



Cognitive & Behavioral Assessment

**Computer-assisted prediction of clinical progression
in the earliest stages of AD**

Hanneke F. M. Rhodius-Meester^{a,*}, Hilka Liedes^b, Juha Koikkalainen^{b,c}, Steffen Wolfsgruber^d,
Nina Coll-Padros^e, Johannes Kornhuber^f, Oliver Peters^g, Frank Jessen^h, Luca Kleineidam^d,
José Luis Molinuevo^{e,i}, Lorena Rami^e, Charlotte E. Teunissen^j, Frederik Barkhof^k,
Sietske A. M. Sikkes^a, Linda M. P. Wesselman^a, Rosalinde E. R. Slot^a, Sander C. J. Verfaillie^a,
Philip Scheltens^a, Betty M. Tijms^a, Jyrki Lötjönen^c, Wiesje M. van der Flier^{a,l}

^aAlzheimer Center, Department of Neurology, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands

^bVTT Technical Research Centre of Finland Ltd., Tampere, Finland

^cCombinostics Ltd., Tampere, Finland

^dDepartment for Neurodegenerative Diseases and Geriatric Psychiatry, University of Bonn, Germany, and German Center for Neurodegenerative Diseases, Bonn, Germany

^eAlzheimer's Disease and Other Cognitive Disorders Unit, Hospital Clínic, IDIBAPS, Barcelona, Spain

^fDepartment of Psychiatry and Psychotherapy, Universitätsklinikum Erlangen and Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

^gDepartment of Psychiatry, Charité Berlin, Campus Benjamin Franklin, Berlin, Germany

^hDepartment of Psychiatry, University of Cologne, Cologne, Germany

ⁱBarcelona Beta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain

^jNeurochemistry Lab and Biobank, Department of Clinical Chemistry, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands

^kDepartment of Radiology and Nuclear Medicine, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands and Institutes of Neurology and Healthcare Engineering, UCL, London, United Kingdom

^lDepartment of Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands

Abstract

Introduction: Individuals with subjective cognitive decline (SCD) are at increased risk for clinical progression. We studied how combining different diagnostic tests can help to identify individuals who are likely to show clinical progression.

Methods: We included 674 patients with SCD (46% female, 64 ± 9 years, Mini-Mental State Examination 28 ± 2) from three memory clinic cohorts. A multivariate model based on the Disease State Index classifier incorporated the available baseline tests to predict progression to MCI or dementia over time. We developed and internally validated the model in one cohort and externally validated it in the other cohorts.

Results: After 2.9 ± 2.0 years, 151 (22%) patients showed clinical progression. Overall performance of the classifier when combining cognitive tests, magnetic resonance imaging, and cerebrospinal fluid showed a balanced accuracy of 74.0 ± 5.5 , with high negative predictive value (93.3 ± 2.8).

Discussion: We found that a combination of diagnostic tests helps to identify individuals at risk of progression. The classifier had particularly good accuracy in identifying patients who remained stable.

© 2018 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords:

Alzheimer's disease; Prognosis; Diagnostic test assessment; Clinical decision support system; Subjective cognitive decline

1. Background

In the setting of a memory clinic, patients with subjective cognitive decline (SCD) are highly relevant [1]. Most of

The authors have declared that no conflict of interest exists.

*Corresponding author. Tel.: +31204440816; Fax: +31204448529.

E-mail address: h.rhodius@vumc.nl

<https://doi.org/10.1016/j.dadm.2018.09.001>

2352-8729/© 2018 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

them are “worried well”, yet a small proportion of these patients is likely to suffer from preclinical Alzheimer’s disease (AD) [2,3]. For both the patient and the clinician, it is important to know who will progress to mild cognitive impairment (MCI) or dementia and who will remain stable [2,4,5].

At this point, cerebrospinal fluid (CSF) and magnetic resonance imaging (MRI) markers, and to a lesser extent cognitive tests, are associated with decline in SCD [3,6–13]. These findings have been translated into the “SCD plus”—criteria that have been developed to identify individuals who are more likely to harbor preclinical AD [2,14]. Translation to clinical practice is hampered because a set of recommendations for what the diagnostic workup and follow-up for patients with SCD should look like is currently lacking [15,16].

Clinical decision support systems based on modern machine-learning technologies are being developed to support clinicians to integrate multiple determinants in daily practice [17]. We have previously developed the Disease State Index (DSI) classifier, which is a technology that integrates patient data from multiple modalities to support the clinician in decision-making [18]. In previous studies, we showed that the DSI can distinguish different types of dementia and discriminate between stable and progressive MCI patients [18–21].

In this study, we aimed to investigate and validate in independent cohorts the prognostic ability of the DSI classifier to identify patients with SCD at risk for progression, by combining and visualizing all available data on baseline characteristics, neuropsychology, CSF biomarkers, and automated MRI features.

2. Methods

2.1. Patients

We included 674 patients with SCD with baseline neuropsychology available and a minimal follow-up of 1 year, from three different memory clinic-based cohorts: 354 from the Amsterdam Dementia Cohort (ADC) from the VU Medical Center [22–24], 51 from Barcelona [25], and 269 from the German Dementia Competence Network (DCN), consisting of nine memory clinics [26,27]. We used the ADC cohort to develop and internally validate our model and the pooled data of the Barcelona and DCN cohorts to externally validate our model. The study was approved by the local medical ethical committees. All patients provided written informed consent for their clinical data to be used for research purposes.

2.2. Clinical assessment

All patients went to the memory clinics seeking medical help. At baseline, they received a standardized and multidisciplinary work-up, including medical history and neuropsychological examination. CSF and MRI were

performed in a subset of patients. In multidisciplinary consensus meetings, patients were labeled as having SCD when the cognitive complaints could not be confirmed by cognitive testing using a neuropsychological battery and criteria for MCI, dementia, or other neurologic or psychiatric disorder known to cause cognitive complaints were not met.

Annual follow-up took place by routine clinical visits, in which medical and neuropsychological examinations were repeated. As outcome measure, we defined clinical progression as conversion to MCI, AD, or another type of dementia as diagnosed at follow-up. Time to follow-up was defined as time in years from baseline SCD diagnosis to progression or, if stable, time to most recent follow-up date. In the ADC and Barcelona cohort, MCI was diagnosed using Petersen’s criteria; in addition, all patients fulfilled the core clinical criteria of the NIA-AA for MCI [28,29]. In DCN, MCI patients met the Jak and Bondi criteria [30]. Patients were diagnosed with probable AD using the criteria of the NINCDS-ADRDA in all centers; all patients also met the core clinical criteria of the NIA-AA for AD dementia [31,32].

2.3. Neuropsychological tests

Cognitive functions were assessed with a standardized test battery, and we selected those tests that overlapped between the three centers. We used the Mini-Mental State Examination for global cognitive functioning [33]. For measuring executive functioning, we used Trail Making Test A (TMT-A) and Test B (TMT-B), and also for measuring language, category fluency (animals) [34,35]. For episodic memory, we included the tests that resembled each other most. In ADC, the Rey Auditory Verbal Learning Task (RAVLT) immediate and delayed recall were included [36]. In the Barcelona cohort, the Free and Cued Selective Reminding Test (FSCRT) immediate and delayed total recall were used [37]. In DCN, the Consortium to Establish a Registry for Alzheimer’s Disease word list immediate and delayed recall were used [38]. To pool the different memory tests, we standardized RAVLT, FSCRT, and Consortium to Establish a Registry for Alzheimer’s Disease per center to z-scores using group mean (details on distribution can be found in [Supplementary Fig. A1](#)). Missing data varied per test, details can be found in [Table 1](#).

2.4. MRI

In ADC, patients were scanned routinely on a 1.0 T (n = 183), 1.5 T (n = 26), or 3.0 T (n = 123) MRI scanners. Images were acquired on a 3.0 T scanner in Barcelona (n = 49) and on 1.5 T scanners in DCN (n = 93). A set of computed MRI imaging biomarkers were extracted using an image quantification tool (Combinostics Oy, Tampere, Finland, www.cneuro.com/cmri/) [19]. We included four features in the current analysis: hippocampal volume,

Table 1
Baseline characteristics according to outcome at follow-up for the separate centers

Variable	ADC			Barcelona			DCN		
	n	Stable SCD, n = 291	Progressive SCD, n = 63	n	Stable SCD, n = 46	Progressive SCD, n = 5	n	Stable SCD, n = 186	Progressive SCD, n = 83
Demographics									
Female, n (%)*	354	138 (47)	26 (41)	51	34 (74)	4 (80)	269	71 (38)	34 (41)
Age in years	354	61.2 ± 9.6	69.0 ± 7.1	51	64.9 ± 6.4	70.2 ± 8.3	269	64.5 ± 7.8	68.0 ± 8.4
Education in years	354	13.3 ± 4.3	14.0 ± 4.4	51	10.8 ± 4.2	11.6 ± 4.3	269	12.5 ± 2.8	13.3 ± 3.3
Follow-up in years	354	3.4 ± 2.2	3.8 ± 3.2	51	3.7 ± 1.8	2.8 ± 1.8	269	2.3 ± 0.9	1.6 ± 0.7
MCI/AD/non-AD, n			42/15/6			2/2/1			53/21/9
APOE status									
APOE ε4 carrier, n (%)*	317	92 (35)	27 (54)	49	10 (22)	2 (50)	226	56 (35)	32 (47)
Neuropsychology									
MMSE	351	28.4 ± 1.7	28.0 ± 1.5	51	28.3 ± 1.5	26.8 ± 1.9	265	28.2 ± 1.6	27.6 ± 1.8
Memory, immediate recall	304	41 ± 9	37 ± 8	51	42 ± 5	38 ± 6	269	20 ± 3	18 ± 4
Memory, delayed recall	303	8 ± 3	6 ± 3	51	14 ± 6	13 ± 2	269	7 ± 2	5 ± 2
TMT-A, seconds	318	40 ± 19	44 ± 14	50	44 ± 16	47 ± 18	264	42 ± 15	51 ± 20
TMT-B, seconds	318	97 ± 51	113 ± 48	50	135 ± 87	163 ± 103	264	102 ± 41	127 ± 52
Category fluency	312	22 ± 6	21 ± 5	51	21 ± 5	17 ± 4	269	21 ± 5	20 ± 5
MRI									
Hippocampal volume, mL	332	7.96 ± 0.83	7.49 ± 0.81	49	8.20 ± 0.80	7.77 ± 1.12	93	7.92 ± 0.84	7.19 ± 1.12
cMTA	332	0.37 ± 0.46	0.54 ± 0.54	49	0.22 ± 0.43	0.40 ± 0.54	93	0.54 ± 0.53	1.08 ± 0.86
cGCA	332	0.75 ± 0.65	0.87 ± 0.62	49	0.10 ± 0.24	0.22 ± 0.36	93	0.49 ± 0.64	1.17 ± 0.90
Grading	332	0.22 ± 0.19	0.36 ± 0.22	49	0.09 ± 0.12	0.23 ± 0.22	93	0.21 ± 0.23	0.44 ± 0.32
CSF									
Aβ ₄₂ , pg/mL	227	875 ± 235	638 ± 279	41	771 ± 221	637 ± 194	87	846 ± 300	670 ± 305
Total tau, pg/mL	227	266 ± 146	456 ± 370	41	333 ± 227	645 ± 694	87	286 ± 152	454 ± 281
p-tau, pg/mL	227	46 ± 18	65 ± 34	41	55 ± 28	83 ± 65	87	48 ± 20	63 ± 35

Abbreviations: SCD, subjective cognitive decline; ADC, Amsterdam Dementia Cohort; DCN, Dementia Competence Network; AD, dementia due to Alzheimer's disease; FTD, frontotemporal dementia; VaD, vascular dementia; DLB, Lewy body dementia; MMSE, Mini-Mental State Examination; RAVLT, Rey Auditory Verbal Learning Task; FSCRT, Free and Cued Selective Reminding Test; CERAD, Consortium to Establish a Registry for Alzheimer's Disease; TMT, Trail Making Test; cGCA, computed cortical atrophy score, estimated using gray matter concentration; cMTA, computed medial temporal lobe atrophy score, (left + right)/2, derived from volumes of hippocampus and lateral ventricles; Aβ₄₂, amyloid-β 1-42; p-tau, tau phosphorylated at threonine 181.

NOTE. Follow-up in years: time to conversion to MCI/dementia or follow-up time for nonconverters. Non-AD cases consisted of (1) ADC: 3 FTD and 3 VaD; (2) Barcelona: 1 DLB; and (3) DCN: 1 FTD, 1 VaD, 3 DLB, and 4 nonspecified dementia.

NOTE. Memory, immediate recall: data on immediate recall using RAVLT (ADC), FSCRT (Barcelona), and CERAD (DCN); memory, delayed recall: data on delayed recall using RAVLT (ADC), FSCRT (Barcelona), and CERAD (DCN); hippocampal volume: left plus right hippocampus (in mL), normalized for head size and gender; grading: computed using a region of interest around the hippocampus, describing the intensity similarity of test image and training set images.

NOTE. Raw data are presented as mean ± SD or n (%). Group differences per center according to outcomes were calculated using Student's t-test for continuous variables. Bold represents *P* values < .05.

*For categorical variables, the chi-square test was used.

a computed medial temporal lobe atrophy (cMTA) score, a computed global cortical atrophy (cGCA) score, and region-of-interest (ROI)-based grading. They were derived as follows: first, whole-brain segmentation into 136 structures was performed using multi-atlas segmentation method [39]. From these structures, total (left + right) hippocampal volume was used in the classification. In addition, cMTA score was derived from the volumes of the hippocampus and inferior lateral ventricles [40]. Similarly, cGCA score was estimated using voxel-based morphometry [40]. Finally, the ROI-based grading method measures the similarity of the patient image to patient images from a certain diagnostic group. In practice, an ROI from the patient is represented as a linear combination of the corresponding ROIs from a database of reference images. As each reference image contains also information about the patient's diagnostic label, the grading feature is defined as the share of the weights from images with a certain diagnostic label. In this work, we

used an ROI centered around the hippocampus [41]. See [Supplementary Fig. A2](#), for a schematic presentation of this method. For classification, the volume of the hippocampus was normalized first for head size [42] and then for gender using the LMS method (referring to smooth curve [L], mean [M], and coefficient of variation [S]) [43]. The grading feature was also normalized for gender.

2.5. CSF

CSF samples from both ADC (n = 227) and Barcelona (n = 41) were analyzed at the Neurochemistry Laboratory at the Department of Clinical Chemistry of the VUmc, the Netherlands. In DCN (n = 87), samples were analyzed at the laboratory of the University of Erlangen, Germany. All centers measured amyloid-β 1-42 (Aβ₄₂), total tau, and tau phosphorylated at threonine 181 (p-tau) with commercially available ELISAs (Innotest; Fujirebio, Ghent, Belgium).

2.6. APOE genotyping

In ADC (n = 317), the apolipoprotein E (APOE) genotype was determined with the LightCycler APOE mutation detection method (Roche diagnostics GmbH, Mannheim, Germany). In Barcelona (n = 49), the APOE genotype was determined with PCR amplification and Sanger sequencing (ThermoFisher, USA). In DCN (n = 226), leukocyte DNA was isolated with the Qiagen Isolation Kit (Qiagen, Hilden, Germany). Patients were dichotomized into APOE ϵ 4 carriers (heterozygous and homozygous) and noncarriers.

2.7. Disease State Index

For classifying patients at risk of progression or not, we used a modification of the PredictND tool that was previously developed in the European FP7 project PredictND (www.predictnd.eu). The tool is based on the DSI classifier [17]. When presented with a new patient, the DSI estimates the similarity of measurement values from this patient to observed values from reference patients with and without a certain medical condition, in this case similarity to patients with stable SCD and patients progressing to MCI or dementia [17]. Similarity is estimated in the following way: (1) Each measurement value of an individual person is compared with the reference data using a fitness function defined as $f(x) = FN(x)/(FN(x)+FP(x))$, where FN is the false-negative error rate, and FP is the false-positive error rate in the reference data when using the individual's measurement value x as a cutoff value in classification. (2) The "relevance" of each determinant is defined as sensitivity + specificity - 1. (3) Finally, a composite DSI is defined as a weighted average of fitness values: $DSI = \Sigma (\text{relevance} \cdot \text{fitness})/\Sigma \text{relevance}$. DSI is a continuous value between zero and one, reflecting how similar an individual is to patients who have previously progressed. A cutoff value of 0.5 is used to classify whether an individual patient is more likely to remain stable ($DSI < 0.5$) or progress to MCI or dementia ($DSI \geq 0.5$) at follow-up. In addition, we studied whether the performance is improved for a subset of patients with high ($DSI > 0.7$ and $DSI > 0.8$) or low ($DSI < 0.3$ and $DSI < 0.2$) DSI values. This could enable detecting patients with very low risk or very high risk of progression for clinical counseling. The classifier also provides a visual representation of how different features contribute to the DSI in a so-called disease state fingerprint (see Fig. 2, for details). As the DSI combines multiple independent classifiers (fitness functions), there is no need to impute data or exclude cases with incomplete data. More mathematical details can be found in the study by Mattila et al. [17].

2.7.1. Development and internal validation

We developed the model on the ADC data and internally validated this model on the same cohort using 10 iterations

of three-fold cross-validation. We assessed the different data sources separately (demographics, APOE status, neuropsychology tests, CSF biomarkers, and computed MRI imaging markers) and then combined them, independent of missing data. Owing to the technical differences between scanners, we excluded MRI features from patients scanned with 1.0 T devices (n = 183) from the training set and tested using all field strength and only >1.0 T. In this way, the classifier is able to better learn the differences between diagnostic groups without the excess variation from the scanner differences (for details, see [Supplementary Table A1](#)). We used the following performance metrics in the evaluation of the DSI: the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). Although DSI balances the results by default, we also estimated balanced accuracy that is typically defined as mean of sensitivity and specificity. As an outcome measure, we defined progression to MCI or dementia, and we also repeated the analyses including only progression to MCI or dementia due to AD (excluding other dementias).

2.7.2. External validation

For external validation, we tested our developed model on new, unseen cases from pooled Barcelona and DCN data. To understand why the performance decreased with the independent validation cohort, we repeated the analyses by training the model and performing cross-validation with the Barcelona and DCN data, and using the ADC data as a separate validation cohort.

2.7.3. Comparison to other machine-learning algorithms

Earlier studies have performed comprehensive comparisons between the DSI classifier and other machine-learning algorithms [17,44,45]. We add on to this by comparing the classifier to Naïve Bayes and Random Forest classifiers. Details can be found in [Appendix](#).

2.8. Other statistical analyses

We investigated differences in baseline characteristics in each center according to outcome, using Student's t-test and the chi-square test when appropriate, using SPSS, version 22 (IBM, Armonk, NY, USA). $P < .05$ was considered significant. The DSI analysis was performed using MATLAB toolbox in MATLAB, version R2015b (MathWorks, Natick, MA, USA) [46].

3. Results

3.1. Baseline characteristics

After a mean of 2.9 ± 2.0 years, 151 (22%) patients showed clinical progression to MCI or any type of dementia (Table 1). Patients who showed progression were older, more frequent APOE ϵ 4 carriers, performed somewhat worse on

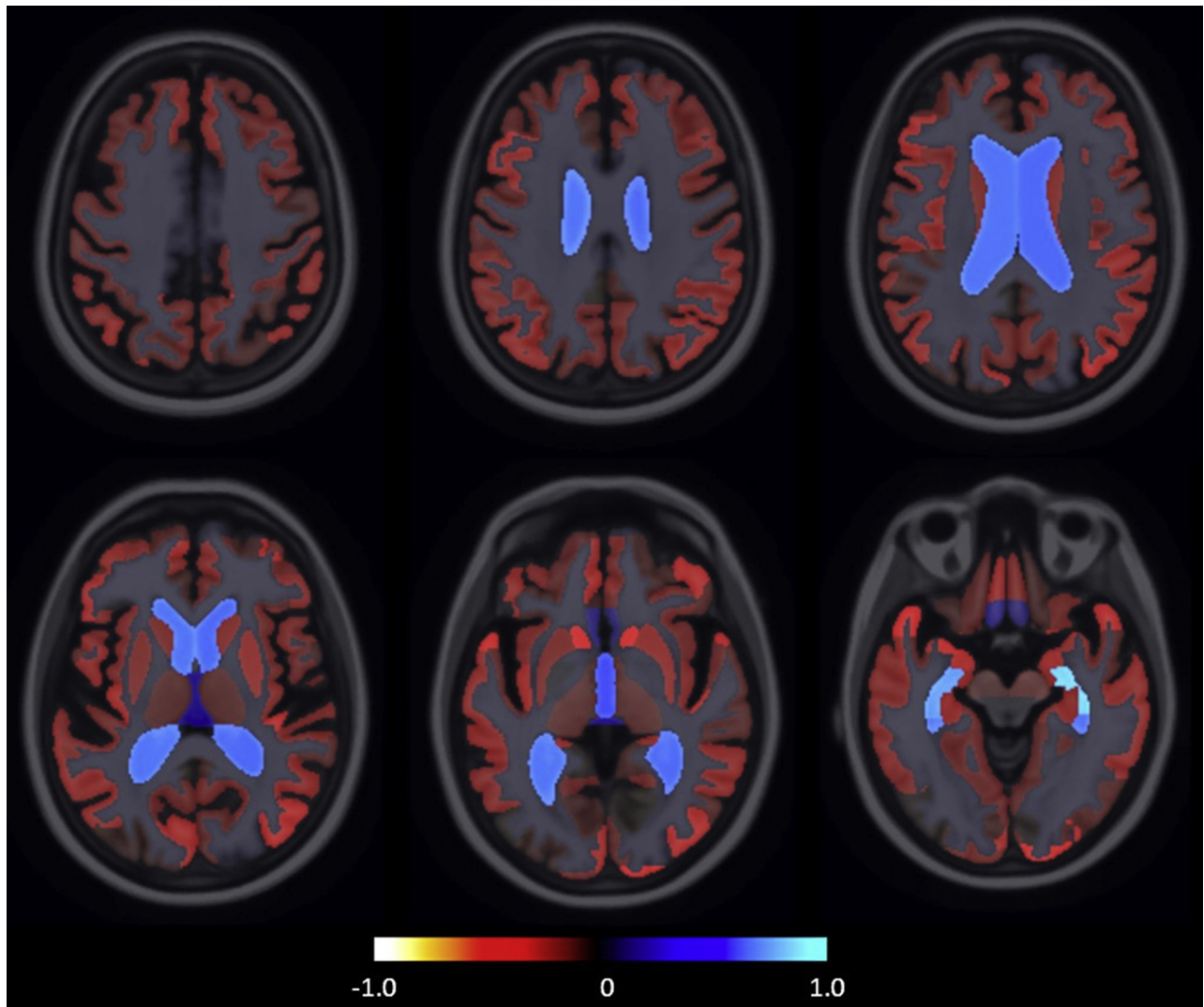


Fig. 1. The visualization of group-wise volume differences between stable subjective cognitive decline (SCD) and progressive SCD groups. The map visualizes the relative volume difference: $\frac{V_p - V_s}{0.5 \times (V_p + V_s)}$, where V_p and V_s are the mean volumes for progressive and stable groups, respectively. Blue indicates the structures on MRI that were larger in the progressive group, and red indicates the structures that were smaller.

neuropsychological tests, and had smaller hippocampal volumes and more abnormal CSF biomarkers. Patients in ADC were younger as compared with Barcelona and DCN. Patients in Barcelona were more often female, had less education, and showed less progression as compared with ADC and DCN. Duration of follow-up was longest in Barcelona and shortest in DCN. There were no differences in percentage of APOE $\epsilon 4$ carriers and baseline Mini-Mental State Examination across the centers.

3.2. Development and internal validation of the model

Table 2 shows performance of the DSI using the different data sources, for progression to MCI or dementia. As single data source, CSF showed highest balanced accuracy, followed by the automatic MRI features. Fig. 1 shows the group-wise volume differences between stable SCD and

progressive SCD groups, with the clearest differences observed in the medial temporal region.

When we used all the data sources together, performance improved (balanced accuracy: $74.0 \pm 5.5\%$). The model had high NPV (93.3 ± 2.8), whereas PPV was only modest (37.7 ± 5.5). This indicates that the DSI classifier was most useful to identify patients who remained stable. When we repeated the analyses for progression to MCI or dementia due to AD (excluding other dementias) as an outcome measure, results were comparable (Supplementary Table A2).

Table 2 also presents performance of the DSI classifier for subgroups having high or low DSI values, to aid the clinician on how to interpret the DSI values. We observed extreme DSI values, that is, below 0.3 or above 0.7, in $48 \pm 6\%$ of the patients. When $DSI < 0.3$, NPV was 97.0 ± 2.6 , indicating that the probability of progression is very low in

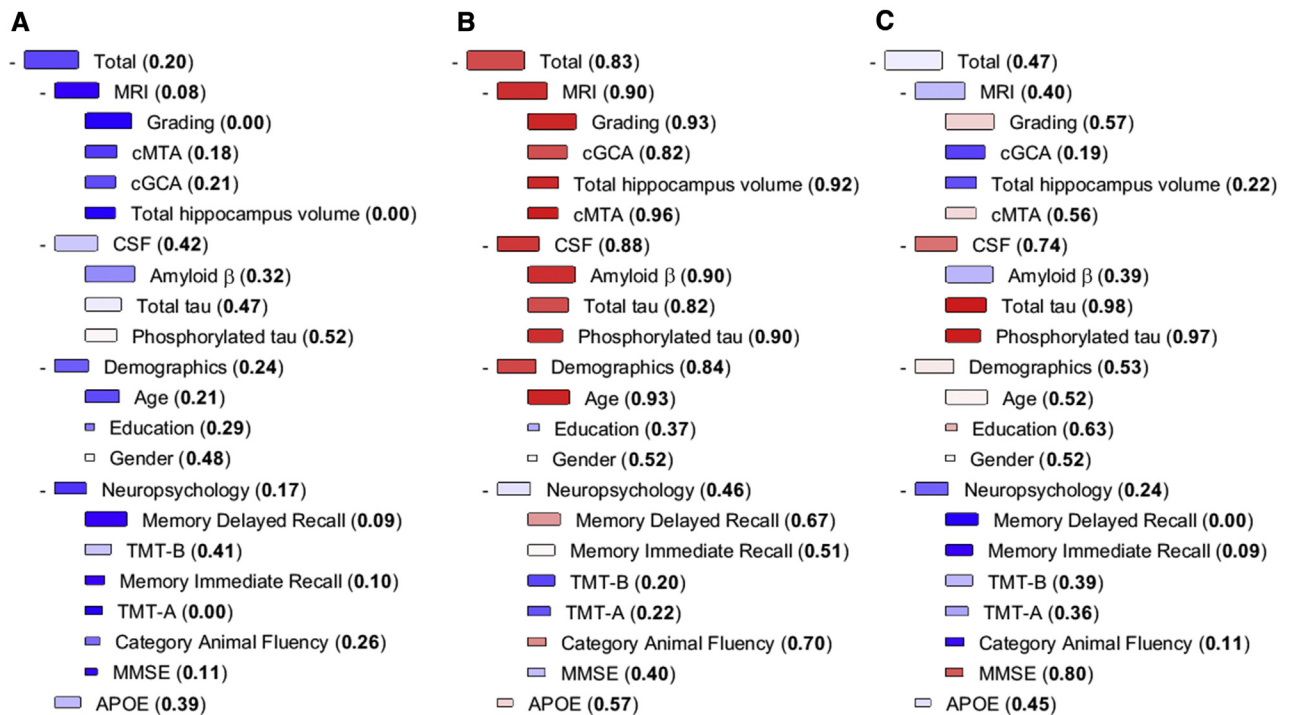


Fig. 2. Examples of DSI fingerprints: patient A and patient C remained stable, and patient B progressed to MCI. The DSI fingerprint combines all data available from one patient and displays it in a visually attractive format to the clinician. The DSI value is presented both numerically and visually with color. The color changes from blue to red when DSI increases from zero (high similarity to the stable group) to one (high similarity to the progressive group). The relevance is visualized by the size of the box. The larger the box, the better the specific marker discriminates the stable and progressive SCD patients. Abbreviations: MMSE, Mini-Mental State Examination; TMT, Trail Making Test; cGCA: computed cortical atrophy score, estimated using gray matter concentration; cMTA, computed medial temporal lobe atrophy score, (left + right)/2, derived from volumes of hippocampus and lateral ventricles; Amyloid β, amyloid-β 1–42; Phosphorylated tau, tau phosphorylated at threonine 181; DSI, Disease State Index.

this subset and the clinician could reassure these patients with high confidence. For comparison, if NPV is computed for all patients without using any prediction model, it is 82.0 [291/(291 + 63)], showing that DSI can clearly help in stratifying patients. When DSI > 0.7, PPV was not very high, only 50.8 ± 12.9. Although the progression of an individual cannot be predicted accurately even in this subgroup, the risk of conversion is clearly elevated. The risk ratio is 2.8 in this subgroup compared with the whole patient population meaning that the clinician might start applying more rigorous follow-up and lifestyle intervention measures to these patients. This means that for roughly half of SCD patients, the DSI could have practical use to aid in individualized prognosis.

Fig. 2 shows the DSI fingerprints for three example patients to illustrate how the tool integrates and visualizes available data. Patient A is a 60-year-old female, with a DSI of 0.20, meaning the clinician can reassure her with high accuracy. Nearly all the boxes in the fingerprint are blue, which fits with the good outcome in this patient; she remained stable during three years of follow-up. Patient B is a 74-year-old female with a DSI value of 0.83, mainly attributable to her values on MRI and CSF (visible as red boxes). This implies her risk of progression is clearly elevated, and follow-up should be discussed. This patient progressed to MCI after 3 years. Patient C is a 66-year-old

female, who remained stable during a follow-up period of 4 years. The fingerprint shows both red and blue boxes, implying that interpretation is inconclusive and a reliable prognosis cannot be made, further illustrated by an inconclusive DSI value of 0.47.

3.3. External validation

When we externally validated our model by testing it in pooled data of Barcelona and DCN, we found an overall lower performance (balanced accuracy 65.1, NPV 83.7; Table 3). Balanced accuracy increased to 78.5 in the more extreme DSI values. To evaluate what caused the lower performance on external validation, we also trained the model on pooled data of Barcelona and DCN and tested it in ADC data (Supplementary Table A3). Even when we developed the model in Barcelona and DCN cohorts, performance was still better in ADC (balanced accuracy 73.3, NPV 92.4).

3.4. Comparison to other machine-learning algorithms

For comparison, other machine-learning algorithms were also tested. The performance of the Naïve Bayes classifier was corresponding to and the Random Forest classifier lower than what was obtained by the DSI classifier (Supplementary Table A4).

Table 2

Performance of DSI to predict conversion to MCI or any type of dementia in the ADC cohort, for the total cohort and for patients with extreme DSI values

Variable	%	Stable SCD, n	Progressive SCD, n	AUC	Balanced accuracy	Sensitivity	Specificity	PPV	NPV
Demographics		291	63	0.74 ± 0.04	66.0 ± 5.0	66.0 ± 11.7	65.9 ± 6.1	29.7 ± 3.8	90.1 ± 2.8
APOE		267	50	0.60 ± 0.05	59.7 ± 4.9	53.9 ± 8.4	65.5 ± 4.5	22.7 ± 4.4	88.4 ± 2.4
Neuropsychology		290	62	0.69 ± 0.06	62.7 ± 4.4	61.6 ± 10.8	64.3 ± 5.3	26.9 ± 3.3	88.7 ± 2.4
CSF		194	33	0.77 ± 0.07	69.9 ± 5.0	66.1 ± 11.3	73.6 ± 6.5	30.3 ± 5.8	92.8 ± 2.6
MRI (1 T, 1.5 T, 3 T)		277	55	0.68 ± 0.05	61.4 ± 4.3	80.1 ± 10.1	42.8 ± 6.8	21.8 ± 2.5	91.9 ± 3.5
MRI (>1 T)		123	25	0.73 ± 0.09	69.1 ± 7.8	64.9 ± 15.2	73.3 ± 7.6	33.6 ± 9.6	91.3 ± 4.0
Demographics + APOE + Neuropsychology + CSF + MRI (1 T, 1.5 T, 3 T)		291	63	0.80 ± 0.05	74.0 ± 4.2	82.9 ± 8.4	65.1 ± 5.8	34.2 ± 3.8	94.7 ± 2.4
Demographics + APOE + Neuropsychology + CSF + MRI (>1 T)		291	63	0.81 ± 0.06	74.1 ± 5.8	75.7 ± 11.2	72.6 ± 4.8	37.7 ± 5.5	93.3 ± 2.8
DSI < 0.2 or DSI > 0.8									
Demographics + APOE + Neuropsychology + CSF + MRI (1 T, 1.5 T, 3 T)	14 ± 4	12 ± 5	5 ± 2	0.81 ± 0.10	83.3 ± 7.4	98.9 ± 4.2	67.7 ± 13.6	59.0 ± 17.4	99.4 ± 2.3
Demographics + APOE + Neuropsychology + CSF + MRI (>1 T)	21 ± 6	20 ± 8	5 ± 2	0.83 ± 0.11	84.1 ± 9.6	85.4 ± 17.6	82.8 ± 7.1	56.2 ± 17.1	96.3 ± 4.6
DSI < 0.3 or DSI > 0.7									
Demographics + APOE + Neuropsychology + CSF + MRI (1 T, 1.5 T, 3 T)	37 ± 6	34 ± 7	10 ± 3	0.84 ± 0.06	80.7 ± 6.0	89.6 ± 11.4	71.8 ± 9.4	47.8 ± 10.9	96.8 ± 3.0
Demographics + APOE + Neuropsychology + CSF + MRI (>1 T)	48 ± 6	47 ± 7	9 ± 3	0.84 ± 0.09	84.1 ± 7.3	84.9 ± 14.2	83.3 ± 5.5	50.8 ± 12.9	97.0 ± 2.6

Abbreviations: AUC, area under the receiver operating characteristic curve; SCD, subjective cognitive decline; PPV, positive predictive value; NPV, negative predictive value; APOE, apolipoprotein E; DSI, Disease State Index.

NOTE. For the extreme DSI values, n: number of patients in a cross-validation fold having the DSI value in the given range; %: percentage of patients in a test set (n = 118) of a cross-validation fold having the DSI value in the given range. Values are presented as mean ± standard deviation over 10 iterations of three-fold cross-validation.

Table 3

External validation: Performance of DSI to predict conversion to MCI or any type of dementia when tested in the pooled data of Barcelona and DCN cohorts, for the total cohort and for patients with extreme values

Variable	%	Stable SCD, n	Progressive SCD, n	AUC	Balanced accuracy	Sensitivity	Specificity	PPV	NPV
Demographics		232	88	0.63	57.8	61.4	54.3	33.8	78.8
APOE		203	72	0.57	57.4	47.2	67.5	34.0	78.3
Neuropsychology		232	88	0.69	63.9	63.6	64.2	40.3	82.3
CSF		90	39	0.69	61.7	59.0	64.4	41.8	78.4
MRI		100	42	0.77	67.4	73.8	61.0	44.3	84.7
Demographics + APOE + Neuropsychology + CSF + MRI		232	88	0.72	65.1	68.2	62.1	40.5	83.7
DSI < 0.2 or DSI > 0.8									
Demographics + APOE + Neuropsychology + CSF + MRI	21	38	30	0.81	78.5	83.3	73.7	71.4	84.8
DSI < 0.3 or DSI > 0.7									
Demographics + APOE + Neuropsychology + CSF + MRI	45	94	50	0.79	74.2	76.0	72.3	59.4	85.0

Abbreviations: AUC: area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; APOE, apolipoprotein E; DSI, Disease State Index; SCD, subjective cognitive decline.

NOTE. For the extreme DSI values, n: number of patients having the DSI value in the given range; %: percentage of patients having the DSI value in the given range. Values are presented as mean.

4. Discussion

In this large memory clinic study, we found that after an average follow-up of almost 3 years, 22% of the individuals with SCD showed clinical progression to MCI or dementia. The DSI classifier combining cognitive test results, automated MRI features, and CSF biomarkers accurately classified 74% of the patients, with especially high NPV. Nearly half of the patients had a clearly positive or negative DSI of <0.3 or >0.7 , where balanced accuracy was as high as 84%.

Although many individuals with SCD may indeed be “worried well,” a minority visits the memory clinic because they actually experience cognitive decline, which the clinician is not yet able to verify. We show that a computer-aided decision tool could support clinicians in identifying that minority of individuals who are at high risk of clinical progression. Moreover, for a larger group of individuals, reassurance can be even more explicit, backed up by negative findings on a combination of diagnostic tests. For daily clinical routine, this could imply a paradigm shift; it is current practice to reassure patients with SCD but not disclose results of their particular diagnostic tests. Our results provide support for the notion, however, that we approach an era of personalized medicine, where individuals’ results on diagnostic tests can be used to obtain individualized predictions. Our classifier may aid in providing prognosis or decide to follow up individuals at increased risk for progression. On further scrutinizing the data, we observed that performance was particularly good for roughly half of the population, with a high or low DSI (<0.3 or >0.7), while prognostic performance was suboptimal for those with a medium DSI (0.4-0.6) (data not shown). Yet, overall NPV was very high, reaching up to 97.0 for the cases with $DSI < 0.3$. This implies that patients with a $DSI < 0.3$ can be reassured and do not need follow-up. For patients with a $DSI > 0.7$, a certain prognosis cannot be made, but the risk of clinical progression is clearly elevated and follow-up is warranted. The fingerprint could further aid in this interpretation by visualizing how each of the determinants contributed to the prognosis. Of note, in the present study, we focused on patients who present to a memory clinic with the clinical question whether they have an underlying neurodegenerative disease. In further work, tools like this could also be used for screening patients at risk in the general population, for example, by using blood-based biomarkers [47].

The overall balanced accuracy of the DSI was highest when we combined all different data sources. The discriminative effect of MRI and CSF biomarkers are in line with the additive model, indicating patients with SCD at risk of progression already have more AD-like biomarkers at baseline [48]. Also, neuropsychological assessment at baseline improved the performance of the DSI. It is conceivable that even within normal boundaries, a slight

decline in cognitive performance is associated with progression, which is particularly appreciated when analyzed together with data from other sources. The classifier also provided fully automatically computed MRI features enabling the clinician to extract more information from the images than when using visual interpretation only [19].

The strength of this study was the large size of the cohort in which the model was developed, and the availability of two independent cohorts for external validation. All patients underwent thorough examination and were only included if cognitive complaints could not be confirmed by cognitive testing. We used data that were typical of memory clinics, varied and incomplete. Because we aimed to develop a tool that should be able to support clinicians in daily practice, it is essential the tool can deal with missing data.

However, several potential limitations also need to be discussed. In general, when developing prediction models based on classifiers, comparing training and testing results can be challenging for several reasons. In this study we trained the tool on the ADC data and found that on validation in the Barcelona and DCN data, performance was less optimal. This might indicate that generalizability is limited. When we trained the tool in the Barcelona and DCN data and then performed external validation in ADC, we still found that performance was better in ADC than in the Barcelona and DCN data. This suggests that not the model itself hampers generalizability, but the lower performance is caused by heterogeneity in cohorts. Overall, the following sources can affect generalizability of prediction models: (1) patients in different memory clinics are different (i.e., both referral and definition of SCD), (2) heterogeneity in outcome, (3) patient measurements are done in different ways, and (4) prediction models are not able to generalize. In the field of SCD, heterogeneity between cohorts is an important hurdle [2,5,49]. The field is acknowledging this and working toward more harmonization of research efforts. Nonetheless, it is of the utmost importance to actually perform studies on multiple data sets, both to get to know the differences and how this influences results, and to start harmonizing and bridging data. In this study, we feel there are several important cohort differences: first, patients showed substantial baseline differences between the three memory clinic cohorts. We found differences regarding progression rates and definition of progression; ADC and Barcelona used the Petersen criteria for MCI, whereas DCN used the Jak-Bondi criteria for MCI [28,30]. Also those who remained stable in Barcelona and DCN were older than those in ADC. Second, although follow-up duration in VUmc was longer, more patients showed progression in Barcelona + DCN. Third, although all patients underwent a harmonized work-up, the work-up differed between the centers. We tried to eliminate these differences as much as possible. For the neuropsychology tests, we selected tests that overlapped or resembled each

other. Also, CSF analyses of ADC, DCN, and Barcelona were performed in the two laboratories, as part of the Euro-SCD collaboration, minimizing, but not excluding, interlaboratory variability. MRI scans were acquired on systems with different field strengths, yet the automatic analyses of these scans were all performed by the same software [19]. However, 1.0 T images have worse gray matter–white matter contrast than 1.5 T and 3.0 T images. Consequently, we decided to use only 1.5 T and 3.0 T images in training to have a robust classifier and then reported the results separately for different field strengths to demonstrate the differences between 1.0 T and >1.0 T images with roughly similar performance. In this study, we did not perform feature selection and choose a set of features maximizing prediction accuracy. We included diagnostics tests and features that are either familiar to clinicians or which we found to be good features in other studies. Had we used an optimal set of features, this would probably increase the performance of our model, at the risk of overfitting.

In conclusion, this study shows that it is feasible to extract and combine information from routine diagnostic tests into a measure that can be used within a clinical decision support system, supporting clinicians to identify individuals at risk of progression who need follow-up and individuals who are likely to remain stable and can be reassured and discharged. This implies that it is possible to think about a personalized medicine approach, also in patients with SCD. Recent research has shown that patients would like to be actively involved in decisions about prognostic testing, but they feel they often lack important information on the implication of the tests [15,50]. Tools such as the DSI classifier can provide a first step in taking personalized medicine in SCD to a next level.

Acknowledgments

Research of the VUmc Alzheimer Center is part of the neurodegeneration research program of the Amsterdam Neuroscience. The VUmc Alzheimer Center is supported by Alzheimer Nederland and Stichting VUmc Fonds. The clinical database structure was developed with funding from Stichting Dioraphte. For the development of the PredictAD tool, the VTT Technical Research Centre of Finland has received funding from European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreements 601055 (VPH-DARE@IT), 224328 (PredictAD), and 611005 (PredictND). The Euro-SCD project has been funded by the EU Joint Program–Neurodegenerative Disease Research (JPND_PS_FP-689-019). DCN has been funded by a grant from the German Federal Ministry of Education and Research (BMBF): Kompetenznetz Demenzen (01GI0420). Hanneke FM Rhodius-Meester is appointed on PredictND, a grant from the European Seventh Framework Program project PredictND under grant agreement 611005.

Frederik Barkhof is supported by the NIHR UCLH Biomedical Research Center. Sietske AM Sikkes is supported by an Off Road grant (ZonMw #451001010). Wiesje M. van der Flier is a recipient of a research grant from Gieskes-Strijbis Fonds. Betty M. Tijms receives grant support from ZonMw (#73305056 and #733050824).

Author disclosures: Hanneke F.M. Rhodius-Meester, Hilkka Lieder, Steffen Wolfsgruber, Nina Coll-Padros, Johannes Kornhuber, Luca Kleinedam, Lorena Rami, Sietske A. Sikkes, Linda MP Wesselman, Rosalinde E.R. Slot, Sander C.J. Verfaillie, and Betty Tijms report no disclosures. Juha Koikkalainen and Jyrki Lötjönen all report that the VTT Technical Research Center of Finland owns the following IPR related to the article: (1) J. Koikkalainen and J. Lotjonen—a method for inferring the state of a system, US7,840,510 B2, PCT/FI2007/050277; and (2) J. Lotjonen, J. Koikkalainen, and J. Mattila—state inference in a heterogeneous system, PCT/FI2010/050545, FI20125177. Juha Koikkalainen and Jyrki Lötjönen are shareholders in Combinostics Oy. Oliver Peters has received speaker honoraria from Eli Lilly & Company, Novartis, Affiris, and Roche and has received research support from Axon, Axovant, Biogen, Eli Lilly and Company, Lundbeck, Pharmatrophix, Probiobrug, Novartis, Roche, Janssen, Piramal, Takeda, and TRX Pharmaceuticals. Frank Jessen has received fees for advisory boards of Eli Lilly, Biogene, MSD, Janssen Cilag, Roche, AC Immune, and Novartis. José Luis Molinuevo is the PI of trials funded by Roche, Merck, Novartis, and Janssen and has received speaker or consultant fees from Roche, Roche diagnostics, Biogen, Merck, Novartis, Oryzon, IBL, Axovant, Lundbeck, and Lilly. Charlotte E. Teunissen is a member of the Innogenetics International Advisory Boards of Fujirebio/Innogenetics and Roche. Frederik Barkhof serves/has served on the advisory boards of Bayer-Schering Pharma, Sanofi-Aventis, Biogen-Idec, TEVA, Merck-Serono, Novartis, Roche, Synthon BV, Jansen Research, and Genzyme. He received funding from the Dutch MS Society and EU-FP7 and has been a speaker at symposia organized by the Serono Symposia Foundation and Medscape. Philip Scheltens has served as a consultant for Wyeth-Elan, Genentech, Danone, and Novartis and received funding for travel from Pfizer, Elan, Janssen, and Danone Research. Wiesje M. van der Flier performs contract research for Biogen. Research programs of Wiesje M. van der Flier have been funded by ZonMw, NWO, EU-FP7, Alzheimer Nederland, CardioVascular Onderzoek Nederland, Stichting Dioraphte, Gieskes-Strijbis Fonds, Boehringer Ingelheim, Piramal Neuroimaging, Combinostics, Roche BV, and Janssen Stellar. She has been an invited speaker at Boehringer Ingelheim and Biogen. All funding is paid to her institution.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dadm.2018.09.001>.

RESEARCH IN CONTEXT

1. Systematic review: An increasing number of studies focus on biomarkers that can help identifying patients with subjective cognitive decline (SCD) at risk of progression. Translation to clinical practice is hampered because it remains unclear what the diagnostic workup and follow-up for SCD should look like and what results should be disclosed in daily practice. We cited relevant citations.
2. Interpretation: We used a clinical decision support system to identify patients with SCD at risk for progression. Clinical decision support systems can weigh and combine different diagnostic tests; this multivariate model showed especially a high negative predictive value, meaning the classifier identified patients who will remain stable and can thus be reassured.
3. Future directions: Clinical decision support systems could be useful to aid clinicians in interpreting diagnostic test results and discuss results of these tests with patients with SCD. To take diagnosis and prognosis in SCD to the next level, further knowledge on shared decision-making in SCD is needed.

References

- [1] Buckley RF, Villemagne VL, Masters CL, Ellis KA, Rowe CC, Johnson K, et al. A Conceptualization of the utility of subjective cognitive decline in clinical trials of preclinical Alzheimer's disease. *J Mol Neurosci* 2016;60:354–61.
- [2] Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chetelat G, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement* 2014;10:844–52.
- [3] van Harten AC, Visser PJ, Pijnenburg YA, Teunissen CE, Blankenstein MA, Scheltens P, et al. Cerebrospinal fluid Aβ42 is the best predictor of clinical progression in patients with subjective complaints. *Alzheimers Dement* 2013;9:481–7.
- [4] Cavado E, Lista S, Khachaturian Z, Aisen P, Amouyel P, Herholz K, et al. The Road Ahead to Cure Alzheimer's Disease: Development of Biological Markers and Neuroimaging Methods for Prevention Trials Across all Stages and Target Populations. *J Prev Alzheimers Dis* 2014;1:181–202.
- [5] Molinuevo JL, Rabin LA, Amariglio R, Buckley R, Dubois B, Ellis KA, et al. Implementation of subjective cognitive decline criteria in research studies. *Alzheimers Dement* 2017;13:296–311.
- [6] Fonseca JA, Ducksbury R, Rodda J, Whitfield T, Nagaraj C, Suresh K, et al. Factors that predict cognitive decline in patients with subjective cognitive impairment. *Int Psychogeriatr* 2015;27:1671–7.
- [7] Toledo JB, Bjerke M, Chen K, Rozycki M, Jack CR Jr, Weiner MW, et al. Memory, executive, and multidomain subtle cognitive impairment: clinical and biomarker findings. *Neurology* 2015;85:144–53.
- [8] Hessen E, Nordlund A, Stalhammar J, Eckerstrom M, Bjerke M, Eckerstrom C, et al. T-Tau is associated with objective memory decline over two years in persons seeking help for subjective cognitive decline: A report from the Gothenburg-Oslo MCI Study. *J Alzheimers Dis* 2015;47:619–28.
- [9] Toledo JB, Weiner MW, Wolk DA, Da X, Chen K, Arnold SE, et al. Neuronal injury biomarkers and prognosis in ADNI subjects with normal cognition. *Acta Neuropathol Commun* 2014;2:26.
- [10] Meibeth D, Scheef L, Wolfsgruber S, Boecker H, Block W, Traber F, et al. Cortical thinning in individuals with subjective memory impairment. *J Alzheimers Dis* 2015;45:139–46.
- [11] van der Flier WM, van Buchem MA, Weverling-Rijnsburger AW, Mutsaers ER, Bollen EL, Admiraal-Behloul F, et al. Memory complaints in patients with normal cognition are associated with smaller hippocampal volumes. *J Neurol* 2004;251:671–5.
- [12] Verfaillie SC, Tijms B, Versteeg A, Benedictus MR, Bouwman FH, Scheltens P, et al. Thinner temporal and parietal cortex is related to incident clinical progression to dementia in patients with subjective cognitive decline. *Alzheimers Dement (Amst)* 2016;5:43–52.
- [13] Peter J, Scheef L, Abdulkadir A, Boecker H, Heneka M, Wagner M, et al. Gray matter atrophy pattern in elderly with subjective memory impairment. *Alzheimers Dement* 2014;10:99–108.
- [14] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:280–92.
- [15] Van der Flier W, Kunneman M, Bouwman FH, Petersen R, Smets EMA. Diagnostic dilemmas in Alzheimer's disease: room for shared decision making. *Alzheimers Dement* 2017;3:301–4.
- [16] Frisoni GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *Lancet Neurol* 2017;16:661–76.
- [17] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van GM, Waldemar G, et al. A disease state fingerprint for evaluation of Alzheimer's disease. *J Alzheimers Dis* 2011;27:163–76.
- [18] Mattila J, Soininen H, Koikkalainen J, Rueckert D, Wolz R, Waldemar G, et al. Optimizing the diagnosis of early Alzheimer's disease in mild cognitive impairment subjects. *J Alzheimers Dis* 2012;32:969–79.
- [19] Koikkalainen J, Rhodius-Meester H, Tolonen A, Barkhof F, Tijms B, Lemstra AW, et al. Differential diagnosis of neurodegenerative diseases using structural MRI data. *Neuroimage Clin* 2016;11:435–49.
- [20] Rhodius-Meester HF, Koikkalainen J, Mattila J, Teunissen CE, Barkhof F, Lemstra AW, et al. Integrating biomarkers for underlying Alzheimer's Disease in mild cognitive impairment in daily practice: Comparison of a clinical decision support system with individual biomarkers. *J Alzheimers Dis* 2015;50:261–70.
- [21] Munoz-Ruiz MA, Hartikainen P, Hall A, Mattila J, Koikkalainen J, Herukka SK, et al. Disease state fingerprint in frontotemporal degeneration with reference to Alzheimer's disease and mild cognitive impairment. *J Alzheimers Dis* 2013;35:727–39.
- [22] van der Flier WM, Pijnenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. *J Alzheimers Dis* 2014;41:313–27.
- [23] Slot RER, Verfaillie SCJ, Overbeek JM, Timmers T, Wesselman LMP, Teunissen CE, et al. Subjective Cognitive Impairment Cohort (SCIENCE): study design and first results. *Alzheimer Res Ther* 2018. In press.
- [24] van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. *J Alzheimers Dis* 2018;62:1091–111.
- [25] Valech N, Mollica MA, Olives J, Tort A, Fortea J, Lleó A, et al. Informants' perception of subjective cognitive decline helps to discriminate preclinical Alzheimer's disease from normal aging. *J Alzheimers Dis* 2015;48:S87–98.

- [26] Wolfsgruber S, Polcher A, Koppara A, Kleineidam L, Frolich L, Peters O, et al. Cerebrospinal fluid biomarkers and clinical progression in patients with subjective cognitive decline and mild cognitive impairment. *J Alzheimers Dis* 2017;58:939–50.
- [27] Kornhuber J, Schmidtke K, Frolich L, Pernecky R, Wolf S, Hampel H, et al. Early and differential diagnosis of dementia and mild cognitive impairment: design and cohort baseline characteristics of the German Dementia Competence Network. *Dement Geriatr Cogn Disord* 2009;27:404–17.
- [28] Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004;256:183–94.
- [29] Albert MS, Dekosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. 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* 2011;7:270–9.
- [30] Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis* 2014;42:275–89.
- [31] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. 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* 2011;7:263–9.
- [32] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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* 1984;34:939–44.
- [33] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–98.
- [34] Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Normative data for the animal, profession and letter M naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. *J Int Neuropsychol Soc* 2006;12:80–9.
- [35] Reitan R. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills* 1958;8:271–6.
- [36] Saan RD, BG. De 15-Woorden Test A en B. Een voorlopige handleiding. (in Dutch). Groningen: Afdeling Neuropsychologie, AZG; 1986.
- [37] Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. *Neurology* 1988;38:900–3.
- [38] Berres M, Monsch AU, Bernasconi F, Thalman B, Stahelin HB. Normal ranges of neuropsychological tests for the diagnosis of Alzheimer's disease. *Stud Health Technol Inform* 2000;77:195–9.
- [39] Lotjonen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soinen H, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 2010;49:2352–65.
- [40] Lotjonen J, Koikkalainen J, Rhodius-Meester HFM, van der Flier WM, Scheltens P, Barkhof F, et al. Computed rating scales for cognitive disorders from MRI. *Alzheimers Dement* 2017;13:1108.
- [41] Tong T, Wolz R, Coupe P, Hajnal JV, Rueckert D. Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *Neuroimage* 2013;76:11–23.
- [42] Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* 2004;23:724–38.
- [43] Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.
- [44] Mattila J, Koikkalainen J, Virkki A, van GM, Lotjonen J. Design and application of a generic clinical decision support system for multiscale data. *IEEE Trans Biomed Eng* 2012;59:234–40.
- [45] Tolonen A, Rhodius-Meester HFM, Bruun M, Koikkalainen J, Barkhof F, Lemstra AW, et al. Data-driven differential diagnosis of dementia using multiclass disease state index classifier. *Front Aging Neurosci* 2018;10:111.
- [46] Cluitmans L, Mattila J, Runtti H, van Gils M, Lotjonen J. A MATLAB toolbox for classification and visualization of heterogeneous multiscale human data using the Disease State Fingerprint method. *Stud Health Technol Inform* 2013;189:77–82.
- [47] Verberk I, Slot RE, Verfaillie SC, Heijst H, Prins ND, Van Berckel B, et al. Plasma-amyloid as pre-screener for the earliest Alzheimer's pathological changes. *Ann Neurol* 2018. in press.
- [48] Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* 2013;12:207–16.
- [49] Slot RE, Sikkes S, Berkhof J, Brodaty H, Buckley R, Cavado E, et al. Subjective cognitive decline and rates of incident Alzheimer's disease (AD) and non-AD dementia. *Alzheimers Dement* 2018. in press.
- [50] Kunneman M, Pel-Littel R, Bouwman FH, Gillissen F, Schoonenboom NSM, Claus JJ, et al. Patients' and caregivers' views on conversations and shared decision making in diagnostic testing for Alzheimer's disease: the ABIDE project. *Alzheimer's Dementia (N Y)* 2017;3:314–22.