontal gyrus	Brainstem	p-tau
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		

60 variables in total were included in the OPLS analysis, 34 cortical thickness measures, 23 subcortical volumetric measures and 3 CSF measures.

Supervised techniques such as OPLS use prior information about group information and do not attempt to explain as much variance in the original data as possible, which is the case for unsupervised methods like principal component analysis (PCA). Instead OPLS tries to maximize the covariance between the dependent and the independent variables. This is done via an inner relationship between the latent variables. Trygg and Wold (2002) first described the OPLS method a decade ago. The method combined the existing theory of partial least squares (PLS) regression (Wold et al., 1984; Wold et al., 2001a) and orthogonal signal correction (OSC) (Trygg and Wold, 2003; Wold et al., 2001b). PLS has previously been used in several studies to analyze MR-data (Levine et al., 2008; McIntosh and Lobaugh, 2004; Westman et al., 2009).

One way of describing PLS modeling (the relationship between two blocks of variables) is that it fits and aligns two models simultaneously, one model for X (independent variables) and one for Y (dependent variables). The aim is to model X and Y and to predict Y from X.

$$X = 1 * \bar{x}' + T * P' + E$$

$$Y = 1 * \bar{y}' + U * C' + F.$$

The terms  $1 * \bar{x}'$  and  $1 * \bar{y}'$  originate from the pre-processing step described above and represent the variable averages. Matrices T and U are information related to the observations. They describe if the observations are similar or dissimilar depending on the model and problem at hand. The X-loading matrix P' and the Y-weight matrix C' contain information connected to the variables. The last and final terms E and F are the residual matrices which contain the noise

(data not modeled) (Eriksson et al., 2006). As previously stated, OPLS is an extension of the PLS technique with orthogonal signal correction. This means that the information in Y can be used to decompose the X matrix into two blocks. One part is the structured variation (correlated variation) and the other is the variation orthogonal to Y (uncorrelated variation). To remove the variation in X which is not related to Y, OSC is based on three criteria where each component: 1) should involve the large systematic variations in X, 2) should be predictive by X, 3) should be orthogonal to Y (Trygg and Wold, 2002). OPLS and PLS give the same predictive accuracy, but the advantage of OPLS is that the information related to class separation (correlated variation) 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 (uncorrelated variation). Focusing the information related to class separation on the first component makes data interpretation easier (Wiklund et al., 2008). Similar to the PLS method, the OPLS method can be written as:

$$X = T_p P_p^T + T_o P_o^T + E$$

$$Y = T_p C_p^T + F$$

The  $T_p P_p^T$  block (the Y-predictive block) is the representation of the correlated variation and the  $T_o P_o^T$  block (the Y-orthogonal block) represents the uncorrelated variation.

The OPLS model is characterized by a  $Q^2(Y)$  value that describes the significance for separating groups (the predictability of the model).  $Q^2(Y)$  values  $>0.05$  are regarded as significant (Umetrics, 2008), where

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

where PRESS (predictive residual sum of squares) =  $\sum (y_{\text{actual}} - y_{\text{predicted}})^2$  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 dataset 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. Two groups are always compared using OPLS in this study and Y contains the prior information about group membership. Y is set to 1 for AD cases and 0 for CTL in the AD vs. CTL model and Y is set to 1 for MCI cases and CTL to 0 in the MCI vs. CTL model. The prediction value for a subject to belong to a group is equal to 1 for maximum likelihood, and is equal to 0 for minimum likelihood and vice versa depending on which group the subject belongs to. The cut off value for accepting the observation as correctly predicted is 0.5. When the model is generated each subject receives a predictive Y value while it is omitted from the modeling during the cross-validation rounds and then predicted on to the model.

Altogether 60 variables were used for OPLS analysis (Table 2). No feature selection was performed, meaning all measured variables were included in the analysis. MRI and CSF variables were analyzed separately, which means a separate OPLS model is created for each set of variables. The two models are then combined using hierarchical modeling. This means that the scores obtained from each variable model (the MRI OPLS model and the CSF OPLS model) are used as new latent variables in a hierarchical model.

Three OPLS models were created for AD vs. CTL and MCI vs. CTL, (1) MRI data, (2) CSF data and (3) a hierarchical model containing models (1) and (2).

As previously mentioned the OPLS models create a score for each individual subject, which tends to be close to 1 for AD patients and 0 for CTL subjects. Scores above 0.5 indicate a more AD-like characteristic and scores below 0.5 a more CTL-like characteristic. The AD vs. CTL OPLS model including MRI + CSF baseline data was used as a training set to prospectively predict MCI vs. MCIs at 12, 18, 24 and 36 month follow-up. This produced a discriminant score (the OPLS MRI + CSF based score) for each individual with MCI, reflecting the degree to which the individual's pattern (based on MRI and CSF) resembled the pattern of subjects with AD or the pattern of CTL subjects.

Variables can be plotted according to their importance for the separation of groups (Figs. 2 and 3). The plots show the MRI and CSF measures and their corresponding jack-knifed confidence intervals. Jack-knifing is used to estimate the bias and standard error. The cross-validation results from each model can be fed directly to jack-knifing. By doing so, the various sub-models generated by cross-validation are used to calculate the standard error of the different model parameters, which are then converted into confidence intervals via the t-distribution (Eriksson et al., 2006). Measures with confidence intervals that include zero have low reliability (Wiklund et al., 2008). Covariance is plotted on the y-axis, where

$$\text{Cov}(t, X_i) = t^T X_i / (N - 1)$$

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 and CSF measures below zero in the plots have lower values in AD and MCI subjects compared to CTL subjects, while MRI and CSF measures above zero are higher in AD and MCI subjects compared to CTL subjects in the model.

Sensitivity, specificity and area under the curve (AUC) were determined from the cross-validated prediction values of the OPLS models. The sensitivity was calculated as the percentage of AD subjects classified as AD subjects and the specificity as the percentage of CTL subjects classified as CTL subjects. The positive predictive value (PPV) (true positive/test outcome positive) and negative predictive value (NPV) (true negative/test outcome negative) were also evaluated. Finally, the positive and negative likelihood ratios ( $LR+ = \text{sensitivity} / (100 - \text{specificity})$  and  $LR- = (100 - \text{sensitivity}) / \text{specificity}$ ) 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).

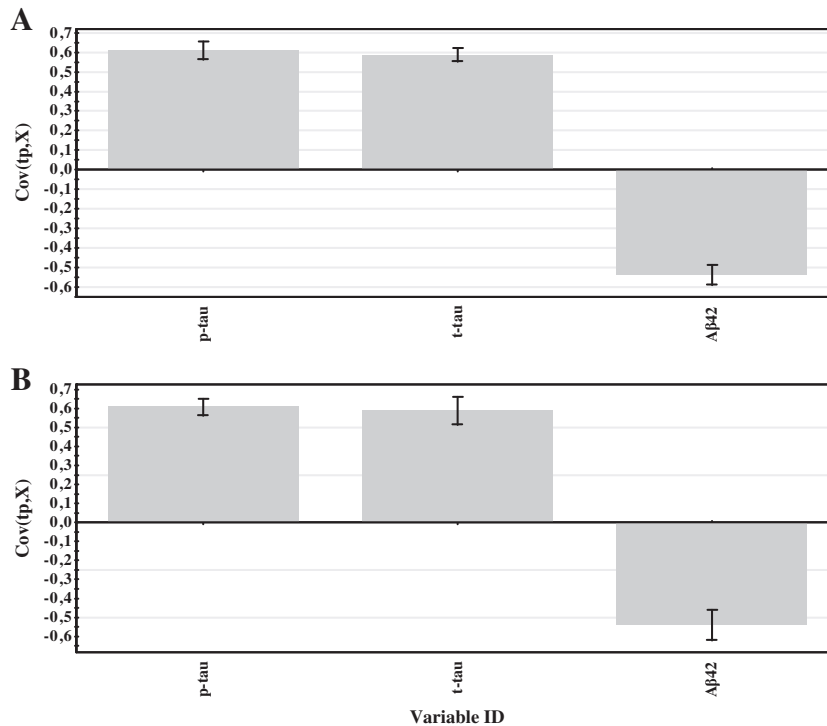
**Results**

*AD and MCI classification*

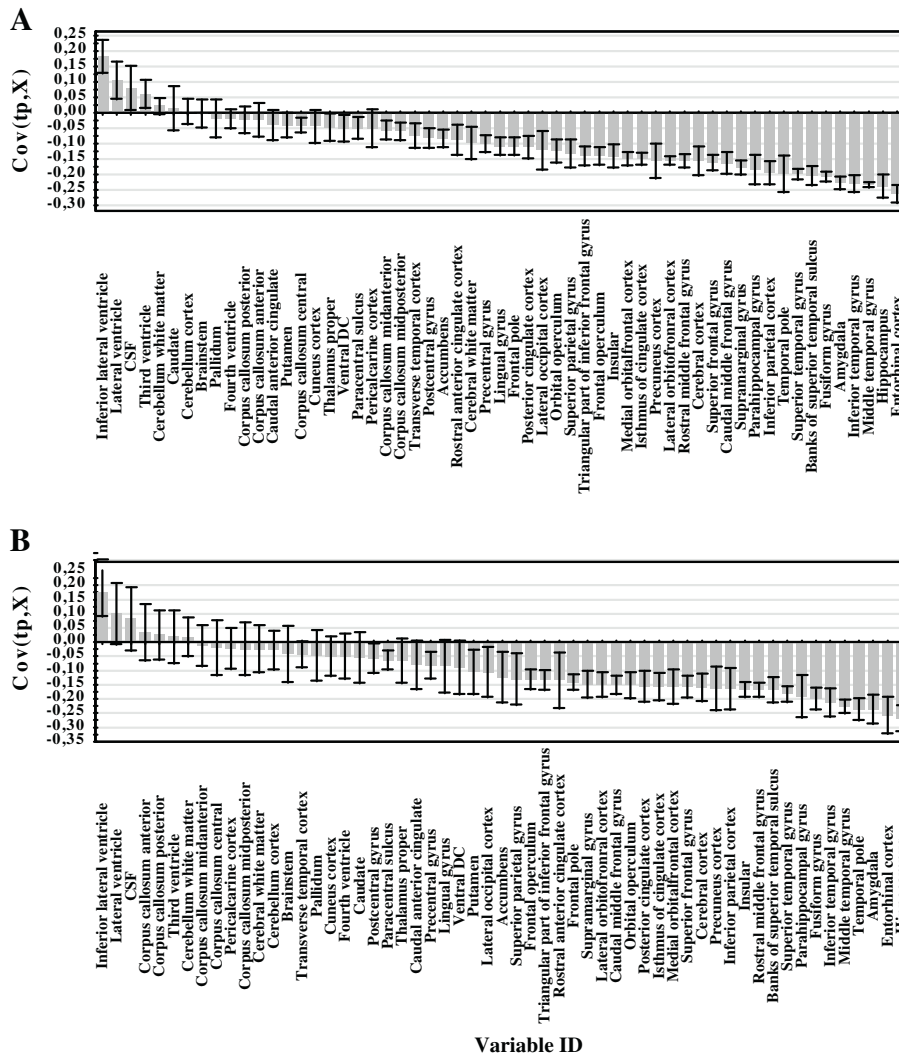
Table 3 shows the classification accuracy, sensitivity, specificity, positive and negative likelihood ratios ( $LR+$  and  $LR-$ ), positive predictive value (PPV), negative predictive value (NPV), AUC and  $Q^2(Y)$  of the different models.

For the AD vs. CTL model, combining the MRI measures with the CSF measures resulted in a classification accuracy of 91.8% (sensitivity = 88.5%, specificity = 94.6%, PPV = 93.4%, NPV = 90.5% and AUC = 0.958) compared to 87.0% for MRI only (sensitivity = 83.3%, specificity = 90.1%, PPV = 87.9%, NPV = 86.2% and AUC = 0.930) and 81.6% for CSF only (sensitivity = 84.4%, specificity = 79.3%, PPV = 77.9%, NPV = 85.4% and AUC = 0.861) at baseline. The positive likelihood ratio,  $LR+$ , doubled when combining the measures, compared to using MRI alone and quadrupled compared to using CSF measures alone.

The same pattern was observed for the MCI vs. CTL model, although less pronounced. Combining the MRI measures with the CSF measures



**Fig. 2.** CSF variables of importance for the separation between AD vs. CTL and MCI vs. CTL. (A) CSF measures for AD vs. CTL model. (B) CSF measures for MCI vs. CTL model. Measures above zero have a larger value in AD and MCI subjects compared to CTL and measures below zero have a lower value in AD and MCI subjects compared to CTL. 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.



**Fig. 3.** MRI variables of importance for the separation between AD vs. CTL and MCI vs. CTL. (A) MRI measures for AD vs. CTL model. (B) MRI measures for MCI vs. CTL model. Measures above zero have a larger value in AD and MCI subjects compared to CTL and measures below zero have a lower value in AD and MCI subjects compared to CTL. 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.

resulted in a higher classification accuracy, 77.6% (sensitivity = 72.8%, specificity = 84.7%, PPV = 87.4%, NPV = 68.1% and AUC = 0.876) compared to 71.8% for MRI alone (sensitivity = 66.7%, specificity = 79.3%, PPV = 82.4%, NPV = 62.0% and AUC = 0.815) and 70.3% for CSF alone (sensitivity = 66.7%, specificity = 75.7%, PPV = 80.0%, NPV = 60.9% and AUC = 0.749). The results from the combined MCI vs. CTL model were examined to investigate how many of the MCIc (n = 81) subjects who

converted at follow-up, were correctly classified as MCI within this model. This resulted in 90.1% of the MCIc being correctly classified as MCI. Only 54.3% of the MCIc group was correctly classified as MCI. Variables of importance for the separation between groups were the three CSF measures (p-tau, t-tau and Aβ<sub>42</sub>) and examples of brain regions contributing include hippocampus, amygdala, entorhinal cortex, inferior temporal gyrus, middle temporal gyrus, temporal pole, parahippocampal

**Table 3**  
Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios and AUC for the AD vs. CTL and MCI vs. CTL models at baseline.

	Accuracy	Sensitivity	Specificity	LR +	LR –	PPV	NPV	AUC	Q <sup>2</sup> (Y)
<i>AD vs. CTL</i>									
CSF	81.6 (75.8–86.3)	84.4 (75.8–90.3)	79.3 (70.8–85.8)	4.1 (2.8–5.9)	0.20 (0.12–0.32)	77.9%	85.4%	0.861	0.395
MRI	87.0 (81.7–90.9)	83.3 (74.6–89.5)	90.1 (83.1–94.3)	8.4 (4.8–14.8)	0.19 (0.12–0.29)	87.9%	86.2%	0.930	0.549
CSF + MRI	91.8 (87.2–94.8)	88.5 (80.6–93.5)	94.6 (88.7–97.5)	16.4 (7.5–35.8)	0.12 (0.07–0.21)	93.4%	90.5%	0.958	0.649
<i>MCI vs. CTL</i>									
CSF	70.3 (64.7–75.4)	66.7 (59.1–73.5)	75.7 (66.9–82.7)	2.7 (1.9–3.9)	0.44 (0.35–0.56)	80.0%	60.9%	0.749	0.185
MRI	71.8 (66.2–76.8)	66.7 (59.1–73.5)	79.3 (70.1–85.8)	3.2 (2.2–4.7)	0.42 (0.33–0.53)	82.4%	62.0%	0.815	0.263
CSF + MRI	77.6 (72.4–82.2)	72.8 (65.5–79.1)	84.7 (76.8–90.2)	4.8 (3.0–7.4)	0.32 (0.25–0.42)	87.4%	68.1%	0.876	0.400

AD = Alzheimer's disease, MCI = mild cognitive impairment, CTL = healthy control, confidence intervals within parenthesis, LR + = positive likelihood ratio, LR – = negative likelihood ratio, PPV = positive predictive values, NPV = negative predictive value and AUC = area under the curve. Q<sup>2</sup>(Y) values > 0.05 are regarded as statistically significant model.

gyrus and cerebral cortex (see Fig. 3 for more detailed information). The same variables were important for the separation between AD vs. CTL and for MCI vs. CTL, though less pronounced in the latter case with the order of importance slightly different.

*Predicting MCI conversion*

Combining MRI and CSF measures proved to give better results than using either MRI or CSF measures alone when distinguishing between AD vs. CTL and MCI vs. CTL. We also wanted to investigate if combining the measures would improve the ability of the AD vs. CTL model to predict conversion from MCI to AD. During the follow-up period of 36 months, 81 MCI subjects converted to AD and 81 remained stable (Table 1). Table 4 shows the classification accuracy, sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR-), positive predictive value (PPV), negative predictive value (NPV) and AUC for the MCI predictions using the MRI and CSF measures separately and combined. Again, the results showed that combining the measures gave the best accuracy, (68.5%) compared to an accuracy of 65.4% using either the MRI or CSF measures separately (PPV = 66.7% for the combined model compared to PPV = 65.4% for MRI and PPV = 62.6% for CSF).

Since the combined MRI + CSF model gave the best results we wanted to further investigate how the model could predict conversion from MCI to AD at different time points separately, but also to observe how the model predicted the subjects who remained stable (Table 5). All MCI subjects were predicted as either AD like or CTL like using the combined OPLS model (AD vs. CTL). The accuracy of predicting both MCI statuses at follow up was 58.6%, 65.8%, 66.4% and 66.1% for the 12, 18, 24 and 36 month time points respectively (Table 5). The positive prediction values showed a similar pattern with values of 32.2%, 44.7%, 55.1% and 61.8% for the same time points. The overall accuracy, PPV, NPV and AUC (Table 5) masked different results for the MCIC and MCIs groups. For the MCIC group the best results using the AD vs. CTL model were obtained at 18 month follow up (86.4% MCIC predicted as AD like) followed by 12 month follow up (82.9% MCIC predicted as AD like). The accuracies for correctly classifying MCIC subjects as AD like at baseline using the follow-up diagnosis at 24 months and 36 months were 75.4% and 68.0% respectively. The inverse pattern was observed for predicting MCIs using the combined model. The best result was observed at 36 month follow up (64.4% predicted as CTL like), followed by 24 month (60.7%), 18 month (57.7%) and 12 month follow up (52.0%). Investigating the total number of MCIC subjects who converted after baseline up to 36 month follow-up, 74.1% were correctly predicted using the AD vs. CTL model.

Table 6 shows how many MCI subject converted at each time point and the respective prediction accuracy. The table also illustrates how many of the converters remained in the study at the different time points.

**Discussion**

The use of biomarkers for diagnosis and prognosis in AD and MCI is of great importance. At least one abnormal biomarker among MRI, PET and CSF should be included alongside a clinical core of early and

**Table 5**  
MCI predictions at 12, 18, 24 and 36 months using the AD vs. CTL model at baseline.

	Number	Predicted as AD like	Predicted as CTL like	
		Sensitivity in bold <sup>a</sup>	Specificity in bold <sup>a</sup>	
<i>MCIC</i>				
12 months <sup>b</sup>	35	<b>82.9%</b> (29)	17.1% (6)	
18 months <sup>b</sup>	44	<b>86.4%</b> (38)	13.6% (6)	
24 months <sup>b</sup>	57	<b>75.4%</b> (43)	24.6% (14)	
36 months <sup>b</sup>	50	<b>68.0%</b> (34)	32.0% (16)	
<i>MCIs</i>				
12 months	127	48.0% (61)	<b>52.0%</b> (66)	
18 months	111	42.3% (47)	<b>57.7%</b> (64)	
24 months	89	39.3% (35)	<b>60.7%</b> (54)	
36 months	59	35.6% (21)	<b>64.4%</b> (38)	
<i>MCIC vs. MCIs</i>				
	Accuracy	AUC	PPV	NPV
12 months	58.6%	0.594	32.2%	91.7%
18 months	65.8%	0.647	44.7%	91.4%
24 months	66.4%	0.610	55.1%	79.4%
36 months	66.1%	0.578	61.8%	70.4%

AD = Alzheimer's disease, MCI = mild cognitive impairment, MCIC = MCI converter, MCIs = MCI stable, CTL = healthy control, AUC = area under the curve, PPV = positive predictive values and NPV = negative predictive value.

<sup>a</sup> Sensitivity at each time point (percentage of MCIC subjects predicted as AD) and specificity at each time point (percentage of MCI subjects predicted as CTL) in bold.

<sup>b</sup> Subjects who converted up to the specific time point (MCI to AD).

significant episodic memory impairment for a diagnosis of AD according to the new criteria (Dubois et al., 2007). A combination of biomarkers may prove to be more useful than single biomarkers for individual classification/prediction of subjects.

MRI and PET generate large volumes of data which can be analyzed using automated techniques. Multivariate data analysis tools such as support vector machines (SVM), linear discriminant analysis (LDA), PLS and OPLS (McEvoy et al., 2011; McIntosh and Lobaugh, 2004; Westman et al., 2011a; Zhang et al., 2011) can be powerful techniques for dealing with large and complex data sets, but automated computerized methods will only be implemented in clinical practice if they are carefully investigated and validated.

We have previously shown that combining a set of automated MRI measures of the brain together with manual measures of hippocampal volume gives good classification accuracy using OPLS (Westman et al., 2011c). Manual measures of different brain regions are time consuming and operator dependent however and are hence not regularly used in clinical settings. Therefore we also investigated the applicability of OPLS using the same automated regional subcortical volumes and cortical thickness measures as applied in the present study (Westman et al., 2011a). To further validate the OPLS technique using fully automated MRI measures as input, we compared the results to the Schelten's scale for visual assessment of medial temporal lobe atrophy, performed by an experienced neuroradiologist (Westman et al., 2011b). This showed that the automated approach gave superior results to the neuroradiologist. We have also previously investigated the value of combining automated regional MRI measures with magnetic resonance spectroscopy using OPLS (Westman et al., 2010) which showed a significant improvement compared with using either set of

**Table 4**  
Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios and AUC for MCI predictions on the baseline AD vs. CTL model.

	Accuracy	Sensitivity	Specificity	LR +	LR -	PPV	NPV	AUC
<i>MCIC vs. MCIs</i>								
CSF	65.4 (57.8–72.3)	76.4 (66.3–84.4)	54.3 (43.5–64.7)	1.7 (1.3–2.2)	0.43 (0.28–0.67)	62.6%	69.8%	0.668
MRI	65.4 (57.8–72.3)	65.4 (54.6–74.9)	65.4 (54.6–74.9)	1.9 (1.3–2.7)	0.53 (0.38–0.74)	65.4%	65.4%	0.731
CSF + MRI	68.5 (61.0–75.2)	74.1 (63.6–82.4)	63.0 (52.1–72.7)	2.0 (1.5–2.7)	0.40 (0.28–0.62)	66.7%	68.0%	0.760

AD = Alzheimer's disease, MCI = mild cognitive impairment, CTL = healthy control, confidence intervals within parenthesis, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, PPV = positive predictive values, NPV = negative predictive value and AUC = area under the curve.



**Table 6**  
MCI conversion and retention.

	12 months	18 months	24 months	36 months
Converted at 12 months	<b>35 (82.9%)</b>	30 (93.3%)	26 (84.6%)	15 (86.7%)
Converted at 18 months	–	<b>14 (85.7%)</b>	12 (83.3%)	8 (75.0%)
Converted at 24 months	–	–	<b>19 (57.9%)</b>	14 (50.0%)
Converted at 36 months	–	–	–	<b>13 (61.5%)</b>

Number of MCI subjects who converted at each time point (in bold) and remained in the study up to 36 month follow-up. Percentage of correctly predicted subjects on the AD vs. CTL baseline model are given in parentheses.

measures individually. The next natural step was to investigate the power of combining MRI and CSF biomarkers (which are both included in the new diagnostic criterion) for AD classification and MCI conversion.

Alternative multivariate techniques such as SVM and LDA have previously been utilized by others. We have previously shown that OPLS provides robust results when comparing the results from two large multi center studies (Westman et al., 2011a) and wished to build on this prior work by combining MRI and CSF measures together. There are many similarities between linear SVM and OPLS. Both methods can handle datasets with more dimensions than samples. Linear SVM weights illustrate the importance of the variables for the classification in descending order in the same way as the loading plots do for OPLS. The unique property of OPLS when compared to other linear regression methods is its ability to separate the modeling of correlated variation from structured noise (uncorrelated variation). The structured noise is defined as orthogonal variation in Y. At the same time the model maximizes the covariance between X and Y.

#### AD and MCI classification

Previous studies have combined different biomarkers for classification of individual subjects using MRI and FDG-PET (Fan et al., 2008; Hinrichs et al., 2009a, 2009b), MRI and CSF (Davatzikos et al., 2011; Kohannim et al., 2010; Nettiksimmons et al., 2010), as well as MRI, FDG-PET and CSF (Kohannim et al., 2010; Zhang et al., 2011). In the current manuscript we used the multivariate OPLS method with MRI and CSF data as input. Compared to previous studies we utilized the largest cohort to date and applied our technique for individual classification and prediction, in contrast to some other studies. We also investigated conversion from baseline MCI to AD at the full range of follow up time points up to 36 months (data downloaded from ADNI database on 14th September 2011). We did not use feature selection since it did not improve the performance of our models, though feature selection does increase computational time (Cuingnet et al., 2011). Further, the use of a limited set of pre-defined features might not reflect the true pattern of the CSF and MRI abnormalities (Fan et al., 2008; Zhang et al., 2011). In contrast to some previous studies which have combined MRI and CSF we aimed to predict conversion, introducing the MCI converters to the OPLS models as new and unseen data. Further, we investigated how conversion rates vary at different future time points (12, 18, 24 and 36 month follow-up).

Combining MRI with CSF resulted in better classification accuracy (91.8%) than using either measure separately (87% and 81.6% respectively) for the AD vs. CTL model at baseline. The positive likelihood ratio doubled when combining the measures, compared to using MRI alone and quadrupled compared to using CSF measures alone, which emphasizes the power of combining the measures (Table 3).

These results are in line with those of Zhang et al. (2011), who found a classification accuracy of 93.2% when combining MRI, FDG-PET and CSF, using ten-fold cross validation. Our results are slightly lower but we only used two biomarkers and the cohort we used was almost twice the size of that of Zhang et al. It is easier to realize higher accuracies with smaller groups, since classification techniques tend to focus on features which discriminate the groups, but may not be representative of the wider population. Since FDG-PET examinations can be expensive

it would have been interesting to see how the method of Zhang et al. performed with just the combination of MRI and CSF, but this data was not presented. On the other hand, obtaining CSF measurements can cause discomfort for some patients, which might warrant the combination of MRI and PET instead of MRI and CSF. However, it has previously been shown that the combination of MRI and CSF significantly improves classification accuracy and Walhovd et al. (2010) have reported that FDG-PET measures provide largely redundant information.

For the comparison between MCI and CTL subjects, again we found that the best accuracy was obtained by combining the two measures (77.6%). Within this model 90.1% of the MCIc subjects who converted within the investigated time period were correctly classified. These results are again similar to those of Zhang et al. (overall accuracy = 76.4% and accuracy for MCIc = 91.5%) but we used just two biomarkers. Kohannim et al. (2010) have also combined MRI, CSF and FDG-PET in a smaller cohort. Using leave-one-out cross-validation they received an accuracy of 88% for AD vs. CTL (present study 91.8%) when combining MRI and CSF. Adding FDG-PET increased the accuracy to 90.7%. For distinguishing between MCI vs. CTL they found an accuracy of 86.5% (present study 77.6%). Adding FDG-PET as a third biomarker did not improve the classification. Kohannim et al. used a limited set of MRI features. We believe that using both global and regional information from the entire brain, rather than just medial temporal lobe structures and ventricular volume adds to the predictive power (Westman et al., 2011c). Further, using information from the whole brain may also be more representative when investigating new and unseen data. We also considered a longer follow up period with full follow-up to 36 months and a larger dataset.

#### Predicting MCI conversion

It is of great importance to be able to build models which can predict conversion from MCI to AD on new and unseen data. Similar to baseline AD vs. CTL and MCI vs. CTL classification, combining MRI and CSF measures for MCI predictions gave the best accuracy, 68.5% compared to 65.4% using either MRI or CSF measures separately (Table 4). Since the combined model gave the best results predicting conversion over the investigated follow-up period of 36 months, we wanted to further investigate how well the AD vs. CTL model could predict conversion at different time points (12, 18, 24 and 36 months). We hypothesized that conversion could be predicted with a higher accuracy the closer the follow up time point was to the baseline. As expected the best results were obtained when predicting conversion from baseline at 12 and 18 month follow-up, where 82.9% and 86.4% of the MCIc were correctly predicted respectively. Looking at the number of subjects who converted up to the 36 month follow-up, 74.1% of the converters were correctly predicted using an AD vs. CTL model. The inverse pattern was observed for predicting MCIc, with the best predictions obtained at the later time points (24 and 36 months). The overall accuracy of predicting both MCIc and MCIs at each time point reflects this inverse pattern, 58.6%, 65.8%, 66.4% and 66.1% for 12, 18, 24 and 36 months respectively (Table 5). This illustrates the complex nature of predicting conversion. The likelihood for MCIc subjects to be predicted as CTL like increases with time. This is because, for example, some subjects predicted as MCIc at 12 months will convert to AD at a future time point and may already demonstrate an abnormal pattern at baseline. This pattern of improving performance with increasing time from baseline is also reflected in the PPV for each time point (32.2%, 44.7%, 55.1% and 61.8% for 12, 18, 24 and 36 months respectively).

Davatzikos et al. showed that with a combination of MRI and CSF measures applied to approximately half the number of subjects considered here they could classify MCIc vs. MCIs with a sensitivity of 84.2% and a specificity of 51.2% at an average follow-up time of 12 months with a standard deviation of 6 months (Davatzikos et al., 2011). However they did not apply their approach to the full ADNI



cohort as we have done here and did not provide results for multiple future time points. We found similar results for the AD vs. CTL model, using the MCI subjects as a test set which we believe is an approach of more practical use. Another recent study also aimed to predict conversion from MCI to AD, resulting in 71.4% correctly predicted MCIs, using a similar approach to this present study (Cui et al., 2011). We found a higher predictive accuracy for the combined MRI + CSF model than this latter study.

This study demonstrates the value of combining MRI and CSF measures for baseline AD vs. CTL classification, baseline MCI vs. CTL classification and predicting MCI to AD conversion using OPLS. However, the conversion prediction accuracy and positive predictive value are not as high as would be needed for clinical use currently and demonstrate the complex nature of MCI predictions. Future areas of research may allow performance to be further improved. This could include combining additional biomarkers, using neuropathologically proven AD cases for the training data and longer term follow up of the ADNI cohort. Finally, methodological improvements may improve classification and prediction further.

## Conclusion

Different biomarkers provide complementary information, which have been shown to be useful in AD and MCI diagnoses when used together (Apostolova et al., 2010; Fjell et al., 2010; Landau et al., 2010; Zhang et al., 2011). We show that the combination of MRI and CSF using OPLS as a tool more accurately classifies AD, MCI and CTL subjects at baseline compared to using either biomarker separately. At the moment there is no universally agreed gold standard for prediction accuracy, though in the broader field of diagnostics an accuracy of >90% could be considered as clinically relevant. We further show that combining MRI and CSF improves the prediction of future MCI conversion, though longer follow up times and further improvements in prediction accuracy are needed.

A potential limitation of combining biomarkers is the increased cost, so further work is needed to assess the cost-benefits of the improved accuracy attained. This may be dependent on whether the measures are to be used to aid diagnosis, as inclusion criteria in a clinical trial, or as primary or secondary clinical trial outcome measures. The ADNI dataset represents one of the largest sample sizes to date, but replication in independent samples is an obvious next step. One drawback of the ADNI data set is that it is not neuropathologically confirmed, however, it is very difficult to obtain large neuropathologically confirmed datasets in practice. Neuropathology as a gold standard in dementia has recently been questioned (Scheltens and Rockwood, 2011), however, with the suggestion that it should be considered as another complementary biomarker rather than the gold standard.

In conclusion this study has shown that combining automated MRI measures and CSF measures with OPLS improves the classification accuracy of AD vs. CTL and MCI vs. CTL, as well as the prediction of MCI to AD conversion when compared to using either MRI or CSF alone. This technique has the potential to serve as a complement to clinical assessment of AD, but further methodological improvement is needed before this is practical for predicting MCI conversion and helping to target appropriate populations for clinical trials.

## Acknowledgment

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. Hoffmann-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](http://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 Hjärnfonden, StratNeuro at Karolinska Institutet and Swedish Brain Power. Andrew Simmons was 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

- Apostolova, L.G., Hwang, K.S., Andrawis, J.P., Green, A.E., Babakhanian, S., Morra, J.H., Cummings, J.L., Toga, A.W., Trojanowski, J.Q., Shaw, L.M., Jack Jr., C.R., Petersen, R.C., Aisen, P.S., Jagust, W.J., Koeppe, R.A., Mathis, C.A., Weiner, M.W., Thompson, P.M., 2010. 3D PIB and CSF biomarker associations with hippocampal atrophy in ADNI subjects. *Neurobiol. Aging* 31, 1284–1303.
- 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.
- Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., Jin, J.S., the Alzheimer's Disease Neuroimaging, I., 2011. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS One* 6, e21896.
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, 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, 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., 2011. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging* 32 (12), 2322.e19–27. Epub 2010 Jul 1.
- 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.
- Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 6, 734–746.
- Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikström, C., Wold, S., 2006. Multi- and Megavariate Data Analysis (Part I – Basics and Principles 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., in press. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol. Aging* 2010 Dec 13. [Epub ahead of print].
- Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C., 2008. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage* 41, 277–285.
- 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), S69–84.

- 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.
- Fjell, A.M., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Walhovd, K.B., 2009. High consistency of regional cortical thinning in aging across multiple samples. *Cereb. Cortex* 19, 2001–2012.
- Fjell, A.M., Walhovd, K.B., Fennema-Notestine, C., McEvoy, L.K., Hagler, D.J., Holland, D., Brewer, J.B., Dale, A.M., 2010. CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease. *J. Neurosci.* 30, 2088–2101.
- Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., 2009a. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* 48, 138–149.
- Hinrichs, C., Singh, V., Xu, G., Johnson, S., 2009b. MKL for robust multi-modality AD classification. *Med. Image Comput. Comput. Assist. Interv.* 12, 786–794.
- Johan Trygg, S.W., 2002. Orthogonal projections to latent structures (O-PLS). *J. Chemometr.* 16, 119–128.
- Kohannim, O., Hua, X., Hibar, D.P., Lee, S., Chou, Y.-Y., Toga, A.W., Jack Jr., C.R., Weiner, M.W., Thompson, P.M., 2010. Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiol. Aging* 31, 1429–1442.
- Landau, S.M., Harvey, D., Madison, C.M., Reiman, E.M., Foster, N.L., Aisen, P.S., Petersen, R.C., Shaw, L.M., Trojanowski, J.Q., Jack Jr., C.R., Weiner, M.W., Jagust, W.J., 2010. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 75, 230–238.
- 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., 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* 31, 1375–1385.
- 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* 32, 1198–1206.
- McEvoy, L.K., Holland, D., Hagler, D.J., Fennema-Notestine, C., Brewer, J.B., Dale, A.M., 2011. Mild cognitive impairment: baseline and longitudinal structural MR imaging measures improve predictive prognosis. *Radiology* 259, 834–843.
- McIntosh, A.R., Lobaugh, N.J., 2004. Partial least squares analysis of neuroimaging data: applications and advances. *Neuroimage* 23, S250–S263.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34, 939–944.
- McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack Jr., C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., Mohs, R.C., Morris, J.C., Rossor, M.N., Scheltens, P., Carrillo, M.C., Thies, B., Weintraub, S., Phelps, C.H., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 263–269.
- Nettiksimons, J., Harvey, D., Brewer, J., Carmichael, O., DeCarli, C., Jack Jr., C.R., Petersen, R., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Beckett, L., 2010. Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline. *Neurobiol. Aging* 31, 1419–1428.
- Qizilbash, S., Chui, T., Tariot, Brodaty, Kaye, Erkinjuntti, 2002. Evidenced-based Dementia Practice. Blackwell Publishing, Oxford. 20–23 pp.
- 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 metabolomic–proteomic studies on a human prostate cancer xenograft model in mice. *J. Proteome Res.* 5, 2642–2655.
- Scheltens, P., Rockwood, K., 2011. How golden is the gold standard of neuropathology in dementia? *Alzheimers Dement.* 7, 486–489.
- 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.
- 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 multicentre MRI assessment of longitudinal changes in Alzheimer's disease: experience from the first 24 months. *Int. J. Geriatr. Psychiatry* 26 (1), 75–82.
- 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., Gunter, 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.
- Trygg, J., Wold, S., 2002. Orthogonal projections to latent structures (O-PLS). *J. Chemometr.* 16, 119–128.
- Trygg, J., Wold, S., 2003. O2-PLS, a two-block (X–Y) latent variable regression (LVR) method with an integral OSC filter. *J. Chemometr.* 17, 53–64.
- Umetrics, 2008. Users guide to SIMCA-P+. [http://www.umetrics.com/Content/Document%20Library/Files/UserGuides-Tutorials/SIMCA-P\\_12\\_UG.pdf](http://www.umetrics.com/Content/Document%20Library/Files/UserGuides-Tutorials/SIMCA-P_12_UG.pdf).
- Walhovd, K.B., Fjell, A.M., Brewer, J., McEvoy, L.K., Fennema-Notestine, C., Hagler Jr., D.J., Jennings, R.G., Karow, D., Dale, A.M., 2010. Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *AJNR Am. J. Neuroradiol.* 31, 347–354.
- Walhovd, K.B., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Fjell, A.M., 2011. Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiol. Aging* 32, 916–932.
- 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. Combining MRI and MRS to distinguish between Alzheimer's disease and healthy controls. *J. Alzheimers Dis.* 22, 171–181.
- Westman, E., Simmons, A., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M.W., Lovestone, S., Spenger, C., Wahlund, L.O., 2011a. AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *Neuroimage* 58, 818–828.
- Westman, E., Cavallin, L., Zhang, Y., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Spenger, C., Lovestone, S., Simmons, A., Wahlund, L.O., 2011b. Sensitivity and Specificity of Medial Temporal Lobe Visual Ratings and Multivariate Regional MRI Classification in Alzheimer's Disease. *PLoS ONE* 6 (7), e22506.
- Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J.S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Spenger, C., Wahlund, L.O., 2011c. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage* 54, 1178–1187.
- Wiklund, S., Johansson, E., Sjöström, 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.
- Wold, S., Ruhe, A., Wold, H., Dunn III, W.J., 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* 5, 735–743.
- Wold, S., Sjöström, M., Eriksson, L., 2001a. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab. Syst.* 58, 109–130.
- Wold, S., Trygg, J., Berglund, A., Antti, H., 2001b. Some recent developments in PLS modeling. *Chemometr. Intell. Lab. Syst.* 58, 131–150.
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867.