Journal of Alzheimer's Disease 44 (2015) 79–92
DOI 10.3233/JAD-140942
IOS Press

Generalizability of the Disease State Index Prediction Model for Identifying Patients Progressing from Mild Cognitive Impairment to Alzheimer's Disease

Anette Hall[a,*,1], Miguel Muñoz-Ruiz[a,b,1], Jussi Mattila[c], Juha Koikkalainen[c], Magda Tsolaki[d], Patrizia Mecocci[e], Iwona Kloszewska[f], Bruno Vellas[g], Simon Lovestone[h,i], Pieter Jelle Visser[j,k], Jyrki Lötjonen[c] and Hilkka Soininen[a,b], for the Alzheimer Disease Neuroimaging Initiative[2], the AddNeuroMed consortium, DESCRIPA and Kuopio L-MCI
[a]*Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio, Finland*
[b]*Department of Neurology, Kuopio University Hospital, Kuopio, Finland*
[c]*VTT Technical Research Centre of Finland, Tampere, Finland*
[d]*Aristotle University of Thessaloniki, Memory and Dementia Centre, "G Papanicolaou" General Hospital, Thessaloniki, Greece*
[e]*Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy*
[f]*Medical University of Lodz, Lodz, Poland*
[g]*UMR INSERM, University of Toulouse, France*
[h]*National Institute for Health Research (NIHR), London, UK*
[i]*King's College London, Institute of Psychiatry, London, UK*
[j]*VU University Medical Center, Amsterdam, The Netherlands*
[k]*Maastricht University, Maastricht, The Netherlands*

Handling Associate Editor: J. Wesson Ashford

Accepted 6 August 2014

Abstract.
Background: The Disease State Index (DSI) prediction model measures the similarity of patient data to diagnosed stable and progressive mild cognitive impairment (MCI) cases to identify patients who are progressing to Alzheimer's disease.
Objectives: We evaluated how well the DSI generalizes across four different cohorts: DESCRIPA, ADNI, AddNeuroMed, and the Kuopio MCI study.

*Correspondence to: Anette Hall, University of Eastern Finland, Institute of Clinical Medicine/Neurology, P.O. Box 1627, 70211 Kuopio, Finland. Tel.: +358 50 5392167; Fax: +358 17 162048; E-mail: anette.hall@uef.fi.

[1]These authors contributed equally to this work.

[2]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). 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. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

ISSN 1387-2877/15/$27.50

Methods: The accuracy of the DSI in predicting progression was examined for each cohort separately using 10 × 10-fold cross-validation and for inter-cohort validation using each cohort as a test set for the model built from the other independent cohorts using bootstrapping with 10 repetitions. Altogether 875 subjects were included in the analysis. The analyzed data included a comprehensive set of age and gender corrected magnetic resonance imaging (MRI) features from hippocampal volumetry, multi-template tensor-based morphometry, and voxel-based morphometry as well as Mini-Mental State Examination (MMSE), APOE genotype, and additional cohort specific data from neuropsychological tests and cerebrospinal fluid measurements (CSF).
Results: The DSI model was used to classify the patients into stable and progressive MCI cases. AddNeuroMed had the highest classification results of the cohorts, while ADNI and Kuopio MCI exhibited the lowest values. The MRI features alone achieved a good classification performance for all cohorts. For ADNI and DESCRIPA, adding MMSE, APOE genotype, CSF, and neuropsychological data improved the results.
Conclusions: The results reveal that the prediction performance of the combined cohort is close to the average of the individual cohorts. It is feasible to use different cohorts as training sets for the DSI, if they are sufficiently similar.

Keywords: Alzheimer's disease, computer-assisted diagnosis, dementia, magnetic resonance imaging (MRI), mild cognitive impairment

INTRODUCTION

Mild cognitive impairment (MCI) is considered to be a stage preceding dementia, in particular Alzheimer's disease (AD). Several studies have attempted to determine the rate of progression from MCI to AD. The meta-analysis conducted by Mitchell et al. indicated that less than half of the MCI cases actually progressed to AD, with an annual rate of conversion of around 10%, and many cases did not convert even after 10 years of follow-up [1]. As the prevalence of MCI is growing, it is important to detect initial subtle changes that could indicate a high risk for dementia in order to identify those patients who could benefit from treatment and preventive measures.

A variety of studies have been performed in order to investigate different aspects of the AD pathology based on proposed classifications and models [2, 3]. Nonetheless, there is no consensus about which biomarker or combination of biomarkers is most accurate for predicting the progression from MCI to AD. In particular, many studies have focused on imaging biomarkers which display brain atrophy with structural magnetic resonance imaging (MRI) techniques [4]. Automated imaging methods, such as multi-atlas segmentation of hippocampal volume (HCV) [5, 6] and morphometric methods, including tensor-based morphometry (TBM) [7–9] and voxel-based morphometry (VBM) [10, 11] have also been studied for the identification of MCI, AD, and frontotemporal dementia (FTD).

DSI is a statistical analysis method developed to function as part of a computer assisted diagnosis tool which integrates different data features from patients and combines them in order to provide global and individual scores for disease classification [12, 13]. The DSI method has previously been applied in predicting progression of MCI to AD [12, 13], as well as differentiating between controls, MCI, FTD, and AD [14]. However, these studies using the DSI have only included patients from a single cohort, making it important to validate the method further by applying it to data drawn from multiple cohorts. The main objective of this study was to evaluate how well the DSI prediction model would generalize to the situation where the model is built from different data as it is applied to. This is a vital issue when the clinical utility of prediction models is being evaluated.

In this study the DSI was applied for predicting the conversion from MCI to AD using four different cohorts: DESCRIPA [15], ADNI [16, 17], AddNeuroMed [18], and the Kuopio MCI cohort [19–22]. The analysis was done using HCV, multi-template TBM, and VBM methods, as well as with a commonly used cognitive screening test Mini-Mental State Examination (MMSE) and APOE genotype. In addition a set of neuropsychological tests and cerebrospinal fluid biomarkers (CSF) were included in intra-cohort analysis. The goal was to assess the cross-cohort performance of DSI with these biomarkers, particularly the imaging biomarkers, for predicting the conversion from MCI to AD.

METHODS

Patient data

Patients with MCI were chosen from four different previously described cohorts: the Alzheimer's Disease Neuroimaging Study (ADNI) [16], the AddNeuroMed consortium [18], the DESCRIPA study [15], and Kuopio longitudinal-MCI study [20–22]. A total of 875 subjects with MCI were included in this analysis.

Table 1
Demographic and clinical data. Values are mean (standard deviation) or number (percentage)

	ADNI	AddNeuroMed	DESCRIPA	Kuopio MCI
MCI patients	370	123	237	145
Age (years)[A]*	75.3 (7.3)	73.8 (5.7)	70.2 (7.9)	71.5 (5.0)
Gender (female)[B]*	134 (36%)	61 (50%)	136 (57%)	97 (67%)
Education (years)[C]*	15.6 (3.1)	8.9 (4.3)	9.3 (4.0)	7.0 (2.5)
Follow-up time (years)	2.9 (0.6)	1	2.2 (1.1)	2.6 (1.8)
Converted to AD	163 (44%)	23 (19%)	57 (24%)	54 (37%)
MMSE score[A]*	27.0 (1.8)	27.1 (1.7)	27.1 (2.3)	23.1 (3.6)
CSF $A\beta_{42}$ (pg/ml)	161 (52)	–	529 (272)	572 (220)
CSF t-tau (pg/ml)	104 (61)	–	507 (375)	456 (232)
CSF p-tau (pg/ml)	–	–	77 (51)	77 (25)
Hippocampal volume[A]*	3753 (612)	3912 (624)	3891 (652)	3840 (567)
APOE ε4 0/1/2[B]*	167/158/45 (45%/42%/12%)	70/35/5 (64%/32%/5%)	111/75/19 (54%/37%/9%)	76/51/16 (53%/36%/11%)

[A]One-way ANOVA, [B]Pearson Chi-Square test, [C]Kruskal-Wallis test *$p < 0.05$.

Informed consent was obtained from all subjects and the protocols and procedures were approved by the relevant Institutional Review Board at each data acquisition site and data coordination site.

Cohorts

Table 1 shows the demographic and clinical characteristics of the cohorts included in this study.

ADNI

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). The 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, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1,500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see http://www.adni-info.org.

Inclusion criteria for MCI patients were MMSE score between 24 and 30, memory problems with objective memory loss, Clinical Dementia Rating Scale (CDR) score of 0.5, not impaired significantly in other cognitive domains, preservation of activities of daily living and absence of dementia. In this study from this cohort we included 370 subjects, 36% female, with an average follow-up time 2.9 years.

AddNeuroMed

AddNeuroMed is a multi-center European study aimed at validating and identifying plasma-based and neuroimaging biomarkers for AD [18]. The data was collected from six centers and contained control, MCI, and AD patients; from this database 123 MCI patients (50% female) with MRI results were included in this study. In addition to the MRI, the MMSE, Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery, and APOE genotype were assessed. The follow-up time for MCI to AD conversion was one year and AD diagnosis was made according to the National Institute of Neurologic and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable AD. At baseline, all MCI subjects fulfilled the following components of the diagnostic

criteria for amnestic MCI: 1) memory complaint by patient, family, or physician; 2) normal activities of daily living; 3) MMSE score range between 24 and 30; 4) Geriatric Depression Scale score less than or equal to 5; 5) subject aged 65 years or above; 6) CDR memory score of 0.5 or 1; and 7) absence of dementia according to the NINCDS-ADRDA criteria. Patients with subjective memory complaints, but with a CDR memory and total score of 0 were not included in the MCI group.

DESCRIPA

The DESCRIPA study is a multicenter study conducted by the European Alzheimer's Disease Consortium on the development of screening guidelines and clinical criteria for the diagnosis of pre-dementia AD [15]. Memory clinics across Europe recruited patients aged 55 years and older with cognitive complaints, but without a diagnosis of dementia or a somatic, psychiatric or neurological disorder that could have caused the cognitive impairment. We included 237 subjects (57% female), who displayed an observable cognitive impairment, MCI, defined as a z-scored result of -1.5 or lower in neuropsychological tests. The average follow-up time was 2.2 years. The diagnosis of AD was based on the NINCDS-ADRDA criteria.

Kuopio L-MCI

The Kuopio MCI cohort consisted of 145 subjects diagnosed with MCI from the Kuopio longitudinal MCI study. These patients were pooled from two population based studies gathered in the University of Eastern Finland [19, 21–23]. One of the two studies used the MMSE for screening, with patients scoring under 24 or less invited to participate in the clinical phase [22]. MCI was diagnosed using the following criteria originally proposed by the Mayo Clinic Alzheimer's Disease Research Center: 1) memory complaint by patient, family, or physician; 2) normal activities of daily living; 3) normal global cognitive function; 4) objective impairment in memory or in one other area of cognitive function as evident by scores >1.5 S.D. below the age appropriate mean; 5) CDR score of 0.5; and 6) absence of dementia. In a few cases where all these criteria could not be evaluated due to missing data, MCI diagnosis was based on performing below the age-adjusted norms in at least one cognitive domain and having a score of 0.5 on the CDR scale. In this study were included 145 subjects (67% female). The mean follow-up time was 2.6 years. The diagnosis of AD was based on the NINCDS-ADRDA criteria.

Neuropsychological tests

From the ADNI dataset we chose the Alzheimer's Disease Assessment Scale-Cognitive Subscale as the neurological test set to be included in the DSI analysis for ADNI.

In AddNeuroMed the patients were evaluated with CERAD tests.

In DESCRIPA, the neuropsychological tests were divided into domains of learning, memory, language, executive function, and visuo-construction. As the subjects were collected from several different countries, the cognitive tests administered varied from center to center. The results were compared to the average scores of healthy controls according to age, gender, and education and a z-score was calculated.

The neuropsychological data after feature selection by relevance for Kuopio MCI included the Logical Memory Test from the Wechsler Memory Scale–Revised, Block Design and Vocabulary from the Wechsler Adults Intelligence Scale, Buschke Selective Reminding Test, Copy a Cube-test, and Constructional Praxis for CERAD. Full details on the battery of neuropsychological tests used in Kuopio MCI can be found elsewhere [21–23].

CSF analysis

In ADNI, levels of CSF concentrations of $A\beta_{42}$, t-tau, and p-tau were measured by flow cytometry using the monoclonal antibodies in the INNOBIA Alz Bio3 immunoassay kit (Innogenetics, Ghent, Belgium) and assayed with xMAP technology (Luminex, Austin, TX) [17].

In DESCRIPA, all CSF samples were analyzed in Sahlgrenska University Hospital, Mölndal, Sweden. The concentrations of $A\beta_{42}$, t-tau, and p-tau were measured with single parameter ELISA (Innotest β-amyloid 1–42; Innotest hTAU-Ag; Innogenetics, Ghent, Belgium).

In Kuopio MCI, the CSF sample was obtained by lumbar puncture at the baseline visit and stored in polypropylene tubes at –70°C until the analyses. The CSF $A\beta_{42}$, t-tau and p-tau levels were measured by commercial ELISA (Innogenetics, Ghent, Belgium) [24].

No CSF samples were available for the AddNeuroMed cohort.

APOE genotype

In ADNI, the APOE genotyping was performed using DNA extracted from EDTA blood. Polymerase chain reaction amplification was followed by HhaI restriction enzyme digestion, resolution on 4%

Metaphor Gel, and visualization by ethidium bromide staining [25].

In AddNeuroMed, the APOE genotype was determined from blood leukocytes with a standardized method [26].

In DESCRIPA, the *APOE* genotype was determined on genomic DNA extracted from EDTA blood with the polymerase chain reaction as described earlier [27].

In Kuopio MCI, the APOE genotype was determined from blood leukocytes. DNA was extracted by a standard phenol-chloroform extraction, and APOE genotypes were analyzed by the polymerase chain reaction and *Hha*I digestion as described previously [28].

MRI

MRI acquisition

The imaging protocol in ADNI and AddNeuroMed included a high-resolution sagittal 3D T1–weighted MPRAGE volume and axial proton density/T2-weighted fast spin echo images. In AddNeuroMed, the MRI data was acquired with 6 different 1.5T MRI systems [29]. The data acquisition was designed to be compatible with ADNI [30]. In DESCRIPA, the MRI scans were conducted in 9 different centers, each with their own scanners and protocols. The scanning was performed at either 1.0 or 1.5 Tesla and included 3D T1 weighted gradient echo and a fast fluid attenuated inversion recovery (FLAIR) sequence. In Kuopio MCI, the MR images were acquired with two different 1.5 T MRI scanners in the Department of Clinical Radiology, Kuopio University Hospital [19]. Anatomical high-resolution T1-weighted images were acquired using a 3D-MPRAGE sequence.

MRI feature extraction

Several feature extraction methods were used in the analysis of the T1-weighted MRI-images: VBM, multi-template TBM [7, 8], and multi-atlas segmentation of the HCV [31]. In the VBM analysis, the mean gray brain matter concentrations over the regions of interest (ROIs) were used as the features. In TBM, the MRI-images were non-rigidly registered to a mean anatomical reference MRI-image, and the mean values of the logarithm of the Jacobian (i.e., the local volume change as compared to the reference MRI-image) computed over the ROIs were used as the TBM features. The ROIs used were the 83 structures of the Hammer's atlas [32–34] and the combination of all of the ROIs. In multi-template TBM analysis, thirty templates (10 controls, 10 MCI patients, and 10 AD patients) from the ADNI [16, 17] were used, and the mean of the Jacobians of the 30 templates was used in the feature computation. Baseline HCV were measured using fast and robust multi-atlas segmentation presented elsewhere [5, 31].

The MRI features were corrected for age and gender as described previously [36]. The correction was performed using linear regression models determined between each MRI feature and age and gender using the control subjects of the ADNI study. By using the control subjects it was possible to remove any variations related to normal ageing and gender differences but all the disease specific variations were preserved.

Disease State Index

The DSI is measure for the similarity of a patient's test data with regard to diagnosed populations [12]. In this work, DSI values close to zero correspond to a similarity with the stable MCI cases, while values close to one indicate a progression toward AD. The DSI method can utilize any quantifiable data, and the resemblance of measurement values to the AD-converting population is calculated through a *fitness function*.

The fitness for feature i, as a function of measurement value x, is defined as

$$f_i(x) = \frac{FN_i(x)}{FN_i(x) + FP_i(x)}$$

where $FN_i(x)$ is the false negative errors and $FP_i(x)$ the false positive errors in the training data, when using x as the classification threshold.

For each measurement type, or *feature*, a *relevance* value is calculated. The relevance value ranges from 0 to 1 and indicates the feature's ability to differentiate between stable and progressive populations. The relevance is calculated from the sensitivity and specificity of the feature i:

$$relevance_i = sensitivity_i + specificity_i - 1,$$

with negative values set to a minimum value of zero. A relevance of zero denotes a feature completely unable to differentiate between the two populations, and relevance of one represents a feature for which there is no overlap between the two groups. Patient measurements are combined into a composite DSI value using a weighted average, where the fitness values are weighted according to their relevance. This process of evaluating DSI and relevance and combining features by weighted average can then be repeated recursively, until an overall DSI is obtained through:

$$Total\ DSI = \frac{\sum_i relevance_i \times fitness_i}{\sum_i relevance_i}$$

where i runs over the set of features available. A comprehensive description of the DSI method, including supplementary data detailing the calculations, has been published previously [12].

Data analysis

Due to the large number of available features, a simple feature selection was employed. Any features with a relevance under 0.1—corresponding to average sensitivity and specificity less than 0.55—were omitted from the analysis as they provide little information while increasing statistical noise. The cohorts were analyzed separately using 10×10-fold intra-cohort cross-validation for three different groups: 1) MRI results only; 2) MRI, MMSE scores, and *APOE ε4* allele count; 3) the previous group together with CSF and neuropsychological test results, depending on availability. Each cohort was also tested by using the three other cohorts as the training group using bootstrapping with 10 repetitions. DSI values under 0.5 were classified as stable cases and over 0.5 as patients who had progressed to AD. From the DSI values, we calculated AUC (area under the receiver-operator curve (ROC)), prediction accuracy, sensitivity, and specificity in order to evaluate the performance of the model.

In the cross-validation, fold-wise means, standard deviations and 95% confidence intervals were calculated. The confidence intervals were based on the standard error, assuming a normal distribution. The inter-cohort results were obtained with bootstrapping. From each training group with N cases, N random samples were chosen with replacement so that any particular case may appear several times in the training set. The means and standard deviations were then calculated by repeating the process 10 times for each tested patient.

RESULTS

Demographics

The demographic and clinical data for the study groups are presented in Table 1. There were significant differences between the cohorts in terms of age. The average age of the ADNI, which contained the oldest cases, was about 5 years higher than that for DESCRIPA, the youngest cohort.

The cohorts differed in terms of gender and years of education. There were only 36% of women in the ADNI cohort whereas for Kuopio MCI it was 67%. The ADNI cohort also differed significantly from the others by having subjects with an average of 15.6 years of education, while the Kuopio MCI cohort had the lowest level of education with an average of 7.0 years.

The AddNeuroMed study had a fixed follow-up time of 1 year, while the other three cohorts had average follow-up times of over 2 years. Subsequently, the AddNeuroMed study also had the lowest overall conversion percentage of 19%, while ADNI had the highest with 44%.

The average MMSE scores for ADNI, AddNeuroMed, and DESCRIPA were equivalent, while the average score for the Kuopio MCI cohort was significantly lower. This was also evident in the one-way ANOVA and post-hoc analyses.

HCV differed significantly between the groups; however, in the individual group comparisons, the statistically significant differences were found between the ADNI and DESCRIPA cohorts.

Pearson Chi-Square test detected significant differences between the groups for the presence of the *APOE ε4* allele. The *APOE ε4* alleles were most frequent in ADNI, with 55% carrying at least one allele, and least frequent in AddNeuroMed, where the frequency was only 36%.

Classification results

The DSI model was used to classify the patients into stable and AD-progressive MCI cases. Table 2 shows the AUC, accuracy, sensitivity, and specificity of the DSI prediction results obtained with intra-cohort cross-validation and inter cohort validation using independent training and test sets.

The accuracy of DSI in predicting progression was studied for each cohort separately using 10×10-fold cross-validation and for inter-cohort validation using one cohort as a test set for the model built from the other independent cohorts. Based on the mean AUCs, AddNeuroMed had the highest classification results of the cohorts, while ADNI and Kuopio MCI exhibited the lowest values.

The AUCs obtained for the DSI analyzed with MRI data only were 0.72 for ADNI, 0.73 for Kuopio MCI, 0.77 for DESCRIPA, 0.79 for AddNeuroMed, and 0.74 for the combined cohort. With the addition of the other biomarkers into the analysis, the AUC for the ADNI cohort increased to 0.74 with APOE and MMSE and to 0.76 also with CSF and neuropsychology. For

Table 2
Classification results. AUC, Area under the receiver-operator curve. Results from cross-validated cohorts are in the form: mean (standard deviation) [95% confidence interval]. Inter-cohort results are from ten bootstrapped DSI runs

	AUC	Accuracy	Sensitivity	Specificity
MRI with intra-cohort cross-validation				
ADNI	0.72 (0.07) [0.70–0.73]	0.67 (0.06) [0.66–0.69]	0.69 (0.11) [0.66–0.71]	0.66 (0.10) [0.64–0.68]
AddNeuroMed	0.79 (0.22) [0.75–0.84]	0.81 (0.11) [0.78–0.83]	0.74 (0.34) [0.67–0.80]	0.82 (0.12) [0.80–0.84]
DESCRIPA	0.77 (0.12) [0.74–0.79]	0.72 (0.08) [0.70–0.74]	0.62 (0.22) [0.57–0.66]	0.75 (0.10) [0.73–0.77]
Kuopio MCI	0.73 (0.14) [0.71–0.76]	0.67 (0.12) [0.65–0.70]	0.69 (0.20) [0.65–0.72]	0.67 (0.15) [0.64–0.70]
Combined cohort	0.74 (0.06) [0.73–0.75]	0.69 (0.05) [0.68–0.70]	0.69 (0.08) [0.67–0.70]	0.69 (0.06) [0.68–0.70]
MRI, APOE and MMSE with intra-cohort cross-validation				
ADNI	0.74 (0.07) [0.73–0.76]	0.69 (0.07) [0.67–0.70]	0.71 (0.11) [0.69–0.73]	0.67 (0.11) [0.65–0.69]
AddNeuroMed	0.82 (0.20) [0.78–0.86]	0.82 (0.11) [0.80–0.84]	0.80 (0.28) [0.75–0.86]	0.83 (0.11) [0.81–0.85]
DESCRIPA	0.78 (0.13) [0.75–0.80]	0.75 (0.09) [0.74–0.77]	0.68 (0.22) [0.63–0.72]	0.78 (0.10) [0.76–0.80]
Kuopio MCI	0.74 (0.11) [0.72–0.76]	0.68 (0.10) [0.66–0.70]	0.70 (0.18) [0.66–0.74]	0.67 (0.16) [0.64–0.70]
Combined cohort	0.76 (0.06) [0.75–0.78]	0.70 (0.05) [0.69–0.71]	0.70 (0.08) [0.69–0.72]	0.70 (0.05) [0.69–0.71]
MRI, APOE, MMSE, CSF and Neuropsychology with intra-cohort cross-validation				
ADNI	0.76 (0.08) [0.75–0.78]	0.70 (0.07) [0.68–0.71]	0.74 (0.10) [0.72–0.76]	0.66 (0.10) [0.64–0.68]
AddNeuroMed	0.83 (0.20) [0.79–0.86]	0.82 (0.12) [0.79–0.84]	0.78 (0.30) [0.72–0.83]	0.83 (0.12) [0.80–0.85]
DESCRIPA	0.81 (0.10) [0.79–0.83]	0.75 (0.08) [0.74–0.77]	0.66 (0.20) [0.62–0.70]	0.78 (0.10) [0.76–0.80]
Kuopio MCI	0.76 (0.13) [0.73–0.78]	0.70 (0.11) [0.68–0.72]	0.71 (0.20) [0.67–0.75]	0.70 (0.15) [0.67–0.73]
MRI using other cohorts as a training group				
ADNI	0.71 (0.003)	0.65 (0.01)	0.74 (0.02)	0.58 (0.03)
AddNeuroMed	0.75 (0.005)	0.75 (0.02)	0.71 (0.02)	0.76 (0.02)
DESCRIPA	0.72 (0.005)	0.69 (0.01)	0.60 (0.03)	0.72 (0.02)
Kuopio MCI	0.74 (0.003)	0.69 (0.02)	0.54 (0.04)	0.77 (0.005)
MRI, APOE and MMSE using other cohorts as a training group				
ADNI	0.72 (0.01)	0.67 (0.01)	0.67 (0.03)	0.67 (0.03)
AddNeuroMed	0.76 (0.01)	0.79 (0.01)	0.70 (0.02)	0.81 (0.02)
DESCRIPA	0.73 (0.01)	0.66 (0.03)	0.65 (0.03)	0.67 (0.05)
Kuopio MCI	0.74 (0.01)	0.68 (0.01)	0.56 (0.05)	0.75 (0.04)

DESCRIPA, there was a similar increase to 0.81 with the addition of all biomarker data. There was no discernible change for Kuopio MCI or AddNeuroMed outside of the 95% confidence intervals.

The same data sets were also used to analyze each cohort using the remaining three cohorts as the training group. CSF and neuropsychological data was not available for all cohorts, so no inter-cohort analysis could be made for those parameters.

A comparison of these inter-cohort results with those obtained through intra-cohort analysis revealed that with just the MRI data the only notable decrease in AUC was found for the DESCRIPA cohort. When also taking APOE and MMSE into account, there were small decreases in AUC also for ADNI and AddNeuroMed.

There were also some differences in accuracy, sensitivity, and specificity between the cohorts. Accuracy was slightly lower in the AddNeuroMed and DESCRIPA cohorts for both inter-cohort results, while in ADNI only for the MRI data. The most notable change in sensitivity and specificity were exhibited by the Kuopio MCI cohort, with specificity increasing and sensitivity dropping. Meanwhile, the ADNI cohort exhibited an opposite effect with just the MRI data.

Table 3 displays the AUCs of the feature groups using intra-cohort cross-validation of the DSI. CSF had the highest AUC of the features in both the Kuopio MCI (0.78) and DESCRIPA (0.79) cohorts, but not in ADNI (0.65). The neuropsychological test set was a good classifier for all of the cohorts (0.67–0.71), except in the AddNeuroMed (0.62). The MMSE score was not relevant in either the AddNeuroMed or Kuopio MCI cohorts, and so it was not included in the final analysis. With respect to the imaging techniques it was found that multi-template TBM was the best classifier in all cohorts with the exception of AddNeuroMed, where VBM achieved a higher AUC (0.82).

Figure 1 illustrates the results of the inter-cohort analysis for MRI, MMSE, and APOE, by showing the distribution of DSI values obtained for each cohort, in comparison to the distributions of the training group composed of the three other cohorts. The ADNI results

Table 3
Classification AUC of the feature groups using intra-cohort cross-validation. NR, not relevant (below the relevance threshold of 0.1); NA, not available. Results are in the form: mean (standard deviation) [95% confidence interval]

AUC	ADNI	AddNeuroMed	DESCRIPA	Kuopio MCI	Combined
CSF	0.65 (0.13)	NA	0.79 (0.21)	0.78 (0.28)	NA
	[0.62–0.68]		[0.74–0.83]	[0.72–0.84]	
Neuropsychological	0.71 (0.08)	0.62 (0.24)	0.67 (0.13)	0.70 (0.21)	NA
	[0.69–0.72]	[0.57–0.67]	[0.64–0.69]	[0.65–0.74]	
MMSE	0.59 (0.08)	NR	0.65 (0.15)	NR	0.60 (0.06)
	[0.58–0.61]		[0.62–0.68]		[0.58–0.61]
Genetic (*APOE ε4*)	0.63 (0.08)	0.69 (0.19)	0.62 (0.13)	0.59 (0.15)	0.64 (0.06)
	[0.61–0.64]	[0.66–0.73]	[0.59–0.64]	[0.56–0.62]	[0.63–0.65]
VBM	0.66 (0.09)	0.82 (0.17)	0.68 (0.13)	0.69 (0.15)	0.67 (0.06)
	[0.65–0.68]	[0.79–0.86]	[0.65–0.70]	[0.66–0.72]	[0.66–0.68]
TBM	0.71 (0.08)	0.77 (0.25)	0.74 (0.10)	0.70 (0.14)	0.74 (0.06)
	[0.69–0.72]	[0.72–0.82]	[0.72–0.76]	[0.67–0.73]	[0.72–0.75]
Hippocampal volume	0.66 (0.08)	0.67 (0.24)	0.69 (0.14)	0.66 (0.16)	0.68 (0.07)
	[0.65–0.68]	[0.62–0.72]	[0.66–0.71]	[0.63–0.70]	[0.66–0.69]

seem to closely follow the training group distributions, with a slight shift towards lower DSI values. AddNeuroMed seems to also follow the training group distribution quite closely, even though the S-MCI patients are again shifted to slightly smaller DSI values. DESCRIPA has more S-MCI cases between 0.4 and 0.6 DSI than in the training group. Finally, the Kuopio MCI results differ most from the training group distribution, with S-MCI cases shifted to higher and P-MCI to lower DSI values.

DISCUSSION

Memory problems are common among the older population, and it has been a challenge to identify MCI patients who will progress to dementia. There is a huge amount of data about which biomarkers are reliable predictors of progression to dementia in different longitudinal cohorts. Several studies have also been performed previously for each of the four separate cohorts examined in this study. While most previous studies have focused on different MRI parameters in a single cohort, this study applied a holistic approach, DSI, to a combination of four cohorts containing a total of 873 subjects with MCI. The major finding is that the prediction performance of the combined cohort (AUC = 0.76) is close to the average of the individual cohorts (AUC = 0.74–0.82). This study confirmed that it is feasible to use different cohorts as training sets for the DSI, as long as the cohorts share parameters.

While this study concentrates on the prediction of progression to AD, the DSI measure itself is continuous, running from zero to one. This allows for the estimation of disease severity or the confidence of the prediction. Clearly stable or progressive patients will receive very high or low DSI values, while borderline cases or those with ambiguous test results are expected to receive DSI values closer to 0.5. When integrated into a clinical decision support system, DSI not only provides a prediction of the progression, but also a measure of the severity through the DSI value, as well as information on how different biomarkers are contributing towards that prediction through the relevance and DSI value of each separate biomarker [12, 14].

Comparison to previous studies

The AUC obtained here by analyzing MRI data for the DESCRIPA cohort (0.77) is very similar to that obtained in previous studies. One study found that the AUC was higher for automatic (0.71) and manual hippocampal (0.71) measurements than for a medial temporal atrophy score (0.65) or lateral ventricle measurement (0.60) [37]. Another study, which included temporal lobe atrophy, achieved an AUC value of 0.81 when comparing converters and non-converters [38]. This suggests that automatic MRI techniques work well for predicting MCI to AD conversion.

A recent study report from the AddNeuroMed cohort comparing different multivariate techniques for automated classification of MRI data described an accuracy in the range 0.674 to 0.747 with AUC values between 0.748 and 0.827 for predicting conversion from MCI to AD [39]. As our results with the DSI method (0.79) are closely in line with these results, they further validate the hypothesis that the DSI and MRI processing methods used in this study are clearly comparable to established methods.

Another study attempted to compare stable-MCI and progressive-MCI from ADNI through a combination of

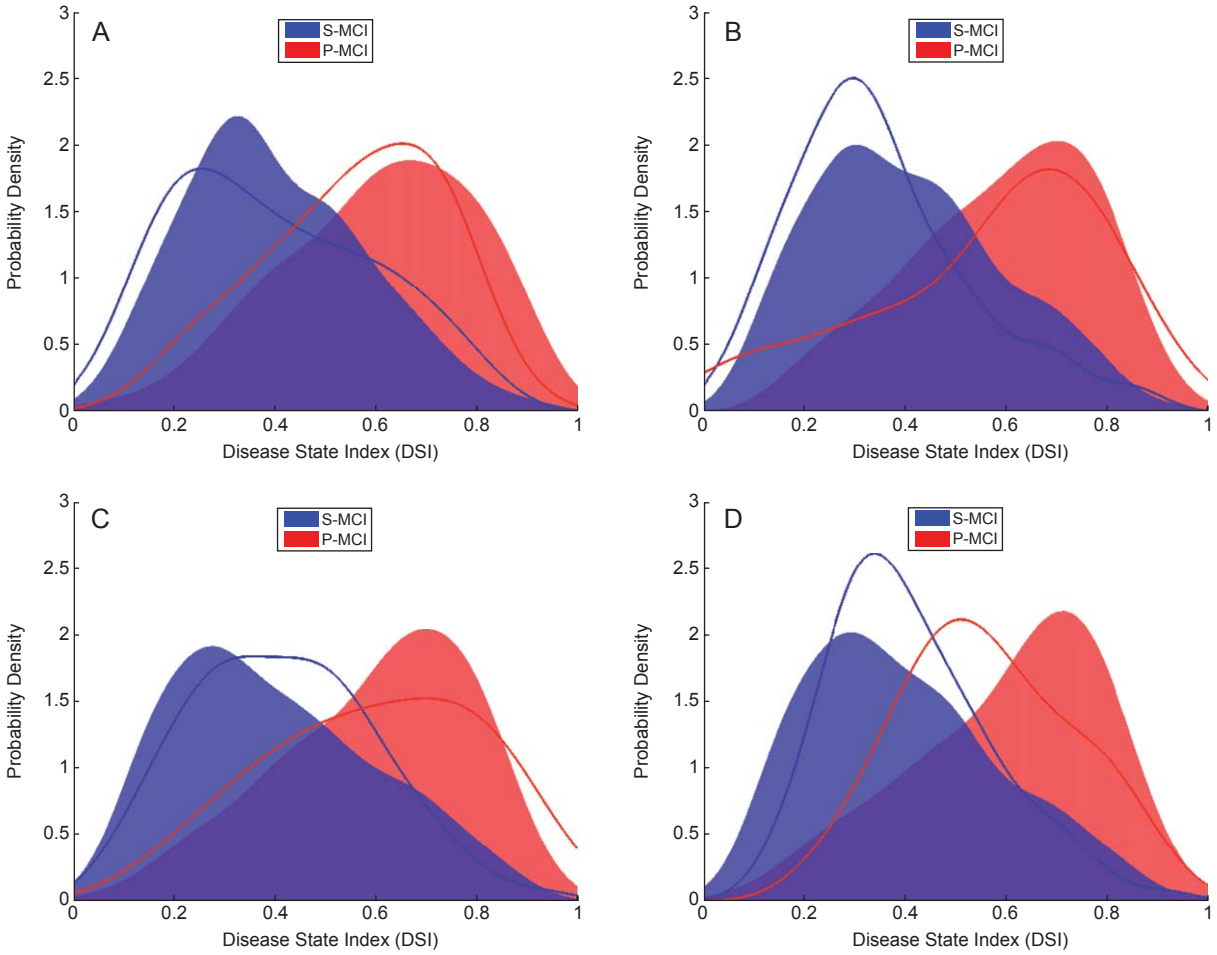

Fig. 1. Inter-cohort probability density distributions of DSI values for stable and progressive MCI. The training groups are shown as colored areas, and results for the tested cohorts as plotted lines. The four tested cohorts are A) ADNI, B) AddNeuroMed, C) DESCRIPA, and D) Kuopio MCI, while the training groups include the three other cohorts.

HCV, manifold based learning, cortical thickness, and TBM data, reaching an accuracy of 0.68 with linear discriminant analysis and an accuracy of 0.60 when using support vector machines [6].

Feature-specific analysis

We aimed to evaluate the overall performance of HCV, multi-template TBM, and VBM as AD biomarkers by applying them to these four large cohorts. The same methods were used as in our previous work comparing AD with MCI and FTD cases [14]. The most accurate imaging method in the AddNeuroMed cohort was VBM, while in the other cohorts it was TBM. HCV was shown to be a robust biomarker, exhibiting very similar accuracies in all cohorts. Our initial analysis included also single template TBM measurements, but these were removed from the final test set as they simply mirrored the results from multi-template TBM and did not affect overall accuracy.

The number of APOE4 alleles, which was the only genetic feature available, was also found to be a good predictor in most of the cohorts.

The MMSE score was relevant only in the ADNI and DESCRIPA cohorts. These results are affected by the fact that both ADNI and AddNeuroMed had MMSE as one of their inclusion criteria, limiting the possible range of values found in the datasets.

MCI patients are a heterogeneous group which is very unlikely to only include cases with impairments in the memory domain [40]. Therefore, the inclusion of tests evaluating all cognitive domains helps to orientate the diagnosis better than a single test or a screening test such as MMSE.

No CSF data was available for the AddNeuroMed cohort. CSF predicted progression only moderately in ADNI, but was particularly effective in both the DESCRIPA and Kuopio MCI cohorts. The method for analyzing CSF varied according to the cohort (ELISA in Kuopio MCI and DESCRIPA or xMAP Luminex in ADNI) and this difference in techniques for analyzing the CSF could have influenced the results [41] and explain the striking differences. The low AUC for CSF in the ADNI cohort was also detected in our previous study [42].

CSF was recommended to be used for predicting conversion from MCI to AD in a previous study from DESCRIPA with a 2 year follow-up [43]. Other cohorts also support using CSF data for predicting conversion from MCI to AD [44, 45].

Limitations and future directions

Since this study combines four large cohorts with different inclusion criteria and data collection techniques, it contains a wide variety of different tests and parameters. For this reason, the overall dataset is not homogenous, e.g., the different cohorts use a wide variety of neuropsychological test batteries, and the AddNeuroMed cohort was lacking CSF data. This is why neuropsychology and CSF could not be included in the inter-cohort analysis. Z-scoring of similar neuropsychological tests according to age and gender, as done in DESCRIPA, could be a way to enable comparison between different cohorts.

There was no standardized follow-up time in the cohorts examined in this work, which influences the results. When a neurodegenerative disease is at its early phase, there is only a small difference in the test results between cases having and not having the disease—normal human variability and measurement noise obscures the difference and makes the classification difficult [46]. Cases which obtain AD diagnosis after a three-year follow-up are more difficult to separate from stable cases than cases obtaining diagnosis after a one-year follow-up. This reasoning seems to agree with the results of Table 2; shorter follow-up time is related to higher classification performance. AddNeuroMed has the shortest follow-up time (1 year) and highest classification performance, while Kuopio and ADNI have the longest follow-up time (2.6 and 2.9 years) and lowest classification performance. On the other hand, when the follow-up time is short, the group of stable cases contains a larger share of progressing cases than when the follow-up time is long. This makes the training data of the classifier less optimal. In summary, if the follow-up time could be standardized in future studies, better results could be obtained when pooling the data.

The number of cases varied considerably across the cohorts. The number of ADNI cases was three times higher than the number of AddNeuroMed cases. The number of progressive cases in AddNeuroMed was also low—23 cases, only 19% of all AddNeuroMed cases. These reasons make the estimates of the classification performance less reliable in AddNeuroMed which is indicated by large confidence intervals in Table 2.

The ADNI population differs strikingly in terms of education from the other three cohorts in this study. This is important since education could be able to preserve cognition by increasing the cognitive reserve. However, data from AddNeuroMed did not reveal any association between education level and cortical thickness for MCI [47]. Instead in controls, more years of education were associated with higher cortical thickness values, whereas in AD more education was associated with a lower level of cortical thickness.

Age is an important factor in the progression of AD, and we have taken this into account by normalizing the MRI results for age and gender using the ADNI control data [36]. As the correction was done using only data from ADNI, it may cause slight bias toward the cohort. The inter-cohort DSI distribution figures show that ADNI matches the training group slightly better than the other cohorts. However, no age or gender correction was done for the MMSE, APOE, CSF, or neuropsychological biomarkers, with the exception of DESCRIPA neuropsychology, which was already Z-scored. Further utilization of corrections for age, gender, education, or genotype could improve classification results. Using genotype for correction would be particularly interesting, since APOE has been shown to be connected to the age of onset for AD [48]. One possibility would be to select subgroups of patients with, for example a certain age and genotype, and comparing patients within their subgroup [36].

With respect to the different methods of CSF analysis, as mentioned before, the lack of standardization may explain the variability of the results between centers [41]. The application of a quality control program [49] and harmonization of sample collection and handling could facilitate future comparison among different centers [50].

Although ADNI and AddNeuroMed both follow the same MRI data acquisition scheme [51], we should presume that there is some scanner variability that could affect the results obtained from the Kuopio MCI and

DESCRIPA cohorts. However, a study using VBM with different scanners did not find this scanner variability to be a confounding factor [52]. In future studies we could widen the list of biomarkers, particularly by including other imaging methods already recommended in the research criteria [53] such as FDG-PET.

In addition, the amount of amyloid deposition in the brain could be studied and compared with the rate of neurodegeneration found in structural MRI methods [54], or functional imaging methods such as resting state functional MRI [55], which presumably could detect changes in the connectivity earlier than can be found with the structural changes based on atrophy. Complementing the current results with more imaging methods or other biomarkers could help in achieving higher prediction accuracies. Furthermore it could also be advisable to survey the presence and rate of *APOE* $\varepsilon 2$, considered to be a protective factor for developing dementia, as done previously for AddNeuroMed [56].

It is important to clearly define the criteria for classifying the different MCI stages and separating the amnestic variants from the non-amnestic subtypes [57]. The use of standardized criteria would homogenize the cohorts and make it possible to achieve less biased results. While the other cohorts included only amnestic MCI cases, the DESCRIPA cohort also had non-amnestic MCI cases, which have a different conversion rate and disease profile.

A longer follow-up in these cohorts could help trace the conversion to AD from MCI [58], particularly with methods for assessing neurodegeneration, which appears later in the disease progression. This would not only help in ascertaining which cases will progress to AD, but also to other types of dementia [59] such as FTD, vascular dementia, or Lewy body dementia, as has been done in previous studies [60].

CONCLUSION

The main goal of this study was to evaluate how well a classification model built for predicting progression from MCI to AD would generalize across four different cohorts with a wide variety of different biomarkers and test results. DSI made it possible to discern not only the differences between the cohorts, but also if the performance of the DSI is altered after inclusion of very large and diverse populations.

The four cohorts in the study are heterogeneous, differing with regard to inclusion criteria, follow-up time, patients' age, and years of education. The results show that the intra-cohort prediction efficiency of the combined cohort is close to the average of the individual cohorts. However, prediction accuracy is slightly lowered in inter-cohort analysis. This highlights that while it is feasible to use different cohorts as training sets for each other, for accurate predictions the cohorts need to be sufficiently similar.

Additionally, this work helps in verifying the performance of the automated MRI image analysis methods, as well as identifying how they perform in these different cohorts. The set of MRI methods was very comprehensive, that including additional features known to be indicative of progressive MCI such as neuropsychological test sets or CSF had only a small effect on the overall predictive capabilities.

ACKNOWLEDGMENTS

This study was partly funded by the European Union 7th Framework Program, PredictAD, 5th Framework Program DESCRIPA study QLK-6-CT-2002-02455, the Health Research council of the Academy of Finland (HS), strategic funding for UEFBRAIN from University of Eastern Finland (AH, HS) and supported by European Medical Information Framework, EMIF, Grant Agreement No: 115372.

Regions of interest included in the MRI analysis were obtained from the Hammer's Atlas, ©Copyright Imperial College of Science, Technology and Medicine, Alexander Hammers and University College London 2011. All rights reserved.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharma-

ceutical Company. The Canadian Institutes of Health Research is providing funds to Rev December 5, 2013 support ADNI clinical sites in Canada. Private sector contributions 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 Southern California.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=2491).

REFERENCES

[1] Mitchell AJ, Shiri-Feshki M (2009) Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand* **119**, 252-265.

[2] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment 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** 270-279.

[3] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, Delacourte A, Frisoni G, Fox NC, Galasko D, Gauthier S, Hampel H, Jicha GA, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Sarazin M, de Souza LC, Stern Y, Visser PJ, Scheltens P (2010) Revising the definition of Alzheimer's disease: A new lexicon. *Lancet Neurol* **9**, 1118-1127.

[4] Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC, Alzheimer's Disease Neuroimaging Initiative (ADNI) (2009) Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr Alzheimer Res* **6**, 347-361.

[5] Lötjönen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, Rueckert D, Alzheimer's Disease Neuroimaging, Initiative (2010) Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* **49**, 2352-2365.

[6] Wolz R, Julkunen V, Koikkalainen J, Niskanen E, Zhang DP, Rueckert D, Soininen H, Lötjönen J, Alzheimer's Disease Neuroimaging, Initiative (2011) Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS One* **6**, e25446.

[7] Koikkalainen J, Lötjönen J, Thurfjell L, Rueckert D, Waldemar G, Soininen H (2011) Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease. *Neuroimage* **56**, 1134-1144.

[8] Brun CC, Leporè N, Pennec X, Lee AD, Barysheva M, Madsen SK, Avedissian C, Chou YY, de Zubicaray GI, McMahon KL, Wright MJ, Toga AW, Thompson PM (2009) Mapping the regional influence of genetics on brain structure variability–a tensor-based morphometry study. *Neuroimage* **48**, 37-49.

[9] Hua X, Leow AD, Parikshak N, Lee S, Chiang MC, Toga AW, Jack CR, Jr., Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging, Initiative (2008) Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: An MRI study of 676 AD, MCI, and normal subjects. *Neuroimage* **43**, 458-469.

[10] Chételat G, Landeau B, Eustache F, Mèzenge F, Viader F, de la Sayette V, Desgranges B, Baron JC (2005) Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *Neuroimage* **27**, 934-946.

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

[12] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van Gils M, Waldemar G, Soininen H, Lötjönen J, Alzheimer's Disease Neuroimaging, Initiative (2011) A disease state fingerprint for evaluation of Alzheimer's disease. *J Alzheimers Dis* **27**, 163-176.

[13] Mattila J, Soininen H, Koikkalainen J, Rueckert D, Wolz R, Waldemar G, Lötjönen J (2012) Optimizing the diagnosis of early Alzheimer's disease in mild cognitive impairment subjects. *J Alzheimers Dis* **32**, 969-979.

[14] Muñoz-Ruiz MÁ, Hartikainen P, Hall A, Mattila J, Koikkalainen J, Herukka SK, Julkunen V, Vanninen R, Liu Y, Lötjönen J, Soininen H (2013) Disease State Fingerprint in frontotemporal degeneration with reference to Alzheimer's disease and mild cognitive impairment. *J Alzheimers Dis* **35**, 727-739.

[15] Visser PJ, Verhey FR, Boada M, Bullock R, De Deyn PP, Frisoni GB, Frolich L, Hampel H, Jolles J, Jones R, Minthon L, Nobili F, Olde Rikkert M, Ousset PJ, Rigaud AS, Scheltens P, Soininen H, Spiru L, Touchon J, Tsolaki M, Vellas B, Wahlund LO, Wilcock G, Winblad B (2008) Development of screening guidelines and clinical criteria for predementia Alzheimer's disease. The DESCRIPA Study. *Neuroepidemiology* **30**, 254-265.

[16] Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, Trojanowski JQ, Toga AW, Beckett L (2005) The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am* **15**, 869-877.

[17] Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E, Morris JC, Petersen RC, Saykin AJ, Schmidt ME, Shaw L, Shen L, Siuciak JA, Soares H, Toga AW, Trojanowski JQ, Alzheimer's Disease Neuroimaging, Initiative (2013) The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimers Dement* **9**, e111-e194.

[18] Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, Spenger C, Tsolaki M, Vellas B, Wahlund LO, Ward M, AddNeuroMed, Consortium (2009) AddNeuroMed–the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. *Ann N Y Acad Sci* **1180**, 36-46.

[19] Julkunen V, Niskanen E, Muehlboeck S, Pihlajamäki M, Könönen M, Hallikainen M, Kivipelto M, Tervo S, Vanninen R, Evans A, Soininen H (2009) Cortical thickness analysis to detect progressive mild cognitive impairment: A reference to Alzheimer's disease. *Dement Geriatr Cogn Disord* **28**, 404-412.

[20] Julkunen V, Niskanen E, Koikkalainen J, Herukka SK, Pihlajamäki M, Hallikainen M, Kivipelto M, Muehlboeck S, Evans AC, Vanninen R, Hilkka, Soininen (2010) Differences in cortical thickness in healthy controls, subjects with mild cognitive impairment, and Alzheimer's disease patients: A longitudinal study. *J Alzheimers Dis* **21**, 1141-1151.

[21] Hänninen T, Hallikainen M, Tuomainen S, Vanhanen M, Soininen H (2002) Prevalence of mild cognitive impairment: A population-based study in elderly subjects. *Acta Neurol Scand* **106**, 148-154.

[22] Kivipelto M, Helkala EL, Laakso MP, Hänninen T, Hallikainen M, Alhainen K, Soininen H, Tuomilehto J, Nissinen A (2001) Midlife vascular risk factors and Alzheimer's disease in later life: Longitudinal, population based study. *BMJ* **322**, 1447-1451.

[23] Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hänninen T, Laakso MP, Hallikainen M, Vanhanen M, Nissinen A, Helkala EL, Vainio P, Vanninen R, Partanen K, Soininen H (2004) Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol Aging* **25**, 303-310.

[24] Herukka SK, Pennanen C, Soininen H, Pirttilä T (2008) CSF Abeta42, tau and phosphorylated tau correlate with medial temporal lobe atrophy. *J Alzheimers Dis* **14**, 51-57. Erratum in: *J Alzheimers Dis* **15**, 347.

[25] Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM, Stein JL, Moore JH, Farrer LA, Green RC, Bertram L, Jack CR Jr, Weiner MW, Alzheimer's Disease Neuroimaging, Initiative (2010) Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimers Dement* **6**, 265-273.

[26] Liu Y, Paajanen T, Westman E, Wahlund LO, 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, AddNeuroMed, Consortium (2010) Effect of APOE ε4 allele on cortical thicknesses and volumes: The AddNeuroMed study. *J Alzheimers Dis* **21**, 947-966.

[27] Norberg J, Graff C, Almkvist O, Ewers M, Frisoni GB, Frölich L, Hampel H, Jones RW, Kehoe PG, Lenoir H, Minthon L, Nobili F, Olde Rikkert M, Rigaud AS, Scheltens P, Soininen H, Spiru L, Tsolaki M, Wahlund LO, Vellas B, Wilcock G, Elias-Sonnenschein LS, Verhey FR, Visser PJ (2011) Regional differences in effects of APOE ε4 on cognitive impairment in non-demented subjects. *Dement Geriatr Cogn Disord* **32**, 135-142.

[28] Tsukamoto K, Watanabe T, Matsushima T, Kinoshita M, Kato H, Hashimoto Y, Kurokawa K, Teramoto T (1993) Determination by PCR-RFLP of apo E genotype in a Japanese population. *J Lab Clin Med* **121**, 598-602.

[29] Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, Kłstrokoszewska I, Wahlund LO, Soininen H, Lovestone S, Evans A, Spenger C, AddNeuroMed, Consortium (2009) MRI measures of Alzheimer's disease and the AddNeuroMed study. *Ann N Y Acad Sci* **1180**, 47-55.

[30] Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell LJ, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DL, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* **27**, 685-691.

[31] Lötjönen J, Wolz R, Koikkalainen J, Julkunen V, Thurfjell L, Lundqvist R, Waldemar G, Soininen H, Rueckert D, Alzheimer's Disease Neuroimaging, Initiative (2011) Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease. *Neuroimage* **56**, 185-196.

[32] Heckemann R, Hajnal R, Aljabar P, Rueckert D, Hammers A (2006) Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* **33**, 115-126.

[33] Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, Mitchell TN, Brooks DJ, Duncan JS (2003) Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp* **19**, 224-247.

[34] Gousias IS, Rueckert D, Heckemann RA, Dyet LE, Boardman JP, Edwards AD, Hammers A (2008) Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *Neuroimage* **40**, 672-684.

[35] Muñoz-Ruiz MÁ, Hartikainen P, Koikkalainen J, Wolz R, Julkunen V, Niskanen E, Herukka SK, Kivipelto M, Vanninen R, Rueckert D, Liu Y, Lötjönen J, Soininen H (2012) Structural MRI in frontotemporal dementia: Comparisons between hippocampal volumetry, tensor-based morphometry and voxel-based morphometry. *PLoS One* **7**, e52531.

[36] Koikkalainen J, Pölönen H, Mattila J, van Gils M, Soininen H, Lötjönen J (2012) Improved classification of Alzheimer's disease data via removal of nuisance variability. *PLoS One* **7**, e31112.

[37] Clerx L, van Rossum IA, Burns L, Knol DL, Scheltens P, Verhey F, Aalten P, Lapuerta P, van de Pol L, van Schijndel R, de Jong R, Barkhof F, Wolz R, Rueckert D, Bocchetta M, Tsolaki M, Nobili F, Wahlund LO, Minthon L, Frölich L, Hampel H, Soininen H, Visser PJ (2013) Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. *Neurobiol Aging* **34**, 2003-2013.

[38] Chincarini A, Bosco P, Gemme G, Esposito M, Rei L, Squarcia S, Bellotti R, Minthon L, Frisoni G, Scheltens P, Frölich L, Soininen H, Visser PJ, Nobili F, Alzheimer's Disease Neuroimaging, Initiative (2013) Automatic temporal lobe atrophy assessment in prodromal AD: Data from the DESCRIPA study. *Alzheimers Dement* **10**, 456-467.

[39] Aguilar C, Westman E, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Spenger C, Simmons A, Wahlund LO (2013) Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Res* **212**, 89-98.

[40] Nordlund A, Rolstad S, Hellström P, Sjögren M, Hansen S, Wallin A (2005) The Goteborg MCI study: Mild cognitive impairment is a heterogeneous condition. *J Neurol Neurosurg Psychiatry* **76**, 1485-1490.

[41] Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, Herukka SK, van der Flier WM, Blankenstein MA, Ewers M, Rich K, Kaiser E, Verbeek M, Tsolaki M, Mulugeta E, Rosèn E, Aarsland D, Visser PJ, Schröder J, Marcusson J, de Leon M, Hampel H, Scheltens P, Pirttilä T, Wallin A, Jönhagen ME, Minthon L, Winblad B, Blennow K (2009) CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* **302**, 385-393.

[42] Muñoz-Ruiz MÁ, Hall A, Mattila J, Koikkalainen J, Herukka SK, Vanninen R, Liu Y, Lötjönen J, Soininen H (2014) Comparing predictors of conversion to Alzheimer's disease using the disease state index. *Neurodegener Dis* **13**, 200-202.

[43] Vos S, van Rossum I, Burns L, Knol D, Scheltens P, Soininen H, Wahlund LO, Hampel H, Tsolaki M, Minthon L, Handels R, L'Italien G, van der Flier W, Aalten P, Teunissen C, Barkhof F, Blennow K, Wolz R, Rueckert D, Verhey F, Visser PJ (2012) Test sequence of CSF and MRI biomarkers for prediction of AD in subjects with MCI. *Neurobiol Aging* **33**, 2272-2281.

[44] Hansson O, Zetterberg H, Buchhave P, Londos E, Blennow K, Minthon L (2006) Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: A follow-up study. *Lancet Neurol* **5**, 228-234.

[45] Zetterberg H, Wahlund LO, Blennow K (2003) Cerebrospinal fluid markers for prediction of Alzheimer's disease. *Neurosci Lett* **352**, 67-69.

[46] Ashford JW, Schmitt FA (2001) Modeling the time-course of Alzheimer dementia. *Curr Psychiatry Rep* **3**, 20-28.

[47] Liu Y, Julkunen V, Paajanen T, Westman E, Wahlund LO, Aitken A, Sobow T, Mecocci P, Tsolaki M, Vellas B, Muehlboeck S, Spenger C, Lovestone S, Simmons A, Soininen H, AddNeuroMed Consortium (2012) Education increases reserve against Alzheimer's disease–evidence from structural MRI analysis. *Neuroradiology* **54**, 929-938.

[48] Raber J, Huang Y, Ashford JW (2004) ApoE genotype accounts for the vast majority of AD risk and AD pathology. *Neurobiol Aging* **25**, 641-650.

[49] Mattsson N, Andreasson U, Persson S, Arai H, Batish SD, Bernardini S, Bocchio-Chiavetto L, Blankenstein MA, Carrillo MC, Chalbot S, Coart E, Chiasserini D, Cutler N, Dahlfors G, Duller S, Fagan AM, Forlenza O, Frisoni GB, Galasko D, Galimberti D, Hampel H, Handberg A, Heneka MT, Herskovits AZ, Herukka SK, Holtzman DM, Humpel C, Hyman BT, Iqbal K, Jucker M, Kaeser SA, Kaiser E, Kapaki E, Kidd D, Klivenyi P, Knudsen CS, Kummer MP, Lui J, Lladó A, Lewczuk P, Li QX, Martins R, Masters C, McAuliffe J, Mercken M, Moghekar A, Molinuevo JL, Montine TJ, Nowatzke W, O'Brien R, Otto M, Paraskevas GP, Parnetti L, Petersen RC, Prvulovic D, de Reus HP, Rissman RA, Scarpini E, Stefani A, Soininen H, Schröder J, Shaw LM, Skinningsrud A, Skrogstad B, Spreer A, Talib L, Teunissen C, Trojanowski JQ, Tumani H, Umek RM, Van Broeck B, Vanderstichele H, Vecsei L, Verbeek MM, Windisch M, Zhang J, Zetterberg H, Blennow K (2011) The Alzheimer's Association external quality control program for cerebrospinal fluid biomarkers. *Alzheimers Dement* **7**, 386-395.

[50] Mattsson N, Zetterberg H, Blennow K (2010) Lessons from multicenter studies on CSF biomarkers for Alzheimer's disease. *Int J Alzheimers Dis* **2010**, pii: 610613.

[51] Westman E, Simmons A, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Soininen H, Weiner MW, Lovestone S, Spenger C, Wahlund LO, AddNeuroMed consortium, Alzheimer's Disease Neuroimaging, Initiative (2011) AddNeuroMed and ADNI: Similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *Neuroimage* **58**, 818-828.

[52] Stonnington CM, Tan G, Klöppel S, Chu C, Draganski B, Jack CR Jr, Chen K, Ashburner J, Frackowiak RS (2008) Interpreting scan data acquired from multiple scanners: A study with Alzheimer's disease. *Neuroimage* **39**, 1180-1185.

[53] Dubois B, Feldman HH, Jacova C, Dekosky ST, 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 PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[54] Petersen R, Aisen P, Boeve B, Geda Y, Ivnik R, Knopman D, Mielke M (2013) Mild Cognitive Impairment Due to Alzheimer Disease in the Community. *Ann Neurol* **74**, 199-208.

[55] Esposito R, Mosca A, Pieramico V, Cieri F, Cera N, Sensi SL (2013) Characterization of resting state activity in MCI individuals. *PeerJ* **1**, e135.

[56] Liu Y, Paajanen T, Westman E, Zhang Y, Wahlund LO, 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, AddNeuroMed, Consortium (2010) APOE $\varepsilon 2$ allele is associated with larger regional cortical thicknesses and volumes. *Dement Geriatr Cogn Disord* **30**, 229-237.

[57] Christa Maree Stephan B, Minett T, Pagett E, Siervo M, Brayne C, McKeith IG (2013) Diagnosing mild cognitive impairment (MCI) in clinical trials: A systematic review. *BMJ Open* **3**, pii: e001909.

[58] Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg H, Blennow K, Londos E, Hansson O (2012) Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: Six-year follow-up study. *PLoS One* **7**, e38639.

[59] Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, Nordberg A, Bäckman L, Albert M, Almkvist O, Arai H, Basun H, Blennow K, de Leon M, DeCarli C, Erkinjuntti T, Giacobini E, Graff C, Hardy J, Jack C, Jorm A, Ritchie K, van Duijn C, Visser P, Petersen RC (2004) Mild cognitive impairment – beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* **256**, 240-246.

[60] Galluzzi S, Geroldi C, Amicucci G, Bocchio-Chiavetto L, Bonetti M, Bonvicini C, Cotelli M, Ghidoni R, Paghera B, Zanetti O, Frisoni GB, Translational Outpatient Memory Clinic Working, Group (2013) Supporting evidence for using biomarkers in the diagnosis of MCI due to AD. *J Neurol* **260**, 640-650.