Published in final edited form as:
Biol Psychiatry. 2008 November 15; 64(10): 871–879. doi:10.1016/j.biopsych.2008.06.020.

Combining early markers strongly predicts conversion from mild cognitive impairment to Alzheimer's disease

D. P. Devanand, M.D., **Xinhua Liu, Ph.D.**, **Matthias H. Tabert, Ph.D.**, **Gnanavalli Pradhaban, M.B.B.S.**, **Katrina Cuasay, B.A.**, **Karen Bell, M.D.**, **Mony J. de Leon, Ed.D.**, **Richard L. Doty, Ph.D.**, **Yaakov Stern, Ph.D.**, and **Gregory H. Pelton, M.D.**

From the Division of Geriatric Psychiatry, New York State Psychiatric Institute, College of Physicians and Surgeons, Columbia University, New York (DPD, MHT, GP, KC, YS, GHP), Gertrude H. Sergievsky Center and Taub Institute for Research in Alzheimer's Disease and the Aging Brain, Department of Neurology, College of Physicians and Surgeons, Columbia University, New York (KB, YS), Department of Biostatistics, Columbia University, New York (XL), Center for Brain Health, Department of Psychiatry, New York University School of Medicine and the Nathan Kline Institute, Orangeburg, New York (MDL), and the Smell and Taste Center, University of Pennsylvania School of Medicine, Philadelphia, PA (RLD)

Abstract

Background—The utility of combining early markers to predict conversion from mild cognitive impairment (MCI) to Alzheimer's Disease (AD) remains uncertain.

Methods—148 outpatients with MCI, broadly defined, were followed at 6-month intervals. Hypothesized baseline predictors for follow-up conversion to AD (entire sample: 39/148 converters) were cognitive test performance, informant report of functional impairment, apolipoprotein E genotype, olfactory identification deficit, MRI hippocampal and entorhinal cortex volumes.

Results—In the 3-year follow-up patient sample (33/126 converters), five of eight hypothesized predictors were selected by backward and stepwise logistic regression: FAQ (informant report of functioning), UPSIT (olfactory identification), SRT immediate recall (verbal memory), MRI hippocampal volume, MRI entorhinal cortex volume. For 10% false positives (90% specificity), this five-predictor combination showed 85.2% sensitivity, combining age and MMSE showed 39.4% sensitivity, and combining age, MMSE, and the three clinical predictors (SRT immediate recall, FAQ, and UPSIT) showed 81.3% sensitivity. Area under ROC curve was greater for the five-predictor combination (0.948) than age plus MMSE (0.821; p =.0009), and remained high in sub-samples with MMSE ≥ 27/30 and amnestic MCI. For the entire patient sample, based on dichotomizing estimated risk at 0.5, positive likelihood ratio was 16.8 (95% CI 6.4, 44.3) and negative likelihood ratio was 0.2 (95% CI 0.1, 0.4).

Conclusions—The five-predictor combination strongly predicted conversion to AD and was markedly superior to combining age and MMSE. Combining only clinically administered measures also led to strong predictive accuracy. If independently replicated, the findings have potential utility for early detection of AD.

Address reprint requests to: D.P. Devanand, M.D. 1051 Riverside Drive, Unit 126, New York, NY 10032, U.S.A. Tel 212-543-5612, Fax 212-543-5088. e-mail: dpd3@columbia.edu.

Introduction

A large number of elderly people have cognitive impairment, which increases with age (1,2). Mild cognitive impairment (MCI) often represents a transition between age-related memory decline and Alzheimer's disease (AD) (3,4). The annual conversion rate from MCI to AD is 8–15%, and most conversions occur within 3 years of presentation (3,5). Identifying predictors of conversion to AD will aid in estimating prognosis and targeting early treatment in likely converters.

In a single-site, long-term study of cognitively impaired outpatients without dementia, specific baseline measures across several domains were hypothesized to predict conversion to AD. Among these measures, our original predictive findings on olfactory identification deficits (6,7), and informant report of functional deficits (8), are consistent with related studies in cognitively impaired subjects (9,10). Our published predictive accuracy of specific neuropsychological deficits (11) and MRI hippocampal and entorhinal cortex atrophy (12) are consistent with the literature (13,14). Each measure significantly predicted conversion to AD after adjusting for age, sex, and education. However, as in other studies, no individual measure was strong enough to use as the sole predictor. Therefore, the value of combining predictors, and their added value to the widely used combination of age and MMSE, was assessed. To aid clinicians in early detection, an initial algorithm to predict conversion to AD was developed.

Methods And Materials

Participants

Patients presented with memory complaints to a Memory Disorders Clinic jointly run by Psychiatry and Neurology at New York State Psychiatric Institute/Columbia University, and met study criteria for cognitive impairment without dementia and without a specific identifiable cause. These patients, and age and sex-matched healthy control subjects, were followed systematically. Healthy controls were recruited primarily by advertisement, had normative MMSE and SRT test scores, met all other inclusion/exclusion criteria, and were group-matched to patients on age and sex. All subjects signed informed consent in this IRB-approved protocol.

Eligibility Criteria

Inclusion criteria were age 41–85 years, cognitive impairment ≥ 6 months and ≤ 10 years, and Folstein MMSE score ≥ 22/30 (15). Neuropsychological screening inclusion guidelines were Folstein MMSE recall ≤ 2/3 objects at 5 minutes, or Selective Reminding Test (SRT) delayed recall score > 1 SD below norms, or Wechsler Adult Intelligence Scale-Revised (WAIS-R) performance IQ score ≥ 10 points below WAIS-R verbal IQ score. Patients without these neuropsychological deficits were considered if they met three criteria: subjective complaint of memory decline, informant's confirmation of memory decline, and modified Blessed Functional Activity Scale score ≥ 1 on the first 8 items that cover memory-related cognitive and functional problems (16). Final determination for inclusion was based on a consensus diagnosis between two expert raters (D.P.D., Y.S.) who reviewed clinical, functional and neuropsychological information, laboratory test results, and MRI clinical report, while remaining blind to all other data.

These inclusion criteria identified a broad group of cognitively impaired non-demented patients. This study began before mild cognitive impairment (MCI) criteria were published (17). Since all patients had subjective memory complaints (diagnostic criterion for MCI) and completed an extensive neuropsychological test battery, baseline MCI subtype was determined post-hoc by using age, education, and sex-based regression norms derived from 83 healthy control subjects (63 from this study and 20 from other studies) who received the same test

battery, as previously described (11). Based on this approach, 73% of patients met criteria for amnestic MCI with or without other cognitive domain deficits, 13.5% had non-amnestic MCI, and 13.5% had cognitive scores < 1.5 SD below norms, i.e., insufficient to meet MCI criteria (11).

Exclusion criteria were a diagnosis of dementia, schizophrenia, current major affective disorder, alcohol/substance dependence, history of stroke, cortical stroke or infarct ≥ 2 cm in diameter based on MRI, cognitive impairment entirely caused by medications, or other major neurological illness, e.g., Parkinson's disease.

Procedures

At baseline, the study neurologist/psychiatrist completed history, physical, neurological, and psychiatric examination. Laboratory tests included complete blood count, serum electrolytes, thyroid, liver and renal function tests, rapid plasma reagin (RPR), serum vitamin B12 and folate levels, and MRI brain scan.

All subjects received an annual neuropsychological test battery, reviewed by an experienced neuropsychologist (Y.S.). Two expert raters (D.P.D. and Y.S.) made a consensus diagnosis at each follow-up, while remaining blind to data from prior visits. The diagnosis of dementia was based on DSM-IV criteria and possible or probable AD on NINCDS-ADRDA criteria (18). The endpoint of conversion to AD required this diagnosis at two consecutive annual assessments during follow-up. The Clinical Dementia Rating (CDR) was the secondary outcome criterion. At each visit, the physician rated an initial CDR and the final CDR was scored at the consensus diagnosis meeting based on the physician's CDR plus neuropsychological test data.

Baseline predictors of future conversion to AD

At study inception, predictors were chosen based on the existing scientific literature. In this sample, among five neuropsychological variables that were examined, SRT immediate recall (12-item, 6-trial sum score) and WAIS-R digit symbol were the strongest predictors (11). Apolipoprotein E genotype was assessed and patients were classified as ε4 carriers (ε3/ε4 or ε2/ε4 or ε4/4) or ε4 non-carriers (19). The University of Pennsylvania Smell Identification Test (UPSIT) employs 40 microencapsulated common odorants, each presented with four written response alternatives (scratch and sniff forced choice format). The test score ranges 0 to 40 (20). The Pfeffer Functional Activities Questionnaire (FAQ) (21) assesses 10 common activities that focus on complex cognitive/social functioning. The informant report's total score (range 0 to 10) was obtained (8).

Brain MRI was acquired on a single GE 1.5T unit. Volumetric ratings utilized a 3D coronal volume spoiled gradient recalled echo (SPGR) sequence, 2 mm thick contiguous slices perpendicular to the temporal horns (12). After establishing intra-rater reliability and inter-rater reliability with experts (ICCs 0.90–0.99) (12), a single rater (G.P.) quantified hippocampal and entorhinal cortex volume by published methods (22), blind to all other information.

Statistical Analyses

Demographic, clinical and predictor variables were compared in converters to AD, non-converters to AD, and controls by χ^2 and Kruskal-Wallis test (several variables were non-normally distributed) for categorical and continuous variables, respectively. ANCOVA was used for group differences in hippocampal and entorhinal cortex volume, adjusting for intracranial volume.

The hypothesized predictor variables were age, SRT immediate recall, WAIS-R Digit Symbol, UPSIT, FAQ, apolipoprotein E ε4 carrier status, MRI hippocampal and entorhinal cortex volumes (adjusting for intracranial volume). Bivariate associations among these variables and baseline age and education were examined by Spearman correlation coefficients.

Since the UPSIT score had a skewed distribution, the log (41-UPSIT score) transformation (23) was applied in regression analyses.

Logistic regression analyses were conducted to assess the incremental effect of predictors (each predictor's individual effect controlled by other predictors' effects) on probability of conversion to AD by 3-year follow-up. The equal follow-up duration for all subjects permitted assessment of sensitivity, specificity, positive and negative predictive values, and likelihood ratios. The 3-year duration was chosen because most conversions occurred within this time frame, consistent with the literature (3), and to evaluate clinically relevant short-term to intermediate-term conversion to AD. For early converters who did not reach 3-year follow-up, the diagnosis was assumed to have remained AD.

From the eight hypothesized predictors, stepwise and backward procedures selected predictors for a final logistic regression model in which all remaining predictors were significantly associated with conversion to AD. For these selected predictors, sensitivity and false positive rate (1-specificity) were calculated for Receiver Operating Characteristic (ROC) curves. Area under the ROC curve was used for classification accuracy. Based on the fitted logistic model and dichotomizing estimated risk at 0.5, likelihood ratios (LR) with 95%CI were calculated for LR(+) as the ratio of true-positive percent to false-positive percent, and LR(−) as the ratio of false-negative percent to true-negative percent. The added value of the predictors to the combination of age and MMSE (age+MMSE) in predicting conversion to AD, based on logistic regression, was examined. These analyses were repeated in subsamples with amnestic MCI (11), MMSE ≥ 27/30, age > 55 years, and the 5-year follow-up sample.

From the final logistic regression model, a predictor algorithm based on dichotomizing the selected predictors was developed. The optimal dichotomized cutoff scores for the selected predictors were derived from their individual ROC curves with each cutoff point minimizing the distance between perfect and actual estimates of sensitivity and specificity. Likelihood ratios (percent of number of abnormal predictors among AD converters to percent among non-converters) and their 95% confidence intervals were calculated. For internal validation, the prediction error of this algorithm was estimated by a five-fold cross-validation procedure (24). The cross-validation procedure randomly partitioned the sample into 5 roughly equal size samples. In each of 5 runs, one of these random sub-samples was the test sample and the rest (combined 4 random sub-samples) was the training sample (24). The training sample was used to estimate parameters of the logistic regression model with the selected predictors to predict the outcome in the test sample. Combining misclassification on the test sample and averaged across the 5 runs, the cross-validation procedure gave the misclassification error rate of the predictor algorithm.

Results

Demographic and Clinical Features

Based on initial statistical power estimation for the neuropsychogical and MRI predictors derived from the literature, 150 consecutive eligible patients were recruited. Two patients diagnosed with other neurological disorders (corticobasal degeneration, and amyotrophic lateral sclerosis presenting with frontal lobe deficits) within 6 months of presentation were excluded, leaving 148 patients in the sample.

Future converters were older and scored lower on baseline MMSE than non-converters or controls, but the groups did not differ significantly in sex distribution and apolipoprotein E ε4 carrier frequency (Table 1). Baseline SRT immediate recall, WAIS-R Digit Symbol, UPSIT, hippocampal and entorhinal cortex volumes decreased from controls to non-converters to converters. FAQ informant scores were greater in converters than non-converters (controls did not have informant interviews). There were no sex differences for any predictor. Table 2 illustrates baseline correlations among the predictors and age, education and MMSE.

Of the 148 patients, 128 met the neuropsychological inclusion guidelines and 20 patients had neuropsychological deficits not severe enough to meet these guidelines but they met the other criteria (subjective memory complaints, informant's confirmation of memory decline, and Blessed rating criteria). These two groups did not differ significantly in age, sex, years of education, or conversion rates to AD. Patients who met neuropsychological inclusion guidelines had lower baseline MMSE scores compared to those who did not (mean 27.3±2.1 versus mean 28.5±2.6, $t=2.3$, $p < .03$).

Follow-up sample

In all instances when dementia was diagnosed, the CDR rating was confirmatory (≥ 1, indicating dementia), and hence the analyses were not repeated for this secondary outcome. Follow-up duration averaged 5 years in controls (Table 1). In patients, converters had shorter follow-up because they exited the study after two consecutive annual AD diagnoses (Table 1). MMSE showed greater annual percent decline in converters (mean ± SD: 3.08±5.8) compared to non-converters (0.11±2.52; $p < .0001$). Of the 39 converters, 31 had probable AD and 8 had possible AD, i.e., AD with concomitant conditions that developed during follow-up (depression n=6, prescription drug abuse n=1, Parkinsonian features without meeting criteria for Lewy Body Dementia n=1). Decline in cognitive scores during follow-up did not differ between probable and possible AD converters who were combined for data analysis. All converters were greater than 55 years old at baseline evaluation.

Of nineteen deaths during follow-up, there were four autopsies, all of which confirmed the clinical diagnoses (3 AD, 1 healthy control).

3-year follow-up sample: variable selection

Thirty-three of 126 patients converted to AD by 3-year follow-up (incidence 9.2% per person year). Twenty-two patients dropped out before the 3-year follow-up time-point. At baseline, they did not differ from the 126 patients who were followed for at least 3 years in age, education, MMSE or any hypothesized predictor (all p's > 0.3).

All eight a priori hypothesized predictors were entered into a logistic regression model with the outcome of AD/no AD by 3-year follow-up. Using either stepwise or backward selection in the model, age, WAIS-R Digit Symbol and apolipoprotein E ε4 did not meet the criterion for significance (alpha=0.1) to be in the final model. Five predictors remained significant and entered the final model: FAQ, UPSIT, SRT immediate recall, hippocampal volume, entorhinal cortex volume.

Sensitivity and specificity

In the 3-year follow-up sample, at 80% specificity (20% false positives), age, MMSE and the five predictors showed individual sensitivities ranging 41–72%. Since a low false positive rate is clinically important (a patient should not be told erroneously that he/she has AD), the individual sensitivities were also calculated at the 10% false positive rate (90% specificity). These sensitivities ranged 28–51% for age, MMSE and the five selected predictors individually (Table 3).

For the combination of age and MMSE (age+MMSE), sensitivity declined from 72.7% for 20% false positives to 39.4% for 10% false positives. For the five-predictor combination, sensitivity was 92.6% for 20% false positives and remained high at 85.2% for 10% false positives (Table 3). For 10% false positives, combining age, MMSE, and the five predictors showed 85.2% sensitivity (Table 3).

Further analyses focused on clinically relevant combinations of the five predictors, together with the widely used measures of age and MMSE. For 10% false positives, the clinically administered measures, SRT immediate recall, UPSIT, and FAQ informant report, each added markedly to the sensitivity and positive predictive value of age+MMSE (Table 3). Combining age, MMSE, SRT immediate recall, FAQ, and UPSIT showed 81.3% sensitivity and 83.3% positive predictive value. Based on estimated risk of 0.5, this combination led to a positive likelihood ratio of 13.0 and negative likelihood ratio of 0.2. Although UPSIT and FAQ informant report are brief, inexpensive and easy to administer, they are not routinely used clinically. Therefore, the combined effect of age, MMSE, SRT immediate recall, and MRI hippocampal and entorhinal cortex atrophy was evaluated. For 10% false positives, this combination showed 70.4% sensitivity and 82.6% positive predictive value, and based on estimated risk of 0.5, positive likelihood ratio of 15.5 and negative likelihood ratio of 0.3 (Table 3). The positive likelihood and negative likelihood ratios (16.8, 0.2, respectively) were strongest for the five predictors plus age and MMSE, but with fairly wide confidence intervals (Table 3).

Both hippocampal and entorhinal cortex volume statistically contributed to prediction. Removing either measure lowered sensitivity for 10% false positives to some degree: all five predictors sensitivity=85.2% (95% CI 63.4–93.8), four predictors excluding entorhinal cortex volume sensitivity=78.6% (95% CI 71.8–98.6), four predictors excluding hippocampal volume sensitivity=81.5% (95% CI 66.8–96.1).

For the five-predictor combination, in ROC analyses the area under the curve (AUC) was high (0.948) and greater than for age+MMSE (0.821; p =.0009; Figure 1 Panel A). AUC was higher for the five-predictor combination added to age+MMSE compared to age+MMSE (p=.0012, Figure 1 Panel B). The five-predictor combination also led to strong prediction in amnestic MCI patients (32/90 converters by 3 years, AUC 0.934), patients with baseline MMSE ≥ 27/30 (19/92 converters by 3 years, AUC 0.941), patients followed for 5 years (37/90 converters, AUC 0.938), and patients over 55 years old (33/112 converters by 3 years, AUC 0.942).

Including education with age and MMSE in similar models, and replacing SRT immediate recall by SRT delayed recall, led to virtually identical results.

Predictor algorithm

Using the 3-year follow-up sample, a predictor algorithm from the logistic regression model was developed. The optimal dichotomized cutoff scores for the 5 predictors were derived from their individual ROC curves with each cutoff point minimizing the distance between perfect and actual estimates of sensitivity and specificity. Dichotomized cutoffs for “abnormality” were SRT immediate recall < 40, UPSIT < 32, FAQ > 1, hippocampal volume < 4.1 cc, entorhinal cortex volume < 0.4 cc (Table 4). These dichotomized variables were then entered in the logistic model, and the estimated risk of AD conversion tabulated for every predictor combination.

In the five-predictor algorithm, the 3-year probability of conversion was very low if only one predictor was abnormal (1.4 to 3%) and low (7.5 to 24%) if only two of five predictors were abnormal. For three abnormal predictors the probability of conversion ranged 34–67%, for four abnormal predictors this probability ranged 84–92%, and if all five predictors were abnormal

this probability was 98% (Table 4). The likelihood ratio for predicting conversion to AD increased with increasing number of abnormal predictors, particularly for three or more abnormal predictors (Table 4).

For internal validation of the five-predictor algorithm, the five-fold cross-validation procedure estimated the misclassification error of the predictor algorithm as 14.95%.

Discussion

In this broadly defined clinical sample of cognitively impaired, non-demented patients presenting with memory complaints, each predictor's effect was consistent with other studies that examined predictors of MCI conversion to AD (25,26), with the exception of apolipoprotein E ε4 carrier status that was not significant when included with other predictors in logistic regression analyses. Sampling differences may account for discrepancies across studies; apolipoprotein E ε4 carrier status is a significant predictor in several studies (25–28) but some recent reports indicate weak prediction (29,30).

In the 3-year follow-up sample, individual baseline measures significantly predicted future conversion to AD but the five-predictor combination markedly improved prediction. The five-predictor combination led to very strong prediction at the MCI stage, exceeding consensus threshold criteria for markers of AD of sensitivity >80% and specificity > 80% (31). The five-predictor combination markedly added predictive value above the combination of age and MMSE at a 10% false positive rate, and positive and negative likelihood ratios confirmed strong prediction. A low false positive rate is necessary when predicting likely conversion to AD, i.e., predicting AD when the patient does not develop AD can lead to adverse clinical consequences. The findings indicate that knowing the patient's age and administering the MMSE, a combination commonly used by practicing physicians, are together insufficient to accurately predict conversion to AD.

The diagnosis of MCI requires lack of impairment in activities of daily living (17). This is the first study to explicitly examine informant report of impairment in complex social/cognitive abilities, which can be impaired in MCI patients, and odor identification deficits together as predictors of conversion to AD. These inexpensive measures do not require prior training to administer: FAQ 2–3 minutes (can be self-report), UPSIT 12–15 min (multiple choice, can be self-administered). In the predictor models, for 10% false positives the combination of SRT immediate recall, FAQ and UPSIT, together with age and MMSE, led to strong predictive accuracy.

The combination of SRT immediate recall with age, MMSE and the MRI variables also led to fairly strong prediction. Neuropsychological testing of episodic verbal memory and other cognitive abilities is time-consuming and can be expensive. MRI volumetric evaluation of hippocampal and entorhinal cortex volumes is expensive and currently limited to research/academic centers. Nonetheless, neuropsychological and MRI assessment improve prediction of conversion to AD, provide other useful diagnostic information, and their serial assessment may clarify the MCI/AD diagnosis. A trained MRI rater, after completing image alignment, can assess hippocampal volume in 30–45 minutes and entorhinal cortex volume in 15–20 minutes using this study's methods (22,32). Since MRI hippocampal and entorhinal cortex volumetric assessment techniques vary by research group, their identified algorithm cutoffs may not apply for other assessment methods. In contrast, the SRT, UPSIT and FAQ employ uniform administration and scoring.

In other studies, combining MMSE/verbal recall measures with MRI hippocampal volume/medial temporal lobe atrophy or apolipoprotein E genotype led to moderately strong accuracy (65–84% correctly classified) in predicting conversion to AD (25,26). We found that adding

olfactory identification deficits and informant report of functional impairment, and entorhinal cortex volume, led to strong predictive accuracy (92.5% correctly classified) that was appreciably higher than prior reports (25,26). The predictor algorithm developed from the data obtained in our study emphasizes the need to examine multiple predictors.

For the five-predictor combination, similar findings were obtained in the amnestic MCI, high baseline MMSE, and 5-year follow-up sub-samples. Strong prediction in the high MMSE (≥ 27/30) group is clinically important because diagnosis and estimation of prognosis is difficult in patients with minimal global cognitive impairment.

The individual effects of olfactory identification deficits and informant report of functional impairment were consistent with related reports (9,10), and the combined effects of episodic memory deficits and MRI volumetric measures were highly consistent with the literature (14,32). For the algorithm, to reduce the risk of over-fitting, the cutoff scores for the 5 predictors were identified from their individual ROC curves and not from their combined ROC curve. Since SRT immediate recall, UPSIT and hippocampal volume declined with increasing age (5,22,33), and SRT immediate recall correlated with education in years, the optimal algorithm cutoffs for the five predictor variables may need alteration when applied in older, less educated samples. A complete independent external validation in a larger sample is needed, but to date no other study has evaluated informant report of functional impairment and odor identification deficits together with neuropsychological and MRI volumetric measures. Further, using the five-predictor combination does not preclude the need for proper subsequent verification of diagnostic classification.

The MRI, olfactory and verbal recall measures tapped into early AD pathology: neurofibrillary tangles begin in trans-entorhinal cortex and progress through limbic areas to neocortex (34). Impairment in complex, integrated brain functions were sensitive to incipient AD: short + long-term verbal memory (SRT immediate recall incorporates encoding with recognition, comparison to a bank of memories, retrieval over 6 consecutive trials), odor identification (UPSIT incorporates odor detection, odor memory, odor naming) and complex cognitive/social functional abilities (FAQ informant report captures the behavioral output of multiple brain regions acting together).

Gender, obesity, vascular risk factors, and family history predict dementia in large-scale epidemiologic studies. Given the known small effect size of these variables, they were not hypothesized predictors in this clinical sample. In secondary analyses, we examined gender, body mass index and family history, and none were significant predictors (data not shown).

This study had some limitations. The study was conducted in a University academic medical center; the results need validation in broader clinical settings and cannot be directly applied to the general elderly population. The sample inclusion/exclusion criteria were broader than the narrowly defined amnestic MCI samples studied by other groups (3,4), but they did identify a clinically relevant, heterogenous group of patients. The findings in our large amnestic MCI subsample were consistent with the results obtained in the entire sample, thus permitting direct comparison to other studies that restricted their samples to amnestic MCI patients (3,4). The raters were not blind to one of the selected predictors, SRT immediate recall, which was part of the neuropsychological test battery. However, the expert raters employed strict diagnostic methods, and the diagnoses were indirectly supported by differences in global cognitive decline in converters versus non-converters. Patients with stroke or MRI cortical or large lacunar infarctions were excluded, reducing the chance of cerebrovascular disease influencing MCI conversion to AD (35).

Other measures that strongly distinguish AD from controls and potentially predict MCI conversion to AD include decreased temporoparietal metabolism with FDG PET (36) and

increased uptake of ligands that label amyloid using PET (37,38). PET has high costs and limited availability. PET amyloid imaging may not show disease progression because maximal brain amyloid saturation occurs early in AD (39). Cerebrospinal fluid (CSF) elevated tau or phosphorylated tau levels, and decreased amyloid ßeta$_{42}$ levels, and their ratio, have been shown to strongly predict MCI conversion to AD (40). Future studies will need to assess these CSF measures when combined with other predictors. Lack of patient acceptability of lumbar puncture can be a limitation. A recent report showed that combining immune and related plasma markers strongly separated AD from controls and predicted MCI conversion to AD in a small MCI sample, but independent replication in a larger sample is needed (41).

In our study, a combination of five measures led to strong predictive accuracy for MCI conversion to AD, and markedly added value to the combination of age and MMSE. These findings clearly need independent replication in larger samples before widespread clinical use is justified. Another potential use is to enrich patient selection in treatment trials.

Acknowledgements

Dr. Devanand had full access to all of the data in this study and takes responsibility for the integrity of the data and accuracy of the data analysis. This study was supported by National Institute of Aging Grant No. R01AG17761; by Grants R01AG12101, P50 AG08702, and P30 AG08051 from the National Institute of Aging; and by Grant Nos. MH55735, MH35636, and MH55646 from the National Institute of Mental Health. This study was presented in part at the 43rd Annual American College of Neuropsychopharmacology Meeting, December 12, 2004; San Juan, Puerto Rico.

Dr. Richard Mayeux's contribution to the interpretation of data is acknowledged.

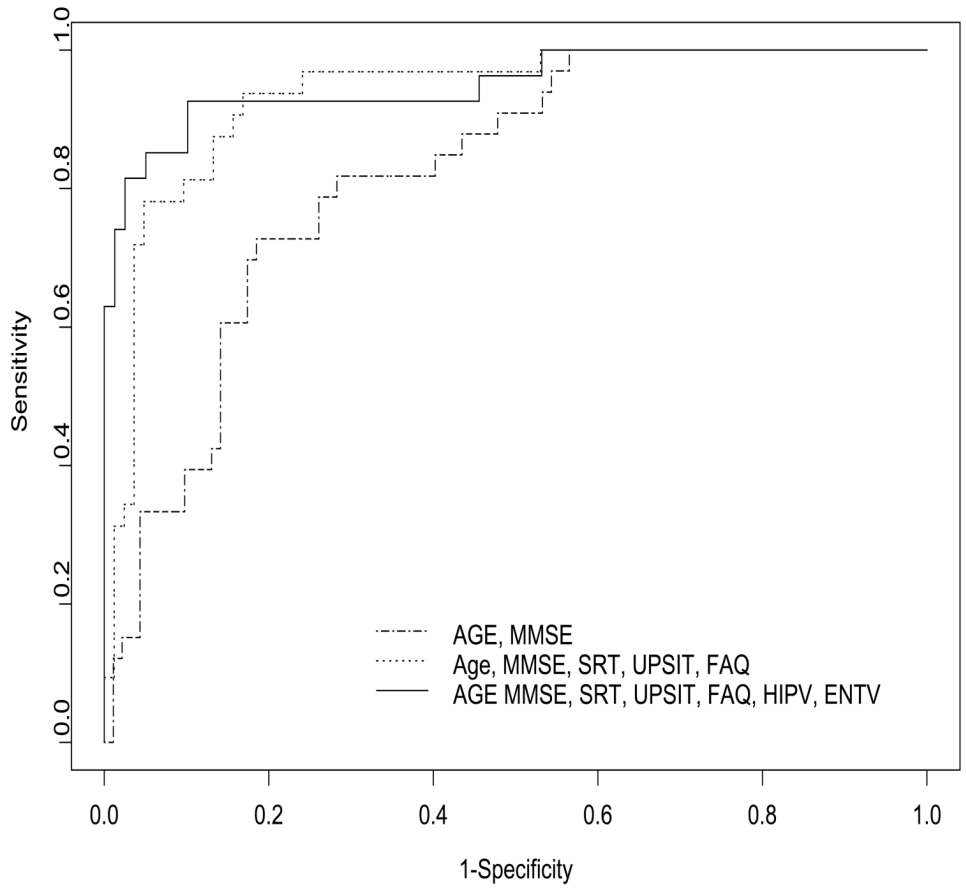
Dr. Doty is a major shareholder in Sensonics, Inc., the manufacturer of the University of Pennsylvania Smell Identification Test. Dr. Devanand received research funding from Eli Lilly, Novartis, and Forest Labs, and consultancy fees from GlaxoSmithKline. Dr. Tabert is now a full-time employee of International Flavors and Fragrances. Dr. Bell is a speaker for Pfizer/Eisai, Novartis, and Bayer, GE, and Merck. Dr. Stern received research funding from Elan, Wyeth, and Eli Lilly, and consultancy fees from Elan, Wyeth, and Bayer. Dr. Pelton received research funding from Forest and Novartis. Drs. Liu, Pradhaban, and Cuasay report no biomedical financial interests or potential conflicts of interest.

References

1. Di Carlo A, Lamassa M, Baldereschi M, et al. CIND and MCI in the Italian elderly: frequency, vascular risk factors, progression to dementia. Neurology 2007;68:1909–1916. [PubMed: 17536047]
2. Larrieu S, Letenneur L, Orgogozo JM, et al. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. Neurology 2002;59:1594–1599. [PubMed: 12451203]
3. Petersen RC. Mild cognitive impairment: current research and clinical implications. Semin Neurol 2007;27:22–31. [PubMed: 17226738]
4. Lopez OL, Jagust WJ, DeKosky ST, et al. Prevalence and classification of mild cognitive impairment in the Cardiovascular Health Study Cognition Study. Part 1. Arch Neurol 2003;60:1385–1389. [PubMed: 14568808]
5. Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. J Intern Med 2004;256:240–246. [PubMed: 15324367]
6. Devanand DP, Michaels-Marston KS, Liu X, et al. Olfactory deficits in patients with mild cognitive impairment predict Alzheimer's disease at follow-up. Am J Psychiatry 2000;157:1399–1405. [PubMed: 10964854]
7. Tabert MH, Liu X, Doty RL, et al. A 10-item smell identification scale related to risk for Alzheimer's disease. Ann Neurol 2005;58:155–160. [PubMed: 15984022]
8. Tabert MH, Albert SM, Borukhova-Milov L, et al. Functional deficits in patients with mild cognitive impairment: prediction of AD. Neurology 2002;58:758–764. [PubMed: 11889240]

9. Graves AB, Bowen JD, Rajaram L, et al. Impaired olfaction as a marker for cognitive decline: interaction with apolipoprotein E epsilon4 status. Neurology 1999;53:1480–1487. [PubMed: 10534255]
10. Portet F, Ousset PJ, Visser PJ, et al. Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease. J Neurol Neurosurg Psychiatry 2006;77:714–718. [PubMed: 16549412]
11. Tabert MH, Manly JJ, Liu X, et al. Neuropsychological Prediction of Conversion to Alzheimer's Disease in Patients with Mild Cognitive Impairment. Arch Gen Psychiatry 2006;63:916–924. [PubMed: 16894068]
12. Devanand DP, Pradhaban G, Khandji A, et al. Hippocampal and entorhinal atrophy in the prediction of MCI conversion to Alzheimer's disease. Neurology 2007;68:828–836. [PubMed: 17353470]
13. Fleisher AS, Sowell BB, Taylor C, et al. Clinical predictors of progression to Alzheimer disease in amnestic mild cognitive impairment. Neurology 2007;68:1588–1595. [PubMed: 17287448]
14. Stoub TR, Bulgakova M, Leurgans S, et al. MRI predictors of risk of incident Alzheimer disease: a longitudinal study. Neurology 2005;64:1520–1524. [PubMed: 15883311]
15. 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;12:189–198. [PubMed: 1202204]
16. Stern Y, Hesdorffer D, Sano M, Mayeux R. Measurement and prediction of functional capacity in Alzheimer's disease. Neurology 1990;40:8–14. [PubMed: 2296387]
17. Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–308. [PubMed: 10190820]
18. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer Disease. Neurology 1984;34:939–944. [PubMed: 6610841]
19. Devanand DP, Pelton GH, Zamora D, et al. Predictive utility of apolipoprotein E genotype for Alzheimer disease in outpatients with mild cognitive impairment. Arch Neurol 2005;62:975–980. [PubMed: 15956169]
20. Doty RL, Shaman P, Dann M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. Physiology and Behavior (Monograph) 1984;32:489–502.
21. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982;37:323–329. [PubMed: 7069156]
22. De Santi S, de Leon MJ, Rusinek H, et al. Hippocampal formation glucose metabolism and volume losses in MCI and AD. Neurobiol Aging 2001;22:529–539. [PubMed: 11445252]
23. Frye RE, Schwartz BS, Doty R. Dose-Related Effects of Cigarette Smoking on Olfactory Function. JAMA 1990;263:1233–1236. [PubMed: 2304239]
24. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning. Springer; New York: 2001.
25. Tierney MC, Szalai JP, Snow WG, et al. A prospective study of the clinical utility of ApoE genotype in the prediction of outcome in patients with memory impairment. Neurology 1996;46:149–154. [PubMed: 8559365]
26. Visser PJ, Verhey FR, Scheltens P, et al. Diagnostic accuracy of the Preclinical AD Scale (PAS) in cognitively mildly impaired subjects. J Neurol 2002;249:312–319. [PubMed: 11993532]
27. Mayeux R, Saunders AM, Shea S, et al. Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. Alzheimer's Disease Centers Consortium on Apolipoprotein E and Alzheimer's Disease. N Engl J Med 1998;338:506–511. [PubMed: 9468467]
28. Tervo S, Kivipelto M, Hänninen T, et al. Incidence and risk factors for mild cognitive impairment: a population-based three-year follow-up study of cognitively healthy elderly subjects. Dement Geriatr Cogn Disord 2004;17(3):196–203. [PubMed: 14739544]
29. Blacker D, Lee H, Muzikansky A, et al. Neuropsychological measures in normal individuals that predict subsequent cognitive decline. Arch Neurol 2007;64:862–871. [PubMed: 17562935]
30. Tyas SL, Salazar JC, Snowdon DA, et al. Transitions to mild cognitive impairments, dementia, and death: findings from the Nun Study. Am J Epidemiol 2007;165:1231–8. [PubMed: 17431012]

31. Frank RA, Galasko D, Hampel H, et al. National Institute on Aging Biological Markers Working Group. Biological markers for therapeutic trials in Alzheimer's disease. Proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease. Neurobiol Aging 2003;24:521–536. [PubMed: 12714109]
32. Killiany RJ, Hyman BT, Gomez-Isla T, et al. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. Neurology 2002;58:1188–1196. [PubMed: 11971085]
33. Doty RL, Shaman P, Applebaum SL, Giberson R, Sikorsky L, Rosenberg L. Smell identification ability: changes with age. Science 1984;226:1441–1443. [PubMed: 6505700]
34. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. Acta Neuropathol (Berl) 1991;82:239–259. [PubMed: 1759558]
35. van de Pol LA, Korf ES, van der Flier WM, et al. Magnetic resonance imaging predictors of cognition in mild cognitive impairment. Arch Neurol 2007;64:1023–1028. [PubMed: 17620494]
36. Silverman DH, Small GW, Chang CY, et al. Positron emission tomography in evaluation of dementia: Regional brain metabolism and long-term outcome. JAMA 2001;286:2120–2127. [PubMed: 11694153]
37. Small GW, Kepe V, Ercoli LM, et al. PET of brain amyloid and tau in mild cognitive impairment. N Engl J Med 2006;355:2652–2663. [PubMed: 17182990]
38. Klunk WE, Engler H, Nordberg A, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol 2004;55:306–319. [PubMed: 14991808]
39. Engler H, Forsberg A, Almkvist O, Blomquist G, Larsson E, Savitcheva I, Wall A, Ringheim A, Långström B, Nordberg A. Two-year follow-up of amyloid deposition in patients with Alzheimer's disease. Brain 2006;129:2856–66. [PubMed: 16854944]
40. Hansson O, Zetterberg H, Bucchave P, et al. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. Lancet Neurology 2006;5:228–234. [PubMed: 16488378]
41. Ray S, Britschgi M, Herbert C, et al. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. Nat Med 2007;13(11):1359–62. [PubMed: 17934472]



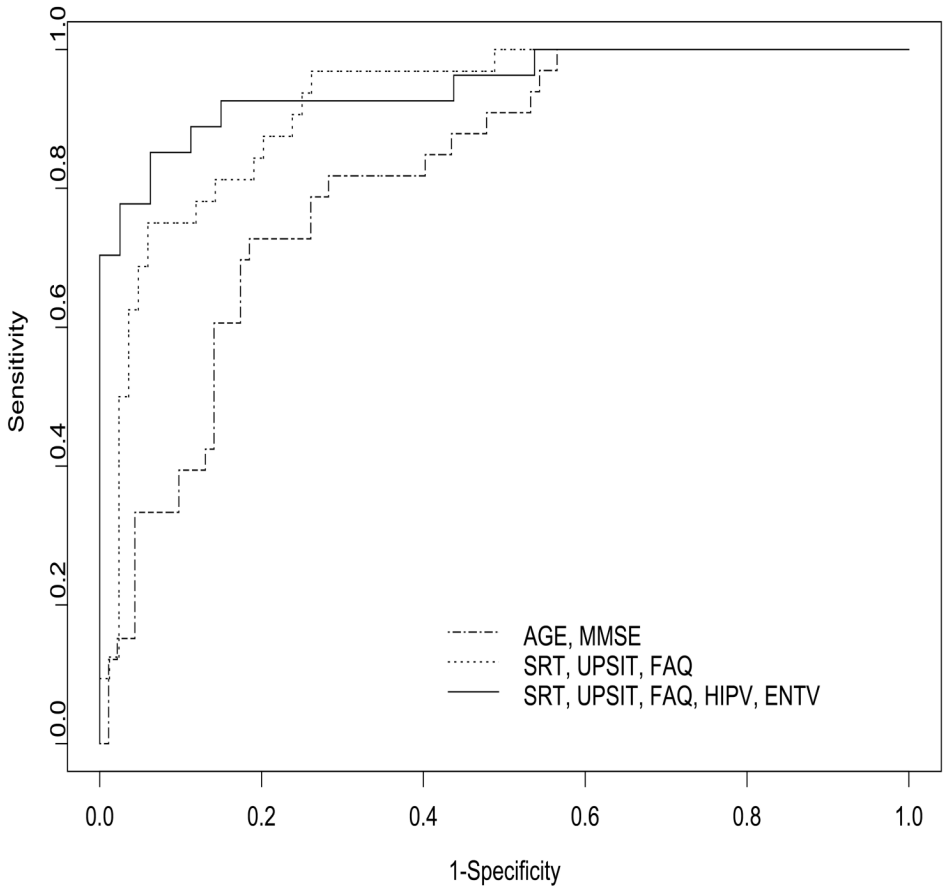


Figure 1.

Comparisons of Receiver Operating Characteristic (ROC) curves in the 3-year follow-up patient sample.

Panel A. Comparison of ROC curves for Age + MMSE (AUC=0.821), SRT + UPSIT + FAQ (AUC=0.920), and the 5 selected predictors: SRT + UPSIT + FAQ + Hippocampal volume (HIPV) + Entorhinal volume (ENTV) (AUC=0.948).

Panel B. Comparison of ROC curves for Age + MMSE (AUC=0.821), Age + MMSE + SRT + UPSIT + FAQ (AUC=0.934), and Age + MMSE + SRT + UPSIT + FAQ + Hippocampal volume (HIPV) + Entorhinal volume (ENTV) (AUC=0.951).

Table 1

Baseline features of 148 MCI patients (converters and non-converters to AD during 1–9 years of follow-up) and 63 healthy control subjects.

Feature	Patients		Healthy Controls	*P value
	39 Converters mean (SD) or %	109 Non-converters mean (SD) or %	63 controls mean (SD) or %	
Sex % female	56.4	55.1	54.0	.971
Baseline Age (years)	73.2 (7.1)	64.9 (9.9)	65.7 (9.3)	<.0001
Education in years	14.0 (4.7)	15.4 (4.1)	16.7 (2.6)	.0014
+Follow-up duration in months	41.5 (18.5)	57.3 (28.3)	62.1 (31.5)	.001
Baseline MMSE	26.3 (2.2)	27.9 (2.0)	29.4 (0.8)	<.0001
Apolipoprotein E ε4 carrier %	34.3	23.6	22.4	.3826
SRT Immediate Recall	34.8 (8.5)	45.2 (8.0)	52.8 (6.6)	<.0001
WAIS-R Digit Symbol	32.5 (10.2)	43.3 (11.6)	50.3 (11.3)	<.0001
UPSIT	25.8 (8.4)	33.2 (4.6)	34.8 (4.2)	<.0001
FAQ	2.76 (2.1)	1.23 (1.84)	----------	<.0001
**Hippocampal volume (R^+ L)	3.77 (0.72)	4.37 (0.62)	4.34 (0.57)	<.0001
**Entorhinal cortex volume (R^+ L)	0.381 (0.093)	0.465 (0.085)	0.548 (0.096)	<.0001

All values are means ± standard deviations, or percentages.

MMSE: Folstein Mini-Mental State Exam, range 0–30.

SRT: Selective Reminding Test, 12-item, 6-trial version.

UPSIT: 40-item University of Pennsylvania Smell Identification Test, range 0–40.

FAQ: Pfeffer Functional Activities Questionnaire score, administered to informant, range 0–10.

+Converters to AD exited the study after 2 consecutive annual AD diagnoses, thereby reducing the follow-up duration in this group.

*Based on $\chi 2$ test or Kruskal-Wallis Test for three-group comparisons in categorical and quantitative variables.

**Hippocampal and entorhinal cortex volumes are in cubic cm, and three-group differences were tested with ANCOVA adjusting for intracranial volume.

Among the 148 patients, missing baseline data for specific measures were as follows. SRT: one patient, UPSIT: one patient was anosmic and one had an upper respiratory infection leading to exclusion, FAQ: 7 patients did not have informant interviews, MRI: 8 patients did not have an MRI scan and one patient had entorhinal cortex volume that could not be rated (poor quality scan), and 8 patients did not have blood drawn for apolipoprotein E genotype.

Table 2

Baseline correlations in MCI patients among age, education, MMSE and specific hypothesized predictors.

Predictor	Age in years r	Education in years r	MMSE r
SRT immediate recall	−0.44 p < .0001	0.31 p < .0002	0.47 p < .0001
WAIS-R Digit Symbol	−0.35 p < .0001	0.53 p < .0001	0.51 p < .0001
UPSIT	−0.50 p < .0001	0.15 p =.079	0.45 p < .0001
FAQ informant	0.02 p = 0.85	−0.15 p=.078	−0.22 p < .01
Hippocampal Volume	−0.32 p < .0001	0.17 p < .05	0.32 p < .0001
Entorhinal cortex volume	−0.25 p < .003	0.14 p = .11	0.09 p = .27

r=Spearman correlation coefficient.

Number of patients with available data for specific variables: age=148, education=148, MMSE=147, SRT=147, WAIS-R Digit Symbol=147, UPSIT=145, FAQ=140, Hippocampal volume=140, Entorhinal cortex volume=139.

Apolipoprotein E carrier status (categorical variable) obtained in 140 patients was not included in these analyses.

Table 3

Effects of individual and combined predictors for conversion to Alzheimer's Disease (AD) in the 3-year follow-up patient sample.

Predictors	Converters/Total n	Logistic Regression P-value*	AUC	Sensitivity % at specificity =90% (95% CI)	Correct classification %	Positive predictive value %	Negative predictive value %	Likelihood ratio LR(+) 95% CI	Likelihood ratio LR(−) 95% CI
Age	33/126	0.0002	0.739	27.7	73.0	46.7	76.6	2.5 (1.0, 6.3)	0.9 (0.7, 1.0)
MMSE	33/125	0.0001	0.778	26.8	76.0	61.5	77.7	4.5 (1.6, 12.7)	0.8 (0.7. 1.0)
SRT Imm Rec	32/125	<0.0001	0.849	50.9	80.0	66.7	82.7	5.8 (2.6, 13.1)	0.6 (0.4, 0.8)
UPSIT	33/124	<0.0001	0.833	48.5	81.4	72.7	83.3	7.4 (3.1, 17.2)	0.6 (0.4, 0.8)
FAQ	33/118	0.0012	0.708	32.8	75.4	62.5	77.4	4.3 (1.7, 10.9)	0.7 (0.6, 0.9)
HIPP	29/118	<0.0001	0.753	41.4	80.5	75.0	81.1	9.2 (2.7, 31.7)	0.7 (0.6, 0.9)
ENTO	28/117	<0.0001	0.773	50.0	80.3	72.7	81.1	8.5 (2.4, 29.8)	0.7 (0.6, 0.9)
Age MMSE	33/125	0.0006 0.0005	0.821	39.4 (22.7, 56.1)	76.0	56.5	80.4	3.6 (1.8, 7.5)	0.7 (0.5, 0.9)
Age MMSE FAQ	33/117	0.0001 0.0006 0.0010	0.873	60.6 (43.9, 77.3)	83.8	76.9	85.7	8.5 (3.7, 19.2)	0.4 (0.3, 0.6)
Age MMSE UPSIT	33/123	0.0936 0.0314 0.0020	0.870	60.6 (43.9, 77.3)	82.1	69.0	86.2	6.1 (3.1, 11.9)	0.4 (0.3, 0.7)
Age MMSE SRT Imm Rec	32/124	0.0170 0.0394 <0.0001	0.887	71.9 (56.3, 87.4)	85.5	76.9	87.8	9.6 (4.2, 21.7)	0.4 (0.3, 0.6)
Age MMSE UPSIT FAQ	33/116	0.0229 0.0236 0.0029 0.0021	0.910	72.7 (57.5, 87.9)	85.3	76.7	88.4	8.3 (3.9, 17.4)	0.3 (0.2, 0.6)
Age MMSE SRT Imm Rec UPSIT FAQ	32/115	0.1145 0.4315 0.0028 0.0145 0.0047	0.934	81.3 (67.7, 94.8)	89.6	83.3	91.8	13.0 (5.4, 30.9)	0.2 (0.1, 0.4)
Age MMSE SRT Imm Rec HIPP ENTO	27/115	0.9874 0.3962 0.0008 0.0124 0.0028	0.913	70.4 (53.2, 87.6)	89.6	82.6	91.3	15.5 (5.8, 41.6)	0.3 (0.2, 0.6)
AGE MMSE SRT Imm Rec UPSIT FAQ HIPP ENTO	27/106	0.7555 0.6538 0.0123 0.0122 0.0293 0.0324 0.0041	0.951	85.2 (71.9, 98.5)	92.5	85.2	94.9	16.8 (6.4, 44.3)	0.2 (0.1, 0.4)
SRT Imm Rec UPSIT FAQ HIPP ENTO	27/107	0.0054 0.0095 0.0283 0.0246 0.0039	0.948	85.2 (71.8, 98.6)	89.7	80.8	92.6	12.4 (5.2, 29.8)	0.2 (0.1, 0.5)

* P-value was based on testing the null hypothesis of zero coefficient of the predictor in logistic regression for conversion by 3 years of follow-up. CI: confidence interval Area under the curve (AUC) was obtained from Receiver Operating Characteristic (ROC) analyses for individual and specified combinations of predictors evaluated as continuous measures.

Percent correct classification used a threshold of 0.5 on predicted risk derived from the logistic regression models.

Based on dichotomizing predicted risk at 0.5, LR(+) is the ratio of true-positive percent to false-positive percent and LR(−) is the ratio of false-negative percent to true-negative percent.

MMSE: 30-item Folstein Mini-Mental State Exam. SRT: Selective Reminding Test. UPSIT: University of Pennsylvania Smell Identification Test.

FAQ: Pfeffer Functional Activities Questionnaire.

HIPP: Hippocampal volume (right+left) in cc. ENTO: Entorhinal volume (right+left) in cc.

Table 4

Probability of MCI conversion to AD in the 3-year follow-up sample (25.2% converters) by the number of abnormal predictors in the predictor algorithm.

Number of abnormal predictors	SRT Immed. Recall <40	UPSIT <32	FAQ >1	Hippocampal Volume <4.1 cc	Entorhinal Volume <0.4 cc	Probability of conversion to AD Percent	Likelihood Ratio (95% CI)
0	No	No	No	No	No	***0.29***	0.2 (0, 1.1)
1–2	Yes	No	No	No	No	1.44	
	No	No	No	Yes	No	1.57	0.27 (0.07,1.07)
	No	No	Yes	No	No	1.82	
	No	No	No	No	Yes	2.93	
	No	Yes	No	No	No	3.00	
	Yes	No	No	Yes	No	7.45	
	Yes	No	Yes	No	No	8.57	
	No	No	Yes	Yes	No	9.29	
	Yes	No	No	No	Yes	13.20	
	Yes	Yes	No	No	No	13.48	
	No	No	No	Yes	Yes	14.25	
	No	Yes	No	Yes	No	14.55	
	No	No	Yes	No	Yes	16.23	
	No	Yes	Yes	No	No	16.56	
	No	Yes	No	No	Yes	***24.35***	
3	Yes	No	Yes	Yes	No	34.09	
	Yes	No	No	Yes	Yes	45.62	3.62 (1.68, 7.79)
	Yes	Yes	No	Yes	No	46.22	
	Yes	No	Yes	No	Yes	49.44	
	Yes	Yes	Yes	No	No	50.04	
	No	No	Yes	Yes	Yes	51.65	
	No	Yes	Yes	Yes	No	52.25	
	Yes	Yes	No	No	Yes	61.91	
	No	Yes	No	Yes	Yes	63.97	
	No	Yes	Yes	No	Yes	***67.42***	
4	Yes	No	Yes	Yes	Yes	84.36	29.63 (3.97, 220.88)
	Yes	Yes	Yes	Yes	No	84.67	
	Yes	Yes	No	Yes	Yes	89.96	
	Yes	Yes	Yes	No	Yes	91.26	
	No	Yes	Yes	Yes	Yes	91.94	
5	Yes	Yes	Yes	Yes	Yes	***98.29***	∞

To develop this model, weighted sum and anti-logit transformation were calculated. The weights were the regression coefficients associated with predictors in the logistic models.

The probability/risk was derived from the following logistic regression model equation:

logit P(AD conversion within 3 years)= −5.8463 + 1.6189(SRT immediate recall < 40) + 2.3685(UPSIT < 32) + 1.8605(FAQ > 1) + 1.7074(Hippocampal volume < 4.1) + 2.3444 (Entorhinal volume < 0.4).

For each of the 5 predictors, the subject's data meets the cutoff criterion for abnormality specified for each predictor (value=1) or does not meet the cutoff criterion for abnormality (value=0). Each row indicates the estimated probability/risk of conversion to AD based on which predictor(s) meets the cutoff criterion.

Likelihood ratio: ratio of percent of number of abnormal predictors in converters to percent in non-converters. For the algorithm, sensitivity=0.82 and specificity=0.93 if subjects are classified by dichotomizing risk at 0.5.