HHS Public Access

Author manuscript

Alzheimers Dement. Author manuscript; available in PMC 2021 November 01.

Published in final edited form as:

Alzheimers Dement. 2020 November ; 16(11): 1524–1533. doi:10.1002/alz.12140.

AD risk score for the early phases of disease based on unsupervised machine learning

Zheyu Wang[a,b,*], **Zhuojun Tang**[a], **Yuxin Zhu**[a,b], **Corinne Pettigrew**[c], **Anja Soldan**[c], **Alden Gross**[d,e], **Marilyn Albert**[c], **BIOCARD research team and the Alzheimer's Disease Neuroimaging Initiative**[f]

[a]Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[b]Department of Biostatistics, Johns Hopkins University School of Public Health, Baltimore, MD, USA

[c]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[d]Department of Epidemiology, Johns Hopkins University School of Public Health, Baltimore, MD, USA

[e]Johns Hopkins University Center on Aging and Health, Baltimore, MD, USA

[f]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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

Abstract

INTRODUCTION—Identifying cognitively normal individuals at high risk for progression to symptomatic Alzheimer's disease (AD) is critical for early intervention.

METHODS—An AD risk score was derived using unsupervised machine learning. The score was developed using data from 226 cognitively normal individuals and included cerebrospinal fluid, magnetic resonance imaging, and cognitive measures, and validated in an independent cohort.

RESULTS—Higher baseline AD progression risk scores (HR=2.70, p<0.001) were associated with greater risks of progression to clinical symptoms of Mild Cognitive Impairment (MCI). Baseline scores had an AUC of 0.83 (95% CI 0.75–0.91) for identifying subjects who progressed to MCI/dementia within 5 years. The validation procedure, using data from the Alzheimer's Disease Neuroimaging Initiative, demonstrated accuracy of prediction across the AD spectrum.

DISCUSSION—The derived risk score provides high predictive accuracy ovh promising predictive accuracy due to ADp dementia on individuals are likely to become symptomatic over

* wangzy@jhu.edu.

Conflict of Interest: Dr. Albert is a consultant to Eli Lilly. All other authors have no conflict of interest.

time without incudifor identifying which individuals with normal cognition are likely to show clinical decline due to AD within 5 years.

Keywords

machine learning; unsupervised learning; risk score; latent variable; multidomain biomarkers; cognitive testing; progression; Alzheimer's disease

1. Introduction

Recognition of the long asymptomatic stage of Alzheimer's disease (AD) has pushed the focus of AD research and therapeutic development to the preclinical phase, when disease-modifying therapies are hypothesized to be more effective [1]. One of the biggest challenges of clinical trials is to identify asymptomatic individuals at risk of developing clinical symptoms of AD over a relatively short time frame, so that potential treatment effects can be observed over the course of a trial. The failure of recent trials, and low progression rates of asymptomatic individuals enrolled in clinical trials, emphasize the importance of improved methods for accurate identification of individuals at high risk of progression at the early phases of disease [2].

Recent efforts to improve AD risk prediction have sought to combine sets of variables together that have been linked to symptom onset. These risk prediction approaches have often combined measures associated with AD pathophysiological processes (e.g., amyloid deposition, neurofibrillary tangles and neurodegeneration) [3] and cognitive test scores. They are thought to be predictive because they quantify the underlying burden of disease.

Prior approaches for combining sets of variables into a risk score have used regression techniques that link biomarkers (predictors) with a clinical diagnosis observed years later (outcome). This approach requires the availability of marker measurements from asymptomatic individuals who have undergone longitudinal follow-up for determining clinical diagnostic outcomes. While the use of clinical diagnostic outcomes in constructing risk scores has been important in establishing the feasibility of this approach, it has recognized limitations. First, prediction models that use clinical diagnoses as the outcome and cognitive test scores as model predictors may overestimate prediction accuracy (especially when the cognitive decline is not due to AD). Although removing cognitive tests as model predictors can avoid this issue, the resulting model would likely have lower accuracy without contribution from the strongly predictive cognitive tests [1]. Secondly, since clinical symptom onset is a manifestation of underlying disease burden, the question has been raised whether prediction models that are both fit and evaluated based on clinical diagnoses are optimal [4]. Additionally, identifying the etiology of the clinical diagnoses is prone to error, especially during the earliest disease stages, causing potential errors in the outcome variable used to build the model [5].

One method for combining biomarkers into a risk score without relying on clinical diagnostic information is through unsupervised machine learning [6]. A small number of prior studies have explored the use of such approaches for modeling AD biomarkers and have obtained risk scores with promising predictive accuracy [4,7,8]. However, the

unsupervised learning approaches used in these prior studies require strong distributional assumptions, such as normality and independence between the biomarkers, which are questionable in the context of AD and may, therefore, affect the accuracy and interpretability of the results.

The present study applies an unsupervised machine learning approach to develop an AD progression risk score, using a recently developed method tailored to AD [9,10]. It allows for synchronizing measurements of different scales and different biomarker domains, and with potentially heterogonous and correlated distributions. Further, this method improves individualized prediction because it more robustly incorporates disease complexity by simultaneously modeling factors that may affect AD progression, as well as factors that may affect an individual's biomarker levels independently of the disease process. For example, this method can incorporate risk factors for AD pathophysiology (such as APOE-ε4 genetic status), as well as individual characteristics that might affect marker levels, independent of AD (such as the aging process or neuronal injury caused by non-AD processes). Here, we applied this new method to data from the BIOCARD study, whose participants were cognitively normal at baseline, had comprehensive biomarker measurement, and up to 21.6 years of clinical follow-up. Using these data, we evaluated the derived risk score's performance in terms of discrimination ability, accuracy for predicting subsequent clinical progression, and comparison to previously published comparable scores. In addition, we performed external validation using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

2. Methods

2.1 Participants and Study Design

The BIOCARD study is a longitudinal, observational study initiated in 1995 at the National Institutes of Health. At baseline, the study enrolled 349 cognitively normal (CN) individuals who were primarily middle-aged (mean age=57.1). By design, about three-quarters of the participants had a first-degree relative with dementia. Clinical assessments and cognitive testing were completed annually; MRI scans, cerebrospinal fluid, and blood specimens were collected approximately every 2 years. The study was stopped in 2005 and reinitiated at Johns Hopkins University in 2009, with clinical and cognitive assessments again occurring annually. All participants received annual consensus diagnoses (using procedures comparable to those employed by the National Institute on Aging Alzheimer's Disease Centers program), according to which they were classified as CN, MCI or dementia (due to AD and/or another etiology; see [11] for details). For individuals who received a diagnosis of MCI or dementia, information from the CDR interview was used to estimate the age at which the clinical symptoms began (see Appendix A for further details of the diagnostic procedures). This study provides over 20 years of longitudinal clinical follow-up from subjects who were initially cognitively normal (median follow up 13.7 years). Data from 226 individuals with complete data for the variables of interest were included in the current study.

ADNI is a multicenter observational study launched in 2003 that has collected clinical, imaging, genetic and biospecimen data from individuals across the spectrum of AD. The

ADNI1 study enrolled 200 CN, 400 MCI and 200 AD subjects between 55 and 90 years of age, classified based on memory criteria, the MMSE and CDR score. Subjects were followed every 6 or 12 months (See http://www.adni-info.org for up-to-date information). This data set complements the BIOCARD data in two respects: 1) ADNI1 participants with normal cognition at enrollment provided a dataset for externally validating the results obtained using BIOCARD data; and 2) inclusion of ADNI1 participants in the symptomatic phase of AD (MCI/dementia) allowed us to apply the prediction model to participants across a wider spectrum of marker ranges. Data from 352 individuals with complete data for the variables of interest who were followed for 4.8 years were included in the validation analyses.

2.2 Measurements

Three domains of interest were included in the analyses: 1) CSF measures; 2) MRI measures; and 3) cognitive test scores. These measures were selected a priori, based on the following criteria: first, each selected marker had been shown to predict progression from normal cognition to MCI or dementia in the BIOCARD cohort [11,12,13]. Second, each domain was measured in both the BIOCARD and ADNI studies. Third, the marker values had adequate variability across study participants. All of the CSF, MRI, and cognitive data were collected at baseline within a mean of 1.7 months of each other.

CSF markers included 1) amyloid-$\beta_{1–42}$, 2) p-tau$_{181p}$ and 3) t-tau. In both BIOCARD and ADNI1, the CSF assays employed the xMAP platform (Luminex Corp, Austin, Texas) and INNO-BIA AlzBio3 research-use-only reagents [14].

MRI markers included 1) entorhinal cortex thickness, 2) entorhinal cortex volume and 3) hippocampus volume, each averaged over the left and right hemispheres. The volumetric measures were adjusted for intracranial cavity volume (ICV) by division. BIOCARD MRI scans were collected on a GE 1.5T scanner. The medial temporal lobe (MTL) volumes were derived using region of interest large deformation diffeomorphic metric mapping (ROI-LDDMM) techniques, whereas ICV was derived from the FreeSurfer pipeline (version 5.1) [15]. ADNI scans were collected on a 3T scanner and processed using the cross-sectional and longitudinal FreeSurfer pipelines (version 5.1) [16,17,18,19], using the 2010 Desikan-Killiany atlas [20]. ADNI MTL volumes and ICV were derived from this pipeline.

Cognitive tests for both studies included: 1) Digit Symbol Substitution Test (DSST) from the Wechsler Adult Intelligence Scale – Revised [21] and 2) Logical Memory (LM) delayed recall from the Wechsler Memory Scale – Revised [22]. Verbal Paired Associates (Paired) immediate recall from the Wechsler Memory Scale – Revised [22] was included for BIOCARD only.

Demographic and genetic variables included age at baseline assessment, sex, years of education, and ApoE-ε4 genetic status (present/absent).

2.3 The model

An illustrative graph of the model is shown in Figure 1. Briefly, the model treats the AD risk as a latent variable (dashed circle) that is calculated using observable measures derived at a

single time point (in this case, baseline; blue solid rectangles). The model provides a risk score reflecting the cumulative burden of disease for each individual at a single point in time. Importantly, follow-up clinical diagnostic information was not used in the development of the model.

The model incorporates demographic variables (green solid rectangles) that may be related to underlying AD pathophysiology (such as age), and those that may be directly related to marker levels, potentially independent of AD (such as APOE genetic status). For example, there is evidence that neurodegeneration may occur independently of AD [23,24,25]. To account for both age- and disease-related marker changes, age was allowed to be related to both the underlying burden of disease (the latent variable) as well as directly to the marker levels (the manifest variables).

Once an overall model structure was generated, based on unsupervised machine learning, it was possible to infer data when only some of the markers were available, based on the partial structure, without refitting or imputing the missing markers. This feature allowed us to: 1) examine the contribution of each marker domain to the underlying burden of disease, and 2) examine the portability of the derived score to an external data set in which only some of the markers are available, as long as marker data were missing at random. For more details about this unsupervised machine learning approach see Appendix B and [9,10].

2.4 Statistical Analysis

2.4.1 Model fitting—We applied the unsupervised machine learning approach to the CSF, MRI, cognitive, demographic and genetic variables specified above. Several possible structures among these variables were examined, including adding/removing correlations between age, ApoE-ε4 status and each marker domain beyond their correlations through the underlying burden of disease latent variable, adding/removing direct correlations between education and cognitive tests, and adding/removing each of the demographic variables. All models were fitted on baseline variables only, except for the exploratory analysis in 2.4.3 examining changes in the risk scores over time. The final structure was selected based on the Akaike Information Criterion (AIC) [26]. All analyses were done in R version 3.5.0 (Vienna, Austria).

2.4.2 Criterion validity of the derived risk score—We first examined the extent to which the derived baseline risk score was associated with time to progression from normal cognition to clinical symptom onset of MCI. The hazard ratio (HR) of the risk score derived from the full model (i.e., including all markers), as well as risk scores based on individual marker domains, were calculated using Cox proportional hazards models. The calculations were based on standardized risk scores on the linear scale to have a mean of 0 and variance of 1, in order to facilitate model comparisons.

We then examined the ability of the risk score to predict later clinical progression on an individual basis, quantified by the area under the ROC curve (AUC). AUCs were calculated based on diagnoses at 5 and 10 years after the baseline and at the last follow-up visit. AUCs were assessed based on two criteria: one used progression to MCI or dementia due to any etiology, and the other used progression to MCI or dementia due to AD (defined as

progression with possible or probable AD as part of the individual's etiology). Separate AUCs for each of the marker domains were also examined to inform the potential predictive accuracy of the risk score when some marker measures are unavailable.

2.4.3 Construct validity of the derived AD risk score—We compared the baseline levels and rates of change in the AD risk score between subjects who remained cognitively normal over time and subjects who progressed to MCI/dementia using linear mixed effect models with random intercepts and slopes, and with natural spline on age (three nodes evenly spread over the range). Models were not adjusted for demographic or genetic variables because those were included in the AD risk score model.

2.4.4 Comparison to other AD risk scores based on BIOCARD data—We then compared our risk score with two comparable risk scores that used data from the BIOCARD study. The first was the AD severity score developed by Gross et al. [7], which we refer to as the "Gross score". This score was obtained using a similar unsupervised learning approach without relying on clinical diagnostic information, and thus provides information about predictive utility using a similar analytic method. The second was the risk score developed by Albert et al. [27], which we refer to as the "Albert score". This score incorporated clinical diagnostic information (i.e., using Cox proportional hazard models to link clinical diagnoses with the domain markers), and thus used more information than the scores based on unsupervised learning approaches. This latter comparison was designed to examine the predictive utility of our risk score relative to a risk score that utilized diagnostic information in modeling fitting and therefore may overestimate predictive accuracy.

To make a direct comparison, we restricted the variables used in our risk score to be the same as those used in the comparison risk scores. Thus, when modeling the Gross score, our risk score markers were reduced to the following set of variables: CSF amyloid-β, CSF p-tau, hippocampal volume, DSST and California verbal learning test [28], as well as the demographic variables age and ApoE-ε4 genetic status. Of note, Functional Activities Questionnaire (FAQ) scores [29] were included in the Gross score, but were not included in our score, since this measure tends to be consonant with clinical diagnoses. When modeling the Albert score, our risk score used the following set of variables: CSF amyloid-β, CSF p-tau, right hippocampal volume, right entorhinal cortex thickness, DSST and Paired Associates immediate recall, as well as the demographic variables age and years of education and ApoE-ε4 genetic status.

2.4.5 External validation using ADNI data—We then applied the AD risk score to ADNI data in order to: 1) provide external validation in an independent cohort with normal cognition at baseline; 2) evaluate the risk score's predictive accuracy in a population with greater heterogeneity in baseline diagnosis and later progression.

We obtained risk scores for ADNI participants using parameters estimated from the BIOCARD model by substituting ADNI subjects' measurements at baseline without refitting the model. Due to the shorter clinical follow-up of the ADNI cohort, the risk scores were compared against clinical diagnoses at 5 years, and the last available follow-up visit. As noted above, since ADNI did not use the Paired Associates test, we applied the model based

on BIOCARD data, but utilized the partial structure without this test. As a comparison, we also calculated the Albert score in ADNI. Directly taking the partial structure without the Paired Associates test may not be appropriate for the Albert score, given calculation of this score involves creating weights for each marker measurement based on a Cox proportional hazard model fit. We therefore updated the Albert score in BIOCARD by refitting the Albert model without this test. The risk scores for ADNI participants were then obtained by substituting ADNI measurements in the updated Albert score, without further refitting of the Albert model.

3. Results

The baseline characteristics of BIOCARD and ADNI participants included in the analyses are summarized in Table 1. Baseline characteristics of the entire BIOCARD and ADNI cohorts are reported in Appendix C.

3.1 The final model

The final structure obtained based on the unsupervised machine learning approach is shown in Figure 1, which depicts the direct relationships among the latent variable and the covariates (e.g., age, genetic status, etc.). Based on AIC criteria, CSF total tau was excluded from the final model (its inclusion results in a similar likelihood value, likely reflecting its high correlation with p-tau).

3.2 Predictive accuracy of the AD risk score based on all or each of the three marker domains

The risk score based on all markers, or on individual marker domains, was significantly associated with progression to clinical symptom onset of MCI, both when all etiologies were included, as well as when the etiologies were restricted to MCI/dementia due to AD (Table 2, left). The full AD risk score had the highest hazard ratio, followed by the cognitive domain and then by the CSF markers. The difference in hazard ratios between the model with all markers and the models with single marker domains were all significant (all $p < 0.05$). However, these tests do not adjust for multiple comparison, and are *ad-hoc* for exploratory purpose only.

The AUC of the risk score for predicting progression from normal cognition to MCI/dementia with any etiology within 5 years was 0.83 (95% CI 0.75 – 0.91). When restricting progression to AD etiology only, the AUC was 0.88 (95% CI 0.81 – 0.96). Among the three individual marker domains, cognitive tests had the highest AUC. Although the differences between the AUCs calculated for the full model (based on all domains) vs. cognitive tests only were numerically small, all but one of these differences were significant (Appendix D). Notably, the accuracy of the MRI markers and the CSF markers were similar when considering all etiologies as the outcome, but were numerically higher (though not statistically significant) when the models were restricted to progression to AD etiology only.

The results of the linear mixed effect models are shown in Figure 2. This is the only model in the paper that requires longitudinal biomarker measurements. The analysis has been done with BIOCARD data only and not with ADNI, due to the limited amount of longitudinal

CSF and MRI measures available in ADNI. Subjects who progressed to MCI/dementia have higher baseline risk scores ($p<0.01$) and tend to have steeper increases in risk score trajectories ($p>0.05$), as compared to the subjects who remained cognitively normal.

3.3 Comparison of the AD risk score with previously published risk scores

Table 3 summarizes the comparisons of the current risk score with the Gross and Albert risk scores. The AUCs for our model, relative to the Albert score, were not statistically different for predicting progression from normal cognition to MCI symptom onset (any etiology) at 5 years ($p =0.48$) or at 10 years ($p=0.12$). The AUC of the Albert score was higher for predicting progression by the last diagnosis ($p=0.01$). In contrast, our score had significantly higher AUCs than the Gross score for predicting progression from normal cognition to MCI symptom onset (Table 3).

3.4 External validation with ADNI

Results from the validation of the risk score using the ADNI data are summarized in Table 4. The 2nd-3rd columns show results for ADNI subjects who were initially cognitively normal. The AUCs for predicting progression to MCI were comparable to those in the BIOCARD study.

We also evaluated the accuracy of the risk scores using the subset of ADNI participants who had a diagnosis of MCI at baseline. Interestingly, the AUCs for predicting progression from MCI to dementia were again comparable. For these individuals, however, the MRI domain tended to have slightly higher AUCs than the cognitive domain, suggesting that MRI markers may be more predictive at later diagnostic time points. In addition, we examined the accuracy of the AD risk score in discriminating groups that had a diagnosis of MCI or AD dementia at any time versus those who were cognitively normal at all study visits. The discrimination accuracy was high, AUC = 0.94 (95% CI 0.92 – 0.97).

Of note, the corresponding Albert score in ADNI still exhibited good accuracy for predicting progression in both the initially normal cohort and those with a diagnosis of MCI at baseline. However, the AUCs were not as high as those in BIOCARD. This is a potential illustration of the over-estimation of accuracy caused by simultaneously using clinical diagnosis as outcome and cognitive tests as predictor, as discussed previously.

4. Discussion

This study examined the utility of a novel unsupervised machine learning approach for generating an AD risk score among cognitively normal individuals to predict onset of clinical symptoms of MCI. In two independent cohorts, the risk score produced high predictive accuracy for later clinical outcomes. The AD risk score performed comparably to a previously published risk score that incorporated follow-up clinical diagnoses (the Albert score), and out-performed another risk score that utilized an unsupervised learning approach (the Gross score). Importantly, in applying this risk score derived from BIOCARD data to ADNI data, the AUCs for predicting progression from normal cognition to MCI were comparable, demonstrating strong external validation. The AD risk score also exhibited comparable accuracy for predicting progression from MCI to dementia in the ADNI dataset.

Taken together, these results indicate that the AD risk score based on unsupervised machine learning (and that excludes follow-up diagnostic information), significantly captures the underlying disease burden, is portable to other cohorts, and can be applied over wide age ranges (given the differences in baseline ages between the BIOCARD and ADNI cohorts), as well as to individuals along the AD continuum.

The analytic approach used here has notable advantages. First, it utilizes more flexible modeling assumptions and techniques, therefore enabling the underlying pathophysiology to be more robustly captured [10]. For example, compared with other unsupervised machine learning approaches, the method does not require that all markers be on a common scale (e.g., all categorical or all continuous). This therefore improves the ability to combine a wide range of AD markers. In addition, the approach can be easily adapted when new biomarkers are identified, for testing whether these further improve the predictive accuracy of the derived risk score. Examples include: (1) medical co-morbidities such as diabetes and hypertension, (2) genetic markers such as polygenic risk scores or outcomes of whole-exome sequencing [34], and (3) additional biomarkers such as inflammation markers [36], that may be identified in the future with the increasing availability of data and the advance of artificial intelligence techniques. Second, the method does not rely on clinical diagnostic information in the construction of the risk score, and therefore avoids the potential overfitting that may occur when clinical diagnoses are used in both risk score derivation and evaluation. Moreover, the method allows risk scores to be derived when only a subset of markers are available. This allows one to study the contribution of different marker domains, and provides information about the model's predictive accuracy in situations where some markers may be unavailable. For example, the full model tended to perform the best, confirming that CSF and MRI markers make important contributions to risk score accuracy. Consistent with this, as shown by Baker et al. [35], a small change in AUC can lead to meaningful improvements in risk prediction for a range of relevant threshold. However, risk scores based on the cognitive domain also performed well, especially among individuals who were initially cognitively normal. This is notable because the estimated predictive accuracy of the individual marker domains can inform study planning in resource limited settings, such as situations in which obtaining study funding and/or participants' consent for biomarker procedures in cognitively normal subjects may be difficult. Additional studies are needed to evaluate the ability of these approaches to refine recruitment in clinical trial settings.

The study has several limitations. Participants from both cohorts are well educated and primarily Caucasian, and the majority of BIOCARD participants have a family history of dementia, therefore limiting the generalizability of these data to the population at large. Moreover, the BIOCARD and ADNI cohorts have modest sample sizes. Future studies with larger, more diverse samples are needed to further validate and potentially refine this approach, before it can be tested in clinical settings. Additionally, unsupervised learning approaches themselves are limited insofar as they require specifying *a priori* the markers related to the underlying construct of interest (e.g., risk for AD). Although we selected a set of measures that have been shown to be associated with AD pathophysiology and risk of progression, the results may vary with the use of different makers (e.g., PET amyloid). Thirdly, although the derived AD risk score exhibits high accuracy for predicting subsequent

clinical outcomes, there is still room for improvement. The overlap in between the individuals who remain normal and those who progress to the symptomatic phase of disease (as shown in Figure 2) suggests that the markers used in the model may not fully capture all of the risk factors or the underlying AD pathophysiological process. Additional risk factors or biomarkers of disease can be included in future analyses to examine this issue further. Finally, a more general issue in AD risk modeling is the need to account for differences in the absolute biomarker values between cohorts (e.g., due to the use of different assays and/or image processing techniques), which can limit the portability of many derived scores. Although the impact of this issue was reduced in our approach due to its flexibility regarding biomarker distributions, a unified protocol for AD biomarkers across studies remains a key issue for further generalization and accuracy improvement.

It is hoped that, with AD clinical trials moving towards earlier disease phases, analytic methods that capture the underlying burden of disease, such as presented here, may help identify individuals at high risk of progression and thereby accelerate the development of effective treatments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We are grateful for the support of the entire BIOCARD study team at Johns Hopkins University, the ADNI study team, and to the dedicated participants who continue to participate in these studies. This study was funded by a pilot project grant from the Johns Hopkins Alzheimer's Disease Research Center (P50-AG00146). Dr. Gross was supported by NIA K01-AG050699. The BIOCARD study is supported by a grant from the National Institute on Aging: U19-AG033655.

The BIOCARD Study consists of 7 Cores with the following members: (1) the Administrative Core (Marilyn Albert, Rostislav Brichko), (2) the Clinical Core (Marilyn Albert, Anja Soldan, Corinne Pettigrew, Rebecca Gottesman, Ned Sacktor, Scott Turner, Leonie Farrington, Maura Grega, Gay Rudow, Scott Rudow), (3) the Imaging Core (Michael Miller, Susumu Mori, Tilak Ratnanather, Anthony Kolasny, Kenichi Oishi, Laurent Younes), (4) the Biospecimen Core (Abhay Moghekar, Jacqueline Darrow, Richard O'Brien), (5) the Informatics Core (Roberta Scherer, David Shade, Ann Ervin, Jennifer Jones, Hamadou Coulibaly, April Patterson), the (6) Biostatistics Core (Mei-Cheng Wang, Daisy Zhu, Jiangxia Wang), and (7) the Neuropathology Core (Juan Troncoso, Olga Pletnikova, Gay Rudow, Karen Fisher).

The ADNI 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: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to 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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1]. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ and Park DC, 2011 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. Alzheimer's & dementia, 7(3), pp.280–292.

[2]. Gauthier S, Albert M, Fox N, Goedert M, Kivipelto M, Mestre-Ferrandiz J and Middleton LT, 2016 Why has therapy development for dementia failed in the last two decades?. Alzheimer's & Dementia, 12(1), pp.60–64.

[3]. Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD and Lesnick TG, 2013 Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology, 12(2), pp.207–216. [PubMed: 23332364]

[4]. De Meyer G, Shapiro F, Vanderstichele H, Vanmechelen E, Engelborghs S, De Deyn PP, Coart E, Hansson O, Minthon L, Zetterberg H and Blennow K, 2010 Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. Archives of neurology, 67(8), pp.949–956. [PubMed: 20697045]

[5]. Zhou XH, McClish DK and Obuchowski NA, 2009 Statistical methods in diagnostic medicine (Vol. 569). John Wiley & Sons.

[6]. Hastie T, Tibshirani R and Friedman J, 2009 Unsupervised learning. In *The elements of statistical learning* (pp. 485–585). Springer, New York, NY.

[7]. Gross AL, Mungas DM, Leoutsakos JMS, Albert MS, Jones RN and Alzheimer's Disease Neuroimaging Initiative, 2016 Alzheimer's disease severity, objectively determined and measured. Alzheimer's & Dementia: Diagnosis, Assessment & Disease *Monitoring*, 4, pp.159–168.

[8]. Leoutsakos JM, Gross AL, Jones RN, Albert MS and Breitner JCS, 2016 'Alzheimer's Progression Score': Development of a Biomarker Summary Outcome for AD Prevention Trials. The journal of prevention of Alzheimer's disease, 3(4), p.229.

[9]. Wang Z, Sebestyen K and Monsell SE, 2017 Model-based clustering for assessing the prognostic value of imaging biomarkers and mixed type tests. Computational statistics & data analysis, 113, pp.125–135. [PubMed: 28966420]

[10]. Wang Z and Zhou XH, 2018 Biomarker assessment and combination with differential covariate effects and an unknown gold standard, with an application to Alzheimer's disease. The Annals of Applied Statistics, 12(2), pp.1204–1227.

[11]. Albert M, Soldan A, Gottesman R, McKhann G, Sacktor N, Farrington L, Grega M, Turner R, Lu Y, Li S and Wang MC, 2014 Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to ApoE genotype. Current Alzheimer Research, 11(8), pp.773–784. [PubMed: 25212916]

[12]. Moghekar A, Li S, Lu Y, Li M, Wang MC, Albert M, O'brien R and BIOCARD Research Team, 2013 CSF biomarker changes precede symptom onset of mild cognitive impairment. Neurology, 81(20), pp.1753–1758. [PubMed: 24132375]

[13]. Soldan A, Pettigrew C, Lu Y, Wang MC, Selnes O, Albert M, Brown T, Ratnanather JT, Younes L, Miller MI and BIOCARD Research Team, 2015 Relationship of medial temporal lobe atrophy, APOE genotype, and cognitive reserve in preclinical A lzheimer's disease. Human brain mapping, 36(7), pp.2826–2841. [PubMed: 25879865]

[14]. Shaw LM, 2009 Alzheimer's Disease Neuroimaging Initiative. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol, 65, pp.403–413. [PubMed: 19296504]

[15]. Miller MI, Younes L, Ratnanather JT, Brown T, Trinh H, Postell E, Lee DS, Wang MC, Mori S, O'Brien R and Albert M, 2013 The diffeomorphometry of temporal lobe structures in preclinical Alzheimer's disease. NeuroImage: Clinical, 3, pp.352–360. [PubMed: 24363990]

[16]. Dale AM, Fischl B and Sereno MI, 1999 Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage, 9(2), pp.179–194. [PubMed: 9931268]

[17]. Dale AM and Sereno MI, 1993 Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. Journal of cognitive neuroscience, 5(2), pp.162–176. [PubMed: 23972151]

[18]. Fischl B and Dale AM, 2000 Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences, 97(20), pp.11050–11055.

[19]. Reuter M, Schmansky NJ, Rosas HD and Fischl B, 2012 Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage, 61(4), pp.1402–1418. [PubMed: 22430496]

[20]. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT and Albert MS, 2006 An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage, 31(3), pp.968–980. [PubMed: 16530430]

[21]. Wechsler D, 1981 Manual for the Wechsler adult intelligence scale-revised. New York: Psychological Corporation, 198.

[22]. Wechsler D, 1987 Manual for the Wechsler memory scale-revised. San Antonio, TX: Psychological Corporation.

[23]. Jack CR Jr, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, Kantarci K, Gunter JL, Senjem ML, Ivnik RJ and Roberts RO, 2012 An operational approach to National Institute on Aging–Alzheimer's Association criteria for preclinical Alzheimer disease. Annals of neurology, 71(6), pp.765–775. [PubMed: 22488240]

[24]. Crary JF, Trojanowski JQ, Schneider JA, Abisambra JF, Abner EL, Alafuzoff I, Arnold SE, Attems J, Beach TG, Bigio EH and Cairns NJ, 2014 Primary age-related tauopathy (PART): a common pathology associated with human aging. Acta neuropathologica, 128(6), pp.755–766. [PubMed: 25348064]

[25]. Resnick SM, Pham DL, Kraut MA, Zonderman AB and Davatzikos C, 2003 Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. Journal of Neuroscience, 23(8), pp.3295–3301. [PubMed: 12716936]

[26]. Akaike H, 1974 A new look at the statistical model identification In Selected Papers of Hirotugu Akaike (pp. 215–222). Springer, New York, NY.

[27]. Albert M, Zhu Y, Moghekar A, Mori S, Miller MI, Soldan A, Pettigrew C, Selnes O, Li S and Wang MC, 2018 Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. Brain, 141(3), pp.877–887. [PubMed: 29365053]

[28]. Delis DC, Kramer JH, Kaplan E and Thompkins BAO, 1987 CVLT: California verbal learning test-adult version: manual. Psychological Corporation.

[29]. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM and Filos S, 1982 Measurement of functional activities in older adults in the community. Journal of gerontology, 37(3), pp.323–329. [PubMed: 7069156]

[30]. DeLong ER, DeLong DM and Clarke-Pearson DL, 1988 Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics, pp.837–845. [PubMed: 3203132]

[31]. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC and Snyder PJ, 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. Alzheimer's & dementia, 7(3), pp.270–279.

[32]. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R and Mohs RC, 2011 The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia, 7(3), pp.263–269.

[33]. Ma Y, Jun GR, Zhang X, Chung J, Naj AC, Chen Y, Bellenguez C, Hamilton-Nelson K, Martin ER, Kunkle BW and Bis JC, 2019 Analysis of whole-exome sequencing data for Alzheimer disease stratified by APOE genotype. JAMA neurology, 76(9), pp.1099–1108.

[34]. Baker SG, Schuit E, Steyerberg EW, Pencina MJ, Vickers A, Moons KG, Mol BW and Lindeman KS, 2014 How to interpret a small increase in AUC with an additional risk prediction marker: decision analysis comes through. Statistics in medicine, 33(22), pp.3946–3959. [PubMed: 24825728]

[35]. Gross AL, Walker KA, Moghekar A, Pettigrew C, Soldan A, Albert MS and Walston J, 2019 Plasma markers of inflammation linked to clinical progression and decline during preclinical AD. Frontiers in aging neuroscience, 11, p.229. [PubMed: 31555121]

[36]. Dempster AP, Laird NM and Rubin DB, 1977 Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society:* Series B (Methodological), 39(1), pp.1–22.

Highlights

1. Applied unsupervised machine learning to derive an AD progression risk score
2. Risk score was examined across the spectrum of preclinical and prodromal stages
3. Risk score had high predictive accuracy for identifying progression within 5 years
4. Results may be relevant to selecting subjects for clinical trials

Research in Context

Systematic review

With AD clinical trials shifting towards earlier disease phases, it is becoming increasingly important to identify asymptomatic individuals at high risk of progression to the symptomatic phase of AD for inclusion in clinical trials. Previous AD risk scores have been developed, combining measures from multiple domains together. These prior efforts incorporated clinical diagnostic outcomes, which may overestimate predictive accuracy. We attempted to address this issue by applying an unsupervised machine learning approach that excluded diagnostic information in model fitting. The model was then validated in an independent cohort.

Interpretation

The AD risk score shows promising predictive accuracy for identifying individuals who are likely to progress clinically along the AD spectrum, from the preclinical and prodromal phases of AD.

Future directions

Further validation studies are needed with larger and more generalizable populations. Potential applications of the risk score in AD clinical trials to facilitate recruitment of high-risk individuals.

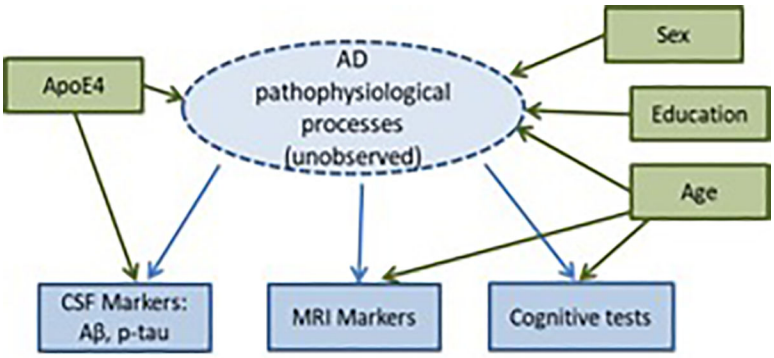

Figure 1:

An illustrative graph of the unsupervised machine learning model. MRI markers are entorhinal cortex thickness, entorhinal cortex volume and hippocampus volume. Cognitive tests included DSST, LM and Paired Associates.

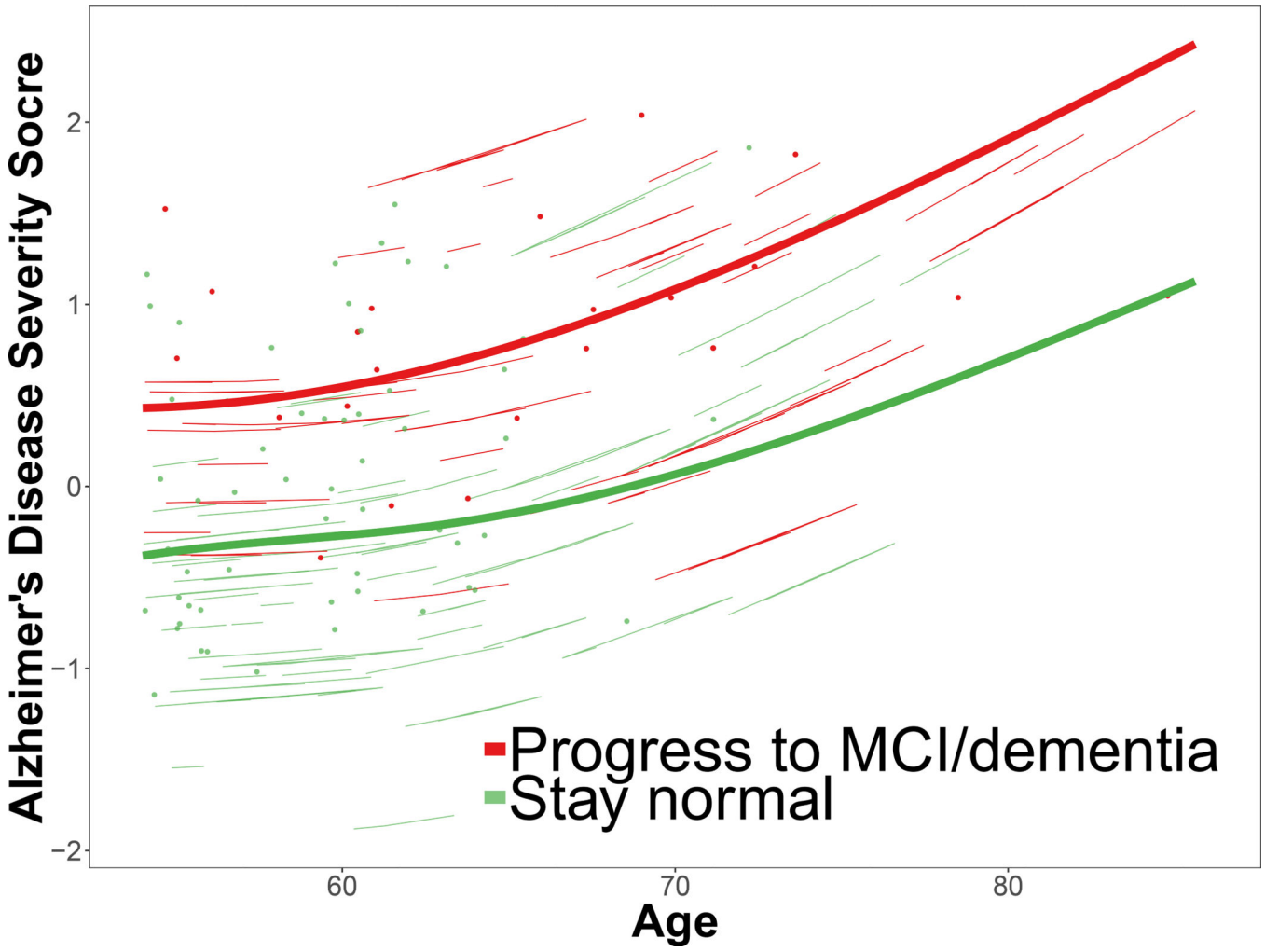

Figure 2:
Trajectory of the AD risk score by age among different diagnostic groups in the BIOCARD study (N=226). The thin lines reflect fitted individual trajectories, and the thick lines are group trajectories.

Table 1:

Baseline characteristics of the participants included in the analyses by diagnostic group (mean, standard deviation and range for continuous variables; counts and percentage for categorical variables). Baseline is defined as the first visit in which a participant has biomarker measures available.

	BIOCARD (N=226)		ADNI1 (N=352)		
Characteristics	**Remained normal (n = 156)**	**Progressed (n = 70)**	**Normal at baseline (n = 107)**	**MCI at baseline (n = 133)**	**Dementia at baseline (n = 112)**
Age, years	56.16 (±6.66)	61.29 (±10.33)	76.42 (±5.32)	75.76 (±7.52)	75.42 (±7.43)
Gender, % Female	95 (60.90%)	38 (54.29%)	51 (47.66%)	46 (34.59%)	44 (39.29%)
Education, years	17.20 (±2.34)	16.77 (±2.52)	15.85 (±2.92)	16.03 (±2.89)	15.28 (3.06)
Ethnicity, % Caucasians	155 (99.36%)	66 (94.29%)	99 (92.52%)	126 (94.74%)	111 (99.11%)
ApoE-ε4 carriers, %	49 (31.41%)	27 (38.57%)	23 (21.50%)	67 (50.38%)	76 (67.86%)
Years of follow-up	13.31 (±4.11)	13.70 (±3.63)	5.78 (±3.63)	3.78 (±2.78)	1.38 (±1.28)
MMSE score	29.65 (±0.66)	29.46 (±0.96)	29.07 (±1.18)	26.86 (±2.49)	22.03 (±4.14)
Progression					
Incident MCI	n/a	50 (71.43%)	9 (8.41%)	n/a	n/a
Incident dementia	n/a	20 (28.57%)	0	49 (36.84%)	n/a
CSF markers					
amyloid-β	417.09 (±93.74)	380.78 (±102.08)	207.21 (±52.29)	165.97 (±53.44)	143.63 (±41.60)
p-tau	33.64 (±11.81)	41.24 (±19.20)	29.38 (±16.08)	40.48 (±29.61)	45.54 (±23.12)
total tau	65.42 (±27.55)	81.17 (±39.17)	72.35 (±29.83)	102.63 (±54.44)	114.89 (59.68)
MRI markers					
Entorhinal cortex thickness (mm)	2.21 (±0.25)	2.11 (±0.26)	3.41 (±0.32)	3.04 (±0.48)	2.69 (±0.45)
Entorhinal cortex volume (mm^3)	458.4 (±115.8)	420.7 (±118.3)	1854.1 (±324.3)	1644.3 (±421.2)	1368.7 (±362.8)
Hippocampus volume (mm^3)	2690.0 (±273.0)	2719.3 (±342.9)	3601.1 (±404.8)	3097.6 (±532.4)	2736.8 (544.9)
Intracranial Volume (cm^3)	1548.5 (194.9)	1583.4 (252.1)	1539.0 (174.7)	1597.9 (167.0)	1568.1 (185.9)
Cognitive tests					
Digit Symbol Substitution	56.58 (±11.15)	46.84 (±8.64)	47.51 (±9.21)	38.29 (±12.70)	24.84 (±13.21)
Logical memory Delayed Recall	13.60 (±3.69)	11.30 (±3.60)	13.27 (±4.48)	4.66 (±4.15)	1.01 (±1.77)
Paired Associates Immediate Recall	20.96 (±2.71)	19.43 (±3.05)	NA	NA	NA
CVLT (BIOCARD) or AVLT (ADNI)	52.85 (±9.50)	50.08 (±10.34)	42.92 (±9.24)	29.92 (8.43)	21.14 (±8.49)

Abbreviations: MMSE - Mini-Mental State Examination; CVLT - California Verbal Learning Test; AVLT - Rey Auditory Verbal Learning Test.

Table 2.

Hazard ratios and predictive accuracy (AUC) of the AD risk score based on all markers, as well as risk scores based on each of the individual marker domains, in relation to future clinical diagnosis in the BIOCARD study

	Hazard ratio (95% CI)	AUC (95% CI*)		
		5 years	10 years	Last follow up
Progress to MCI/dementia (Any etiology) vs. stay normal				
	N progressed= 70	N=194 vs. 24	N=151 vs. 56	N=156 vs. 70
All markers	2.82 (2.15, 3.72)	0.83 (0.75, 0.91)	0.81 (0.75, 0.87)	0.78 (0.71, 0.84)
Cognitive	2.58 (1.97, 3.38)	0.79 (0.70, 0.88)	0.78 (0.71, 0.84)	0.76 (0.69, 0.82)
MRI	1.82 (1.46, 2.27)	0.74 (0.63, 0.85)	0.73 (0.65, 0.81)	0.69 (0.62, 0.77)
CSF	2.15 (1.67, 2.77)	0.73 (0.63, 0.84)	0.72 (0.63, 0.80)	0.69 (0.61, 0.77)
Cognitive + MRI	2.66 (2.03, 3.49)	0.82 (0.74, 0.90)	0.79 (0.73, 0.86)	0.77 (0.70, 0.83)
Cognitive + CSF	2.74 (2.08, 3.61)	0.81 (0.71, 0.90)	0.79 (0.72, 0.85)	0.77 (0.70, 0.83)
MRI+ CSF	2.06 (1.65, 2.56)	0.77 (0.66, 0.88)	0.76 (0.68, 0.84)	0.71 (0.64, 0.78)
Progress to MCI/dementia with AD etiology vs. stay normal *or* progress due to non-AD etiology				
	N progressed = 49	N=202 vs. 16	N=166 vs. 41	N=177 vs. 49
All markers	3.29 (2.34, 4.63)	0.88 (0.81, 0.96)	0.82 (0.75, 0.89)	0.78 (0.70, 0.85)
Cognitive	2.88 (2.07, 4.02)	0.83 (0.73, 0.92)	0.78 (0.71, 0.85)	0.75 (0.68, 0.83)
MRI	2.00 (1.54, 2.60)	0.79 (0.70, 0.89)	0.76 (0.67, 0.84)	0.70 (0.61, 0.78)
CSF	2.52 (1.85, 3.42)	0.76 (0.63, 0.88)	0.73 (0.63, 0.83)	0.71 (0.61, 0.80)
Cognitive + MRI	3.02 (2.17, 4.20)	0.87 (0.79, 0.94)	0.80 (0.73, 0.88)	0.76 (0.69, 0.84)
Cognitive + CSF	3.14 (2.23, 4.42)	0.85 (0.76, 0.94)	0.80 (0.73, 0.87)	0.77 (0.70, 0.84)
MRI + CSF	2.34 (1.80, 3.04)	0.83 (0.74, 0.92)	0.79 (0.71, 0.87)	0.73 (0.64, 0.81)

* 95% CIs for AUCs were calculated using Delong's methods [30].

Table 3.

Model comparison: AUCs (95% CI) of different risk scores for predicting progression to MCI or dementia at different time points using data from the BIOCARD study

	5 years	10 years	Last follow up	Use clinical information?
Stay Normal vs progress to MCI or dementia (Any etiology; N progressed = 70)				
Albert score*	0.84 (0.75, 0.93)	0.84 (0.78, 0.90)	0.83 (0.77, 0.89)	Yes
Our full AD risk score	0.83 (0.75, 0.91)	0.81 (0.75, 0.87)	0.78† (0.71, 0.84)	No
Our AD risk score using markers comparable to the Albert score	0.80 (0.71, 0.89)	0.78† (0.71, 0.84)	0.76† (0.69, 0.82)	No
Gross score*	0.58 (0.45, 0.71)	0.58 (0.49, 0.67)	0.50 (0.41, 0.58)	Used FAQ
Our AD risk score using markers comparable to the Gross score	0.77‡ (0.68, 0.86)	0.77‡ (0.70, 0.85)	0.76‡ (0.69, 0.83)	No
Stay Normal vs progress to MCI or dementia (AD etiology; N progressed = 49)				
Albert score*	0.91 (0.83, 0.99)	0.85 (0.78, 0.92)	0.83 (0.76, 0.89)	Yes
Our full AD risk score	0.88 (0.81, 0.96)	0.82 (0.75, 0.89)	0.78 (0.70, 0.85)	No
Our AD risk score using markers comparable to the Albert score	0.83 (0.73, 0.94)	0.78† (0.70, 0.86)	0.76† (0.68, 0.83)	No
Gross score*	0.62 (0.48, 0.77)	0.60 (0.50, 0.70)	0.51 (0.41, 0.61)	Used FAQ
Our AD risk score using markers comparable to the Gross score	0.73‡ (0.62, 0.85)	0.75‡ (0.66, 0.84)	0.73‡ (0.64, 0.82)	No

*Albert score and Gross score were calculated based on their original methods.

†Significant difference as compared with Albert score based on the DeLong method [30] for comparing AUCs.

‡Significant difference as compared with Gross score based on the DeLong method [30] for comparing AUCs.

Table 4.

External validation of the AD risk score using ADNI data: AUCs (95% CIs*) for distinguishing different diagnostic groups

	Initially normal cohort (N= 107): Stay CN vs progress to MCI or dementia Median follow-up: 4.97 years		Initially MCI cohort (N= 133): Stay MCI or revert to CN vs progress to dementia Median follow-up: 2.89 years		Entire cohort (N=352): CN vs MCI or dementia Median follow-up: 2.05 years	
	5 years	Last follow up	5 years	Last follow up	5 years	Last follow up
	48 vs. 9	98 vs. 9	27 vs. 49	84 vs. 49	51 vs. 41	106 vs. 246
Our full score ǂ	0.81 (0.67, 0.95)	0.78 (0.63, 0.93)	0.81 (0.72, 0.91)	0.70 (0.61, 0.79)	0.88 (0.81, 0.96)	0.94 (0.92, 0.97)
Albert score ǂ	0.75 (0.55, 0.95)	0.74 (0.55, 0.93)	0.79 (0.68, 0.89)	0.73 (0.64, 0.82)	0.78 (0.67, 0.88)	0.88 (0.84, 0.91)
Cognitive	0.80 (0.65, 0.94)	0.77 (0.62, 0.92)	0.68 (0.56, 0.80)	0.60 (0.50, 0.70)	0.87 (0.80, 0.95)	0.93 (0.90, 0.96)
MRI	0.64 (0.43, 0.85)	0.64 (0.43, 0.85)	0.70 (0.57, 0.83)	0.65 (0.55, 0.74)	0.74 (0.64, 0.84)	0.82 (0.77, 0.86)
CSF	0.67 (0.46, 0.88)	0.65 (0.44, 0.85)	0.55 (0.41, 0.69)	0.54 (0.44, 0.64)	0.66 (0.55, 0.78)	0.64 (0.57, 0.70)
Cognitive + MRI	0.82 (0.70, 0.95)	0.79 (0.67, 0.92)	0.79 (0.69, 0.89)	0.68 (0.58, 0.77)	0.88 (0.81, 0.96)	0.94 (0.92, 0.97)
Cognitive + CSF	0.80 (0.64, 0.97)	0.78 (0.62, 0.94)	0.72 (0.61, 0.84)	0.63 (0.54, 0.73)	0.88 (0.80, 0.96)	0.93 (0.90, 0.96)
MRI+CSF	0.66 (0.44, 0.88)	0.66 (0.44, 0.88)	0.76 (0.65, 0.88)	0.69 (0.60, 0.78)	0.76 (0.66, 0.86)	0.84 (0.80, 0.88)

*95% CIs for AUCs were calculated using Delong's methods [30].

ǂNeither our score nor Albert score, as computed here, used the Paired Associates test because this measure is unavailable in ADNI. Both scores were calculated by substituting ADNI subject's measurements into the corresponding models obtained in BIOCARD, without refitting (as described in section 2.4.5).