Diagnostic Assessment & Prognosis

Practical algorithms for amyloid β probability in subjective or mild cognitive impairment

Nancy Maserejian[a,1,*], Shijia Bian[b,1], Wenting Wang[b], Judith Jaeger[c,d], Jeremy A. Syrjanen[e], Jeremiah Aakre[e], Clifford R. Jack, Jr.,[f], Michelle M. Mielke[g], Feng Gao[b], for the Alzheimer's Disease Neuroimaging Initiative and the AIBL research team[2]

[a]Department of Epidemiology, Biogen, Cambridge, MA, USA
[b]Department of Biostatistics, Biogen, Cambridge, MA, USA
[c]CognitionMetrics, Wilmington, DE, USA
[d]Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, NY, USA
[e]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
[f]Department of Radiology, Mayo Clinic, Rochester, MN, USA
[g]Department of Health Sciences Research, Professor of Neurology, Mayo Clinic, Rochester, MN, USA

Abstract

Introduction: Practical algorithms predicting the probability of amyloid pathology among patients with subjective cognitive decline or mild cognitive impairment may help clinical decisions regarding confirmatory biomarker testing for Alzheimer's disease.

Methods: Algorithm feature selection was conducted with Alzheimer's Disease Neuroimaging Initiative and Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing data. Probability algorithms were developed in Alzheimer's Disease Neuroimaging Initiative using nested cross-validation accompanied by stratified subsampling to obtain 1000 internally validated decision trees. Semi-independent validation was conducted using Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing. Independent external validation was conducted in the population-based Mayo Clinic Study of Aging.

Results: Two algorithms were developed using age and normalized immediate recall z-scores, with or without apolipoprotein E ε4 carrier status. Both algorithms had robust performance across data sets and when substituting different recall memory tests.

Discussion: The statistical framework resulted in robust probability estimation. Application of these algorithms may assist in clinical decision-making for further testing to diagnose amyloid pathology.
© 2019 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: ADNI; AIBL; MCSA; Algorithm; Amyloid; Alzheimer's disease; Biomarker; APOE ε4; Immediate recall

1. Introduction

Alzheimer's disease (AD) dementia is a chronic neurode-generative disorder that is both progressive and irreversible [1,2]. Accumulation of brain amyloid beta (Aβ) and tau pathology are defining characteristics of the AD continuum and occur decades before cognitive symptoms are present [1,3–5]. Early intervention to alter the underlying Aβ or tau pathology is considered a potential approach to prevent or delay AD progression, and such treatments are in development [6–9].

Although biomarkers for Aβ and tau pathology are often used to diagnose AD in research settings, these biomarkers are not typically used to diagnose AD in routine clinical practice today, primarily owing to resource limitations and costs [1]. If a new therapy targeting AD pathology were to become available, methods to confirm the presence of AD pathology, including positron emission tomography (PET) imaging—the only US Food and Drug Administration–approved biomarker for AD—and cerebrospinal fluid (CSF) measures, are projected to remain inaccessible to many patients [1,8,9]. As such, practical methods to determine which patients are most likely to benefit from more invasive and costly confirmatory biomarker testing for the presence of AD pathology may be helpful for prioritizing potentially limited resources.

The objective of this work was to provide practical algorithms to estimate the probability that a patient exhibiting cognitive problems possibly due to AD is Aβ positive, using currently available inputs. Prior research has identified factors that are associated with Aβ pathology, such as age, cognitive impairment, apolipoprotein E (APOE) genotype, CSF inflammatory or protein biomarkers [10], and certain lifestyle factors [11–14]. However, risk factor models do not directly translate into clinically useful or practical algorithms. Many models lack external validation, include inputs with small effect sizes, or include inputs that are burdensome or costly (e.g., extensive neuropsychological testing or imaging) [10,15–17]. Recently, more practical algorithms to estimate the likelihood of Aβ positivity among patients with subjective cognitive decline (SCD) or mild cognitive impairment (MCI) were published [11,18,19]. For example, the Swedish BioFINDER study's "optimal" model used data on age, APOE genotype, and delayed recall score [18]. Although performance, as measured by area under the curve (AUC), has been acceptable in these reports, the algorithms published to date have limited flexibility because they require the input of APOE genotype and a specified cognitive test. Moreover, the data sets are composed of patients from highly specialized clinics, and it is unknown whether the performance would remain robust in a broader population of symptomatic individuals.

Our intention was to develop algorithms that would support clinical decision-making regarding future biomarker testing, while also allowing quick administration and flexible inputs, such that providers could select their preferred cognitive measures and use of genetic information. We anticipated that algorithms using currently available inputs would not replace Aβ tests but rather allow providers to more efficiently and confidently send symptomatic patients for more invasive and costly Aβ testing, if needed for diagnosis and treatment planning. To cast a broad net and improve power to detect predictors, we first included all nondemented participants across two data sets in the analysis to identify predictors; in the next phase of deriving probability estimates, we focused on symptomatic patients owing to the current clinical context in which symptoms are ascertained before considering pathology. Given that there are scenarios in which genetic testing is not conducted, we designed two versions: one utilizing APOE ε4 data and another without it. To achieve robust and generalizable algorithms, we developed a multistage statistical framework, using a combination of epidemiologic and random forest decision tree modeling methods, with an independent external validation using a community population-based sample.

2. Methods

We developed a statistical framework and multiphase approach to develop and validate the algorithms using 3 data sources: the Alzheimer's Disease Neuroimaging Initiative (ADNI), Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing (AIBL), and Mayo Clinic Study of Aging (MCSA). The analysis phases were (1) initial feature (i.e., variable) selection in ADNI and AIBL, (2) deep development of probability algorithms in ADNI, (3) semi-independent validation in AIBL, and (4) independent external validation in the population-based MCSA. The goal was to develop an algorithm that was not over-fit to a particular sample of individuals, and then test it in an entirely different sample of individuals with SCD or MCI, achieving good performance in both clinic-based and community-based research cohorts. Therefore, the differences among the cohorts are beneficial as external validation. Before initiating the analysis, we conducted a literature review of factors associated with Aβ to provide critical context to interpret results of the subsequent data-driven approach.

2.1. Initial feature selection in ADNI and AIBL

The goal of this phase was to identify the variables that most strongly predicted elevated brain Aβ among nondemented participants with consistency across two different data sets, ADNI and AIBL. These variables, referred to as "features" hereafter, would then be carried forward to the next phase for the deep development of predicted probabilities.

ADNI is a multisite longitudinal study launched in 2004, observing the impact of aging via clinical assessment,

imaging, and biomarkers in a population largely recruited from memory clinics. AIBL is also a longitudinal study of aging and was launched in 2006 with a focus on cognition, biomarkers, and lifestyle factors in the development of AD. Study designs were previously published for ADNI [20,21] and AIBL [22], and updated methods are available online: http://adni.loni.usc.edu and https://aibl.csiro.au.

ADNI and AIBL participants were included if they had a completed Aβ-PET scan and qualified as cognitively unimpaired, SCD or MCI at their first Aβ-PET scan. Feature selection included cognitively unimpaired participants to maximize the power of feature selection and understand the importance of SCD indicators. Selection of candidate features for the algorithm was driven by both quantitative statistical measures and practical considerations (e.g., length and ease of administration, suitability of the instrument to a range of educational levels). A data-driven approach was used to analyze all available demographic, medical history, physical examination, vital signs, genetic, family history and lifestyle factors, and neuropsychological tests including summary scores, domain-level scores, and item-level scores, as potential features for selection; 654 variables from ADNI and 169 variables from AIBL were analyzed. Because the objective was a quick, low-burden, and low-cost algorithm that could be used in a primary care setting, data from magnetic resonance imaging or CSF biomarkers were not considered as candidate variables. Internally validated decision trees were used to identify the strongest features from ADNI (n = 760 participants) and AIBL (n = 746 participants). For primary analysis of predictors of Aβ pathology, 80% of participant samples were used from each data set and trained 100 × 1000 times. Aβ pathology was classified as present by PET in accordance with the ADNI and AIBL study protocols (Supplementary Material). The importance of the potential features was evaluated by the frequency that they were present in the simulated decision trees and their average position when present. Once the key features were determined, iterations of models were compared on distribution of the AUCs to determine which combination of features would provide the strongest overall performance.

2.2. Deep development of probability algorithm in ADNI

To derive the estimated probability of Aβ positivity, we developed a new simulation framework using nested cross-validation accompanied by stratified subsampling procedures of participants with SCD or MCI in ADNI and decision tree methods. MCI status was ascertained according to the ADNI and AIBL study criteria, which were consistent with accepted clinical methods [23,24] (Supplementary Material). SCD was classified as present in the ADNI data set among participants who did not qualify as having MCI if either the patient or informant Everyday Cognition Questionnaire score reached its respective threshold score (≥1.31 informants or ≥1.36 participants) [25–27]. For AIBL, SCD was classified as present by an IQcode 16-item short form

score of ≥3.38 [28,29]. The methods used to ascertain SCD are not prescriptive components of the algorithms but rather operationalization of the available measures in each study to allow a reasonable representation of a target patient population.

Two algorithms were developed, either omitting or including *APOE ε4* carrier status, which was confirmed to be a strong predictor during the first phase of this analysis (see the Results section). The statistical framework used stratified resampling and maximum AUC–split criteria to obtain internally validated decision trees averaged from 1000 iterations (Fig. 1). The decision tree method estimates probability based on the proportion of amyloid-positive samples in each class/category under each tree branch. The procedure was repeated, with stratified sampling of 250 participants resampled 1000 times, thereby simulating 1000 optimal decision trees trained from the target patient population. Sampling was stratified by 5-year age groups and MCI:SCD proportion within age group to resemble the target clinical scenarios. The predicted probability of Aβ positivity was obtained using the average of 1000 optimal decision trees for each combination of predictors.

2.3. Algorithm validation in AIBL and MCSA

Because AIBL was used in the first phase of feature selection but not used to derive probability estimates, the AIBL data set served as a semi-independent validation of the probability estimation of the algorithms. For this semi-independent validation, the AIBL data set was analyzed using 1000 iterations sampled by 5-year age groups and SCD:MCI proportion within age groups to resemble the target clinical scenarios, consistent with our method of the ADNI sampling.

A fully independent external validation of the algorithms was conducted using the MCSA, an epidemiologic study of aging and MCI in a population-based cohort. The study design was detailed previously [30].The primary validation included 711 participants with SCD (n = 490) or MCI (n = 221) at the most recent Aβ measure occurring from November 2006 to August 2017. MCI was defined using consensus agreement and published criteria [24,30]. SCD was determined to be present if the patient or informant Everyday Cognition Questionnaire score reached its respective threshold (≥1.31 informants or ≥ 1.36 participants) [25–27]. Aβ positivity was defined by Pittsburgh compound B PET standardized uptake value ratio >1.42 [31]. A secondary validation further included cognitively unimpaired participants without SCD (n = 1012) to evaluate the potential for broader clinical application of the algorithms in all nondemented (n = 1723).

Algorithm performance was evaluated using AUC, specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV), and positive likelihood ratio. A probability threshold of 0.5 was applied to the primary
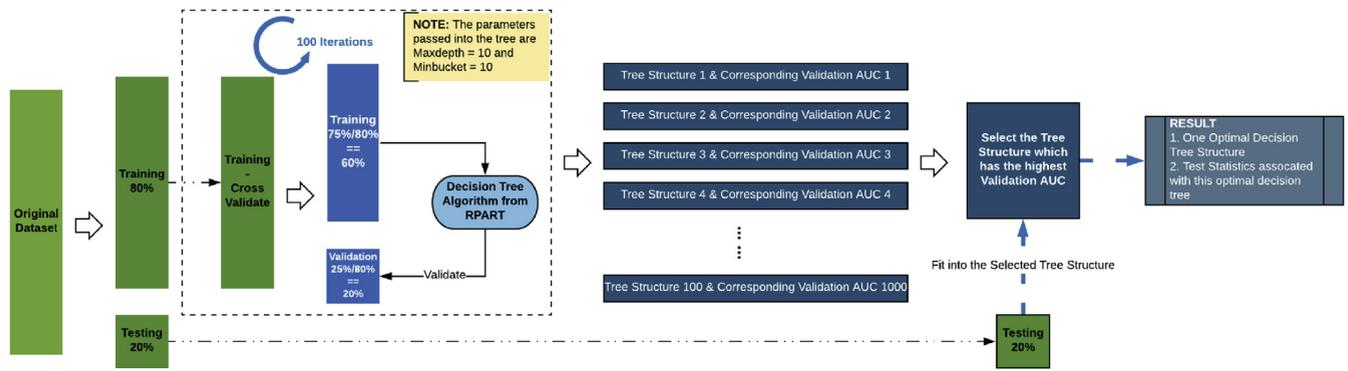
Fig. 1. Statistical framework: procedure for one optimal decision tree. One thousand decision trees were used to derive the probability distribution of the algorithm, with resampling without replacement using age-stratified and subjective cognitive decline:mild cognitive impairment–stratified sampling. Abbreviation: AUC, area under the curve.

analysis of performance, and secondary analyses examined performance using probabilities of 0.4, 0.6, and 0.7 as thresholds to predict positive Aβ status. Sensitivity analyses were conducted by age subgroup and by MCI or SCD status.

Although the algorithm was derived using the normalized immediate recall score on a word list learning task, its performance using other recall measures was also evaluated. Specifically, using the AIBL data, the Rey Auditory Verbal Learning Test (RAVLT) immediate recall z-score was substituted with the z-scores of the California Verbal Learning Test (CVLT) immediate, short-delayed, and long-delayed recall measures; using ADNI and MCSA data, the CVLT z-scores were substituted with the RAVLT short-delayed and long-delayed measures. Z-scores were obtained from published norms that were age-adjusted or age- and sex-adjusted.

To estimate the potential impact of the algorithm on referrals to specialists or Aβ confirmation testing, the performance metrics from MCSA validation were applied to the numbers of patients projected to be in the health care system for possible SCD or MCI due to AD on approval of a disease-modifying therapy for AD [8,9].

3. Results

3.1. Feature selection in ADNI and AIBL

Characteristics of the ADNI and AIBL samples are summarized in Table 1. Of the 654 variables in ADNI and 169 features in AIBL, output from the 100 × 100 decision trees indicated that *APOE ε4* status, age, and cognitive test consistently had the highest predictive value (Supplementary Fig. 1). The combination of all three features provided superior performance compared with combinations of just two or one (e.g., all three AUC = 78.3%; age and *APOE ε4* AUC = 72.9%, ADNI; Supplementary Fig. 2).

Iterative comparison of algorithm performance to select an appropriate cognitive assessment indicated that various cognitive assessments performed similarly well, with median AUCs over 1000 iterations ranging from 0.72 to 0.76 in ADNI and 0.70 to 0.72 in AIBL (Supplementary Fig. 3). Recall measures had slightly higher median AUCs (e.g., 0.75 AVLT in ADNI; 0.72 CVLT in AIBL) than global measures such as the Mini–Mental State Examination (0.73 in ADNI; 0.70 in AIBL), clock drawing (0.72 in ADNI; 0.71 in AIBL), or most measures of verbal fluency, language, attention, and subjective cognitive decline, but differences were not significant. The similarity across cognitive tests remained when the iterative decision trees included *APOE* status and age (AUCs in ADNI: 0.70 Alzheimer's Disease Assessment Scale-Cognitive to 0.72 Boston Naming Test; AUCs in AIBL: 0.68 clock drawing to 0.70 CVLT). Given this solid performance across cognitive tests, recall measures were selected for deriving predicted probabilities, based on consistency with prior research [11,32], relative ease of administration, and evidence that performance on recall tests is less affected by education compared with performance on other cognitive tests [33]. Because the AUC was the same whether the algorithm used immediate or delayed recall (e.g., ADNI AUC = 0.75 for either immediate or delayed; AIBL AUC = 0.72 for either), the immediate recall test was selected for time efficiency. The available recall measures differed in ADNI and AIBL (e.g., ADNI uses RAVLT [34], whereas AIBL uses CVLT) [35]. To allow input of different measures, raw scores were transformed to their normalized z-scores. The final variables included age, *APOE ε4* status, and immediate recall z-score.

3.2. Probability distribution development in ADNI

To derive the predicted probabilities of Aβ positivity for each combination of predictors, 1000 optimal decision trees were run for each algorithm: algorithm 1 using only age and immediate recall z-score (RAVLT) and algorithm 2 also using *APOE ε4* carrier status. For a display of the final probability distributions based on patient characteristics of age,

Table 1
Characteristics of the data sets used for development and validation

	ADNI			AIBL			MCSA		
	Feature selection and probability development			Feature selection and semi-independent validation			Validation		
Characteristic	Aβ+	Aβ−	Total	Aβ+	Aβ−	Total	Aβ+	Aβ−	Total
Participants, n	311	307	618	152	108	260	352	359	711
SCD, n (%)	49 (15.7)	95 (30.9)	144 (23.3)	35 (23.0)	46 (42.6)	81 (31.1)	204 (57.9)	286 (79.7)	490 (68.9)
MCI, n (%)	262 (84.2)	212 (69.1)	474 (76.7)	117 (77.0)	62 (57.4)	179 (68.9)	148 (42.1)	73 (20.3)	221 (31.1)
Female, n (%)	147 (47.3)	143 (46.6)	290 (46.9)	63 (41.4)	44 (40.7)	107 (41.2)	156 (44.3)	141 (39.3)	297 (41.8)
Mean age, y (SD)	72.9 (6.9)	70.1 (7.2)	71.6 (7.2)	76.0 (6.5)	72.5 (7.5)	74.5 (7.1)	79.6 (7.9)	70.6 (10.4)	75.1 (10.3)
Higher education, n (%)*	201 (64.6)	216 (70.4)	417 (67.5)	44 (28.9)	35 (32.4)	79 (30.4)	132 (37.5)	138 (38.4)	270 (38.0)
APOE ε4 status, n (%)									
Noncarrier	113 (36.3)	234 (76.5)	347 (56.1)	62 (40.8)	89 (82.4)	151 (58.1)	181 (51.4)	285 (79.4)	466 (65.5)
Carrier, heterozygous	154 (49.5)	64 (20.9)	218 (35.3)	61 (40.1)	13 (12.0)	74 (28.5)	146 (41.5)	64 (17.8)	210 (28.3)
Carrier, homozygous	43 (13.8)	8 (2.6)	51 (8.3)	15 (9.9)	0	15 (5.8)	18 (5.1)	1 (0.3)	19 (2.7)
Missing *APOE ε4* data	1 (0.3)	1 (0.3)	2 (0.3)	14 (9.2)	6 (5.6)	20 (7.7)	7 (2.0)	9 (2.5)	16 (2.3)

Abbreviations: Aβ, amyloid β; ADNI, Alzheimer's Disease Neuroimaging Initiative; AIBL, Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing; APOE, apolipoprotein E; MCSA, Mayo Clinic Study of Aging; SCD, subjective cognitive decline; MCI, mild cognitive impairment.
*Higher education was defined as years of education ≥16 in ADNI and MCSA and ≥15 in AIBL.

immediate recall test z-score, and, if desired, *APOE ε4* status, the algorithms were expressed as heat maps (Fig. 2). For both algorithms, the probability increased with increasing age and decreasing recall z-score. For algorithm 2, probability also increased in *APOE* ε4 carriers, a strong predictor such that all adult carriers over age 50 y had probability ≥0.5. With the heat map, individuals can be mapped to a combination of their age and recall z-score to obtain the estimated probability; if a certain probability threshold (e.g., ≥0.5) was deemed appropriate given available resources for referral or Aβ confirmation, then that threshold could be overlaid onto the heat map (e.g., dashed red lines in Fig. 2A, representing the ≥0.5 probability threshold). Alternatively, the inputs for a given patient could be entered into a clinical calculator that outputs the predicted probability with CIs. For example, for a patient aged 70 years with a recall z-score of −1.25, the estimated probability is 0.49 (interquartile range 0.35–0.63); if it is known that the patient is an *APOE ε4* carrier, the estimated probability shifts to 0.70 (interquartile range 0.68–0.78).

3.3. Validation

Performance metrics of each algorithm were consistent in the ADNI, AIBL, and MCSA populations, indicating a robust performance across these settings. Fig. 3 shows the performance metrics when a 0.5 probability was applied as the threshold of interest to predict Aβ positivity in the two validation data sets: with just age and recall z-score, the algorithm achieved PPV 66% in MCSA and 67% in AIBL; including *APOE ε4* input, the PPV was 69% in MCSA and 76% in AIBL. The 0.5 probability threshold resulted in the best AUC, at 71%; however, given the high cost of Aβ testing, a higher PPV might be more helpful. Table 2 shows the performance when

different probability thresholds were applied to the MCSA validation data set. For example, a probability threshold of 0.70 provided PPV 83%, with NPV 60%. Using a higher probability threshold also improved the positive likelihood ratio, from 2.3 to 5.0, indicating that the potential impact of the algorithm on clinical decision-making was increased.

The algorithms were further evaluated in subgroups, stratifying the MCSA data by SCD, MCI, or age (50–64.9, 65–74.9, 75–84.9, ≥85.0 years) (Supplementary Table 1). For both algorithms, PPV was higher among individuals with MCI (71% or 76% with *APOE ε4* status) compared with those with SCD (58% or 63%), although specificity was low in MCI (28% or 42%). For both algorithms, PPV increased with increasing age. Additional sensitivity analysis stratified by sex or education showed that AUCs were similar for males and females, although PPV was higher in females and NPV was higher in males, and algorithm 2 performed more consistently across education levels than did algorithm 1 (Supplementary Table 1). In secondary analysis that explored the performance among 1723 nondemented (including cognitively unimpaired without SCD) MCSA participants, AUC decreased slightly (algorithm 1: 66% to 64%; algorithm 2: 71% to 67%), NPV increased, and PPV decreased (Supplementary Table 2). In separate analysis of a small sample of MCSA participants with mild dementia (n = 23 for algorithm 1 and n = 22 for algorithm 2), both algorithms performed with 100% sensitivity and high PPV (82%–83%).

3.4. Substitution of recall tests

Performance remained robust to substitution of different recall measure z-scores, with the AUC, PPV, and NPV stable
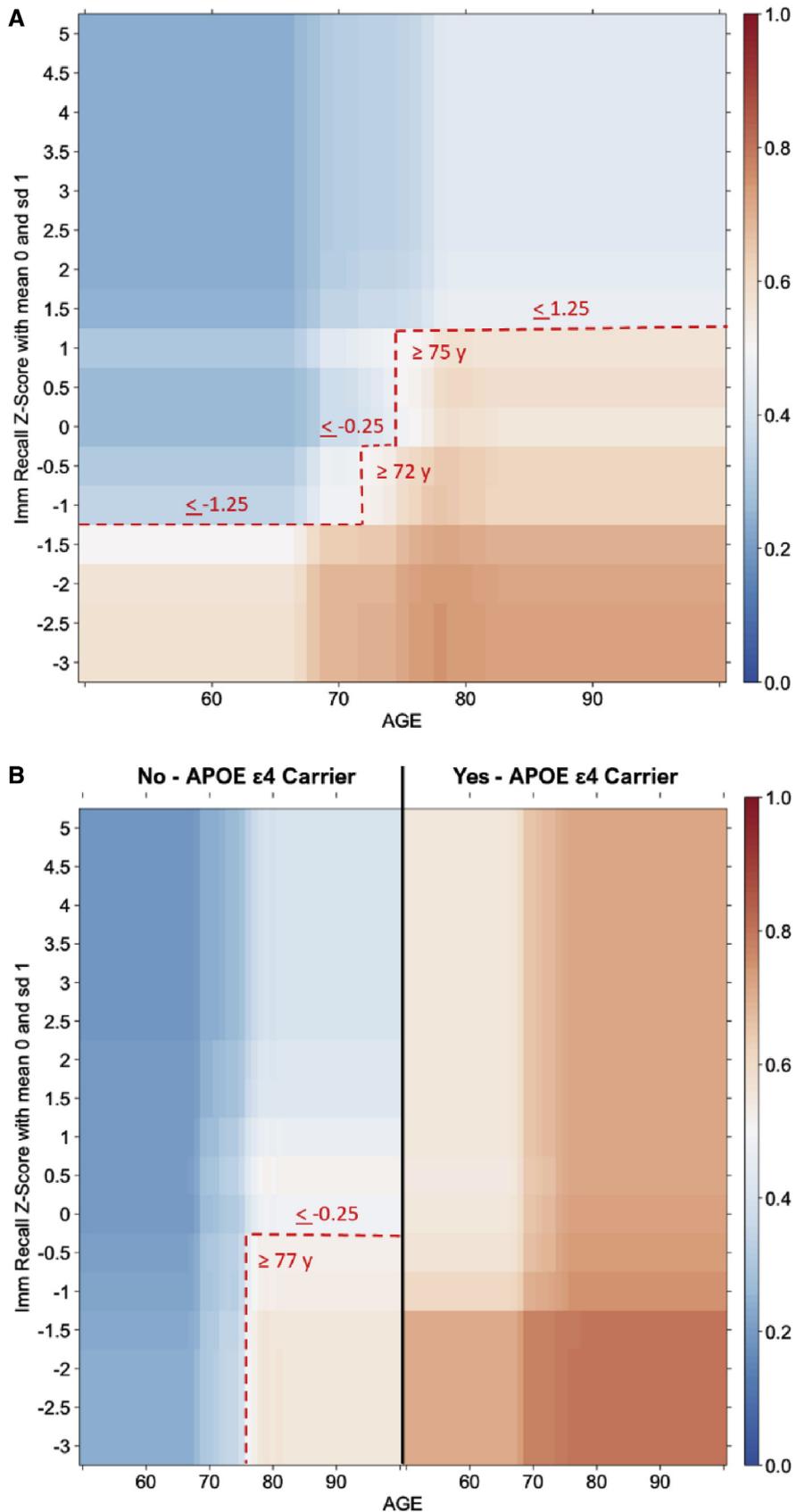
Fig. 2. Average heat maps for predicted Aβ positive status based on 1000 optimal decision trees. Algorithms generated using immediate recall test z-score and age without (A) or with (B) consideration of *APOE ε4* status. Red indicates higher probability of Aβ positivity, blue indicates higher probability of Aβ negativity. The hatched red line indicates the threshold for a probability of >0.5 to be considered predicted positive; different probability thresholds can be applied as appropriate depending on the clinical context and available resources. Abbreviations: Aβ, amyloid β; *APOE*, apolipoprotein E
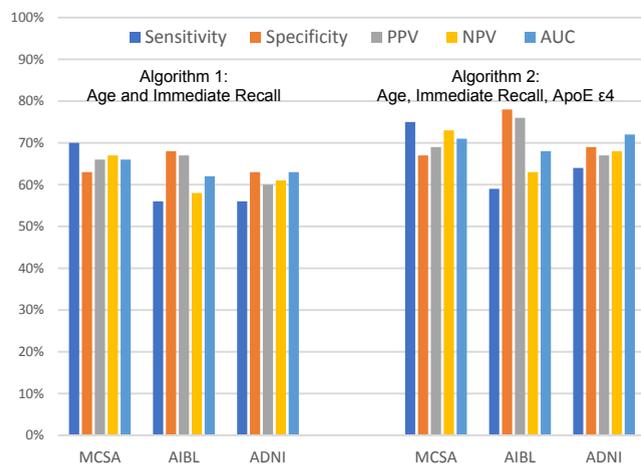
Fig. 3. Performance metrics of algorithm 1 (age and immediate recall) and algorithm 2 (age, immediate recall, and apolipoprotein ε4 status) based on 0.5 probability for predicting positivity in the validation data sets. ADNI validation was an internal validation using resampling for n = 250 over 1000 iterations; AIBL validation was a semi-independent validation using resampling of n = 91 over 1000 iterations; and MCSA was a fully independent validation (n = 711 in algorithm 1 validation; n = 695 in algorithm 2 validation). Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; AIBL, Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing; AUC, area under the curve; MCSA, Mayo Clinic Study of Aging; NPV, negative predictive value; PPV, positive predictive value.

within 1%–2% of the original result for algorithm 2 and within 4%–5% for algorithm 1 (Supplementary Fig. 4).

3.5. Potential impact on projected health care system constraints

On availability of a new AD-modifying therapy, approximately 14 to 15 million patients with possible MCI due to AD are estimated to be eligible for referral or biomarker testing in the US or select European health care systems [8,9]. Applying the algorithms as observed in the MCSA validation to this projected population could prevent an esti-

Table 2
Impact of varying probability thresholds: MCSA validation data set*

Probability	≥0.7	≥0.6	≥0.5	≥0.4
PPV	83%	79%	69%	64%
Specificity	93%	89%	67%	52%
NPV	60%	61%	73%	80%
Sensitivity	35%	43%	75%	87%
Likelihood ratio positive (95% CI)	5.01 (3.32–7.56)	3.84 (2.77–5.31)	2.25 (1.92–2.65)	1.81 (1.61–2.04)
AUC	64%	66%	71%	69%

Abbreviations: AUC, area under the curve; MCSA, Mayo Clinic Study of Aging; NPV, negative predictive value; PPV, positive predictive value.

*Data shown for algorithm using age, recall test z-score, and apolipoprotein ε4 carrier status.

mated 1.0 to 2.8 million negative Aβ confirmation tests while helping identify 0.1 to 3.4 million Aβ-positive symptomatic patients, depending on the desired probability threshold (Table 3).

4. Discussion

We developed and validated two practical algorithms to determine the probability of Aβ positivity in patients with SCD or MCI, using a rigorous statistical framework for probability estimation in both clinical and population-based data sets. Feature selection was guided by the principle that to increase efficiency of biomarker testing, an algorithm ideally should be based on inputs that are quickly administered and readily available while still performing with high test characteristics. As such, algorithm 1 was developed requiring only inputs of age and an immediate recall test, which may be administered in approximately 5 minutes. Algorithm 2 also considered *APOE ε4* carrier status, a quick and often easily accessible genetic test. Both algorithms were robust across clinic-based populations (ADNI, AIBL) and the population-based sample participants (MCSA).

A strength of this study was the creation of a rigorous statistical framework as a foundation for the probability estimation. By using nested cross-validation with stratified subsampling procedures, problems caused by heterogeneity among data sets were reduced and modeling for the specific target population was improved. This framework prevents overfitting and increases reproducibility and model robustness. Indeed, the algorithms' performance metrics were largely similar across ADNI, AIBL, and MCSA, despite the differences in study settings and designs. This statistical structure is generalizable and could easily be extended to apply to different target populations or biomarkers.

Compared with other published practical algorithms for Aβ probability in SCD or MCI, the predictive performance of the current algorithms was similar, while carrying the added advantage of flexibility for the required inputs and validation in an epidemiologic data set. Although AUC was just slightly lower—at best 0.71 in the validation data set using age, recall z-score, and *APOE ε4*, compared with 0.75 to 0.82 for other models [11,18,19]—other models were tested only in specialized clinical sites. The AUCs we observed during feature selection were in the same higher range as other models (e.g., 78% for age, *APOE ε4*, and cognitive test, Supplementary Fig. 2), and after we applied nested cross-validation with stratified subsampling over 1000 iterations to derive probabilities, the AUCs decreased. This observation supports the notion that the performance of the algorithms derived here is tempered to yield more stable performance in various settings. Furthermore, AUC is not necessarily the preferred performance metric when the confirmatory test (e.g., PET) is costly and has limited

Table 3
Projected impact of applying Aβ probability algorithms for the 14.9 million US patients aged ≥55 years projected to screen positive for MCI*

Scenario	RAND report projected number	Applying algorithm[†]					
		≥0.6 Probability threshold		≥0.5 Probability threshold		≥0.4 Probability threshold	
	(No algorithm)	Age, recall	With *APOE ε4*	Age, recall	With *APOE ε4*	Age, recall	With *APOE ε4*
Send to Aβ confirmation	6.7 M	5.1 M	4.0 M	7.7 M	8.0 M	9.9 M	10.0 M
Confirmed (true + sent)	3.0 M	3.6 M	3.1 M	5.0 M	5.5 M	6.1 M	6.4 M
Not confirmed (false + sent)	3.7 M	1.5 M	0.9 M	2.7 M	2.5 M	3.8 M	3.6 M

Abbreviations: Aβ, amyloid β; *APOE*, apolipoprotein E.

*Projected numbers obtained from the RAND report for US health care system readiness for an Alzheimer's disease–modifying therapy; projections for five European countries were of similar magnitude, with an estimated 14.3 M patients in those health care systems screening positive for mild cognitive impairment (data not shown) [8,9].

[†]Algorithm listed as "age, recall" uses age and recall z-scores. Algorithm listed as "with *APOE ε4*" uses age, recall z-score, and *APOE ε4* positive status. Values in the "send to Aβ confirmation" row refer to patients who would be predicted positive with the algorithm for a given threshold for probability (e.g., as displayed in table: 0.6, 0.5, and 0.4 probability). Values are derived from the performance of the algorithms in the Mayo Clinic Study of Aging validation data set using Rey Auditory Verbal Learning Test immediate recall z-score.

availability [36]. Rather, PPV and positive likelihood ratio may be most relevant because a higher PPV more directly reduces the number of Aβ tests returned as negative (reducing unnecessary cost and burden), and a higher positive likelihood ratio conveys a larger impact on the clinician's initial judgment [36,37]. While a 0.5 probability best balances sensitivity and specificity, the probability threshold best suited for a given clinical scenario depends on numerous factors that vary across clinics, such as patient volume and availability of PET scanners or specialists. With this in mind, our analysis considered alternative probability thresholds that may be relevant in different settings based on resource availability and provider preferences.

These algorithms were developed to maintain flexible inputs for application in clinical practice. As such, unlike previously developed algorithms, the algorithms do not require the use of specific cognitive and genetic tests [11,18,19]. Although *APOE ε4* status is a strong predictor of Aβ pathology, there may be scenarios in which genetic counseling is problematic or not easily attainable. Probability values were derived both with and without *APOE ε4* information, resulting in different probability distributions across the two algorithms; *APOE ε4* information is not simply an additive component. Another strength is that the algorithms do not specify which recall test must be used, as a variety of recall tests are effective at detecting MCI in clinical settings [38], with episodic memory most consistently and strongly related to cognitive decline due to AD pathology [39–41]. Recall tests are one of the most commonly documented cognitive assessments in current primary care [42], indicating that these algorithms can fit comfortably into current clinical practice. The algorithms are also not prescriptive for the assessment used for SCD, in line with the 2017 Gerontological Society of America and 2018 Alzheimer's Association tool kits, which have flexible guidelines for ascertaining SCD [43–45].

In clinical practice, these algorithms may be useful to increase the confidence of primary care providers or specialists in their clinical decision-making and furthermore improve efficiency by reducing the number of patients sent for Aβ testing. For patients with MCI, the use of these algorithms could shift the estimated probability of Aβ positivity from a prior probability of 0.45 to 0.50 [8,9,46,47] to approximately 0.65 to 0.75 (Fig. 3). For patients with SCD, the estimated probability may shift from approximately 0.20 to 0.30 [48] to approximately 0.60 (Fig. 3). Confidence intervals provide reassurance on the estimated probability. In light of limited resources and high costs of confirmatory testing, providers could consider a patient's probability of Aβ positivity and send only those patients above a given probability threshold for confirmatory testing. Patients below the threshold might be appropriate for close monitoring (i.e., "watchful waiting") and reassessment at follow-up visits. Such targeted referrals to specialists or Aβ testing may be necessary to reduce burden and increase access to those patients who are most likely to benefit [8,9].

Although these algorithms are designed to help clinical decision-making, they are not perfect predictors of Aβ PET. That is, while decreasing the number of false positives, there will inevitably be patients with Aβ pathology who do not meet the selected probability threshold. For this reason, the algorithms best serve as an adjunct to other considerations in the decision for specialist referral, confirmation testing, or watchful waiting. Follow-up assessments to monitor cognitive decline are important for patient care. The moderately good predictive performance of these algorithms reflects the best of what is currently achievable for practical and low-cost inputs (lacking validated blood-based biomarkers and other potentially emerging technologies). Should a new therapy become approved for AD intervention, an estimated 14.9 million patients over age 55 years may screen positively for MCI in a single year in

the US, with a health care system ill equipped for confirming pathology in this large population, and similar problems in other countries [8,9]. Application of either of these algorithms to this projected population could help diagnose individuals with underlying Aβ pathology while preventing an estimated 1 to 2.8 million negative Aβ confirmation tests. By applying a practical algorithm, there is potential to minimize unnecessary costs and burdens to the patient, provider, and health care system.

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.dadm.2019.09.001.

RESEARCH IN CONTEXT

1. Systematic Review: We reviewed literature on predictive models for cerebral Aβ. Numerous factors, including age, cognitive impairment, *APOE* genotype, CSF inflammatory, or protein biomarkers have been associated with Aβ positivity. Available predictive models are limited by lacking external validation or requiring inputs that are burdensome or not universally available.

2. Interpretation: We developed a multistep statistical framework to obtain robust probability estimates across clinical and nonclinical settings using two different data sources and independently validating in a third, nonclinical population-based cohort. Compared with other published practical algorithms for Aβ probability, the predictive performance of the current algorithms was similar, while carrying the advantage of flexibility regarding the selection of recall test and *APOE* ε4 test.

3. Future directions: While these algorithms may help identify patients for biomarker testing, a validated blood-based or other low-cost, low-burden biomarker that can replace CSF or PET testing would critically improve Alzheimer's disease detection and diagnosis.

References

[1] Alzheimer's Association. 2018 Alzheimer's disease facts and figures. Alzheimers Dement 2018;14:367–429.

[2] Galvin JE, Sadowsky CH, NINCDS-ADRDA. Practical guidelines for the recognition and diagnosis of dementia. J Am Board Fam Med 2012;25:367–82.

[3] Jack CR, Lowe VJ, Weigand SD, Wiste HJ, Senjem ML, Knopman DS, et al. Alzheimer's Disease Neuroimaging Initiative, Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. Brain 2009;132:1355–65.

[4] Dubois B, Hampel H, Feldman HH, Scheltens P, Aisen P, Andrieu S, et al. Proceedings of the Meeting of the International Working Group (IWG) and the American Alzheimer's Association on "The Preclinical State of AD", 2. July 23, U. Washington DC, Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. Alzheimers Dement 2016;12:292–323.

[5] Jack CR J, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al., Contributors, NIA-AA Research Framework. Toward a biological definition of Alzheimer's disease. Alzheimers Dement 2018;14:535–62.

[6] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 2014;13:614–29.

[7] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's

Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–92.

[8] Liu JL, Hlávka JP, Hillestad R, Mattke S. Assessing the Preparedness of the U.S. Health Care System Infrastructure for an Alzheimer's Treatment. Santa Monica, CA: RAND Corporation; 2017. Available at: https://www.rand.org/pubs/research_reports/RR2272.html. Accessed October 7, 2019.

[9] Hlávka JP, Mattke S, Liu JL. Assessing the Preparedness of the Health Care System Infrastructure in Six European Countries for an Alzheimer's Treatment. Santa Monica, CA: RAND Corporation; 2018. Available at: https://www.rand.org/pubs/research_reports/RR2503.html. Accessed October 7, 2019.

[10] Bos I, Vos S, Verhey F, Scheltens P, Teunissen C, Engelborghs S, et al. Cerebrospinal fluid biomarkers of neurodegeneration, synaptic integrity, and astroglial activation across the clinical Alzheimer's disease spectrum. Alzheimers Dement 2019;15:644–54.

[11] Jansen WJ, Ossenkoppele R, Tijms BM, Fagan AM, Hansson O, Klunk WE, et al. Association of cerebral amyloid-β aggregation with cognitive functioning in persons without dementia. JAMA Psychiatry 2018;75:84–95.

[12] Crous-Bou M, Minguillón C, Gramunt N, Molinuevo JL. Alzheimer's disease prevention: from risk factors to early intervention. Alzheimers Res Ther 2017;9:71.

[13] Landau SM, Marks SM, Mormino EC, Rabinovici GD, Oh H, O'neil JP, et al. Association of lifetime cognitive engagement and low β-amyloid deposition. Arch Neurol 2012;69:623–9.

[14] Mathys J, Gholamrezaee M, Henry H, von Gunten A, Popp J. Decreasing body mass index is associated with cerebrospinal fluid markers of Alzheimer's pathology in MCI and mild dementia. Exp Gerontol 2017;100:45–53.

[15] Ten Kate M, Redolfi A, Peira E, Bos I, Vos SJ, Vandenberghe R, et al. MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study. Alzheimers Res Ther 2018; 10:100.

[16] Tosun D, Chen Y, Yu P, Sundell KL, Suhy J, Siemers E, et al. Alzheimer's Disease Neuroimaging Initiative, Amyloid status imputed from a multimodal classifier including structural MRI distinguishes progressors from nonprogressors in a mild Alzheimer's disease clinical trial cohort. Alzheimers Dement 2016;12:977–86.

[17] Kim SE, Woo S, Kim SW, Chin J, Kim HJ, Lee BI, et al. A nomogram for predicting amyloid PET positivity in amnestic mild cognitive impairment. J Alzheimers Dis 2018;66:681–91.

[18] Palmqvist S, Insel PS, Zetterberg H, Blennow K, Brix B, Stomrud E, et al. Alzheimer's Disease Neuroimaging Initiative, Accurate risk estimation of β-amyloid positivity to identify prodromal Alzheimer's disease: cross-validation study of practical algorithms. Alzheimers Dement 2019;15:194–204.

[19] Lee JH, Byun MS, Yi D, Sohn BK, Jeon SY, Lee Y, et al. Prediction of cerebral amyloid with common information obtained from memory clinic practice. Front Aging Neurosci 2018;10:309.

[20] Veitch DP, Weiner MW, Aisen PS, Beckett LA, Cairns NJ, Green RC, et al. Alzheimer's Disease Neuroimaging Initiative, Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's Disease Neuroimaging Initiative. Alzheimers Dement 2019;15:106–52.

[21] Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, et al. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin N Am 2005;15:869–77. xi.

[22] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 2009;21:672–87.

[23] Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, et al. Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. J Intern Med 2004;256:240–6.

[24] Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. Neurology 2018;90:126–35.

[25] Farias ST, Mungas D, Reed BR, Cahn-Weiner D, Jagust W, Baynes K, et al. The measurement of everyday cognition (ECog): scale development and psychometric properties. Neuropsychology 2008;22:531–44.

[26] Rueda AD, Lau KM, Saito N, Harvey D, Risacher SL, Aisen PS, et al. Alzheimer's Disease Neuroimaging Initiative, self-rated and informant-rated everyday function in comparison to objective markers of Alzheimer's disease. Alzheimers Dement 2015;11:1080–9.

[27] van Harten AC, Mielke MM, Swenson-Dravis DM, Hagen CE, Edwards KK, Roberts RO, et al. Subjective cognitive decline and risk of MCI: the Mayo Clinic Study of Aging. Neurology 2018; 91:e300–12.

[28] Grober E, Wakefield D, Ehrlich AR, Mabie P, Lipton RB. Identifying memory impairment and early dementia in primary care. Alzheimers Dement (Amst) 2017;6:188–95.

[29] Jorm AF. A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. Psychol Med 1994;24:145–53.

[30] Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, et al. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology 2008;30:58–69.

[31] Jack CR Jr, Wiste HJ, Weigand SD, Therneau TM, Lowe VJ, Knopman DS, et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. Alzheimers Dement 2017;13:205–16.

[32] Hedden T, Oh H, Younger AP, Patel TA. Meta-analysis of amyloid-cognition relations in cognitively normal older adults. Neurology 2013;80:1341–8.

[33] Lipnicki DM, Crawford JD, Dutta R, Thalamuthu A, Kochan NA, Andrews G, et al. Cohort Studies of Memory in an International Consortium (COSMIC), Age-related cognitive decline and associations with sex, education and apolipoprotein E genotype across ethnocultural groups and geographic regions: a collaborative cohort study. PLoS Med 2017;14:e1002261.

[34] Schmidt M. Rey Auditory Verbal Learning Test: A Handbook. Los Angeles, CA: Western Psychological Services; 1996.

[35] Delis D, Kramer J, Kaplan E, Thompkins B. California Verbal Learning Test-Adult Version: Manual. San Antonio, TX: Psychological corporation; 1987.

[36] van Stralen KJ, Stel VS, Reitsma JB, Dekker FW, Zoccali C, Jager KJ. Diagnostic methods I: sensitivity, specificity, and other measures of accuracy. Kidney Int 2009;75:1257–63.

[37] Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet 2005;365:1500–5.

[38] Tsoi KK, Chan JY, Hirai HW, Wong A, Mok VC, Lam LC, et al. Recall tests are effective to detect mild cognitive impairment: a systematic review and meta-analysis of 108 diagnostic studies. J Am Med Dir Assoc 2017;18:807.e17–807.e29.

[39] Bischof GN, Rodrigue KM, Kennedy KM, Devous MD S, Park DC. Amyloid deposition in younger adults is linked to episodic memory performance. Neurology 2016;87:2562–6.

[40] Ramanan S, Flanagan E, Leyton CE, Villemagne VL, Rowe CC, Hodges JR, et al. Non-verbal episodic memory deficits in primary progressive aphasias are highly predictive of underlying amyloid pathology. J Alzheimers Dis 2016;51:367–76.

[41] E Baker J, Ying Lim Y, Jaeger J, Ames D, T Lautenschlager N, Robertson J, et al. Episodic memory and learning dysfunction over an 18-month period in preclinical and prodromal Alzheimer's disease. J Alzheimers Dis 2018;65:977–88.

[42] Maserejian N, Wang J, Krzywy H, Juneja M, Jaeger J. Cognitive and other neuropsychological assessments documented in electronic health records prior to or at Alzheimer's disease diagnosis. Alzheimers Dement 2018;14:P423.

[43] Alzheimer's Association. Cognitive impairment care planning toolkit. 2018. Available at: https://alz.org/media/Documents/cognitive-impairment-care-planning-toolkit.pdf. Accessed October 7, 2019.

[44] Gerontological Society of America. Kickstart, Assess, Evaluate, Refer (KAER), a 4-step process to detecting cognitive impairment and earlier diagnosis of dementia: Approaches and tools for primary care providers. 2017. Available at: https://www.geron.org/programs-services/alliances-and-multi-stakeholder-collaborations/cognitive-impairment-detection-and-earlier-diagnosis. Accessed October 7, 2019.

[45] Ossenkoppele R, Jagust WJ. The complexity of subjective cognitive decline. JAMA Neurol 2017;74:1400–2.

[46] Doraiswamy PM, Sperling R, Johnson K, Reiman EM, Wong T, Sabbagh M, et al. Florbetapir F 18 amyloid PET and 36-month cognitive decline: a prospective multicenter study. Mol Psychiatry 2014;19:1044.

[47] Ong KT, Villemagne VL, Bahar-Fuchs A, Lamb F, Langdon N, Catafau AM, et al. Abeta imaging with 18F-florbetaben in prodromal Alzheimer's disease: a prospective outcome study. J Neurol Neurosurg Psychiatry 2015;86:431–6.

[48] Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FR, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. JAMA 2015; 313:1924–38.