Cognitive & Behavioral Assessment

A classification algorithm for predicting progression from normal cognition to mild cognitive impairment across five cohorts: The preclinical AD consortium

Alden L. Gross[a,b,*], Jason J. Hassenstab[c], Sterling C. Johnson[d,e], Lindsay R. Clark[d,e], Susan M. Resnick[f], Melissa Kitner-Triolo[f], Colin L. Masters[g], Paul Maruff[g], John C. Morris[c], Anja Soldan[h], Corinne Pettigrew[h], Marilyn S. Albert[h]

[a]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[b]Johns Hopkins University Center on Aging and Health, Baltimore, MD, USA
[c]Charles F. and Joanne Knight Alzheimer's Disease Research Center, Washington University School of Medicine, St. Louis, MO, USA
[d]Wisconsin Alzheimer's Institute and Alzheimer's Disease Research Center, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI, USA
[e]Geriatric Research Education and Clinical Center, William S Middleton Veterans Memorial Hospital, Madison, WI, USA
[f]Laboratory of Behavioral Neuroscience, National Institute on Aging, NIH, Baltimore, MD, USA
[g]The Florey Institute, University of Melbourne, Australia
[h]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA

Abstract

Introduction: We established a method for diagnostic harmonization across multiple studies of preclinical Alzheimer's disease and validated the method by examining its relationship with clinical status and cognition.

Methods: Cognitive and clinical data were used from five studies ($N = 1746$). Consensus diagnoses established in each study used criteria to identify progressors from normal cognition to mild cognitive impairment. Correspondence was evaluated between these consensus diagnoses and three algorithmic classifications based on (1) objective cognitive impairment in 2+ tests only; (2) a Clinical Dementia Rating (CDR) of ≥ 0.5 only; and (3) both. Associations between baseline cognitive performance and cognitive change were each tested in relation to progression to algorithm-based classifications.

Results: In each study, an algorithmic classification based on both cognitive testing cutoff scores and a CDR ≥ 0.5 provided optimal balance of sensitivity and specificity (areas under the curve: 0.85–0.95). Over an average 6.6 years of follow-up (up to 28 years), $N = 186$ initially cognitively normal participants aged on average 64 years at baseline progressed (incidence rate: 15.3 people/ 1000 person-years). Baseline cognitive scores and cognitive change were associated with future diagnostic status using this algorithmic classification.

Discussion: Both cognitive tests and CDR ratings can be combined across multiple studies to obtain a reliable algorithmic classification with high specificity and sensitivity. This approach may be applicable to large cohort studies and to clinical trials focused on preclinical Alzheimer's disease because it provides an alternative to implementation of a time-consuming adjudication panel.

© 2017 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: Preclinical Alzheimer's disease; Harmonization; Longitudinal follow-up; Cognitive testing; Diagnostic classification

*Corresponding author. Tel.: +1 410-474-3386; Fax: +1 410-614-9625.

E-mail address: agross14@jhu.edu

1. Introduction

In middle-aged and older adults, Alzheimer's disease (AD) pathology begins up to 20 years before the onset of any clinically recognizable symptoms [1,2]. This long preclinical period provides an opportunity for clinical trials of interventions designed to slow or even halt AD-related pathological changes and thereby slow or halt the onset of cognitive impairment or dementia [3,4]. However, it is essential to understand the natural history of developing AD to guide decisions about the effectiveness and safety of interventions in the preclinical stage [5,6]. Although there are natural history studies that seek to understand relationships between changes in clinical symptoms, cognition, and biomarkers in preclinical AD, information derived from each of these studies alone may be biased by issues related to inclusion/exclusion criteria, methods for sample ascertainment, sample size, biomarker methodologies, clinical and cognitive outcome measures, as well as the length of follow-up and the number of times individuals are assessed. The preclinical AD consortium was established to develop methods and strategies for combining existing longitudinal study data to generate brain-behavior models that contain minimal bias. These models can then provide a strong foundation for understanding early AD and developing clinical trials in preclinical AD.

Although combining data across multiple preclinical AD cohorts can inform the selection of optimal cognitive and clinical end points for clinical trials of preclinical AD, such as magnitude of decline in memory among asymptomatic individuals or time to progression to the earliest symptomatic stage of AD, commonly operationalized as mild cognitive impairment (MCI), several challenges must be overcome before relevant information can be integrated.

The original criteria for MCI [7] emphasized that the diagnosis of MCI should be based on clinical judgment, as did the recent revision of the MCI criteria based on the recommendations of the National Institute on Aging/Alzheimer's Association (NIA-AA) workgroup [8]. Consensus diagnoses based on this approach have been shown to be highly feasible within a single study. However, evidence suggests that different adjudication processes across studies may cause similar cases to be classified differently depending on how panel members weight clinical information, cognitive testing, and informant reports [9–11]. To improve the standardization of diagnoses for multisite studies, performance below a specific cut-point on an episodic memory test has been incorporated into the evaluation, but this has still resulted in groups of individuals who vary considerably in outcome [12]. Some studies have developed algorithms that utilize more than one cognitive test score in generating a diagnosis and have reported that this improves the likelihood that the individual will progress from MCI to dementia over time [13]. One method that can be used to combine data pertaining to classification of outcomes is to develop and apply a common algorithm to clinical and cognitive data from the different studies and thereby overcome study-specific approaches. No study, to our knowledge, has examined these issues from the perspective of progression from normal cognition to MCI.

In addition, although most studies agree on the domains of cognition that warrant study in preclinical AD, they may choose different neuropsychological tests to operationalize those domains. Thus, it is not possible to simply aggregate information from different studies. This challenge can be addressed by applying approaches to harmonize data from different neuropsychological tests into a common metric (e.g., [14–18]).

The purpose of this study was to use prospective clinical and cognitive data from five cohorts designed to study preclinical AD to develop and validate a common classification algorithm. Since all participants were cognitively normal at baseline, it was possible to evaluate each algorithm's criterion and convergent validity in each study, as well as relationship between each alternative algorithm and the change in study-assigned clinical classifications over time. The relationship between each alternative algorithm and cognitive decline over time in the pooled sample was further examined, using cognitive factor scores across the data sets for general cognitive performance, memory, and executive function, as well as individual test scores.

2. Methods

2.1. Participants

Clinical and cognitive data were used from the Adult Children Study (ACS, $N = 360$; 19), the Australian Imaging, Biomarker, and Lifestyle study (AIBL study, $N = 767$; 20), the Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD) study ($N = 301$; 21), the Baltimore Longitudinal Study of Aging (BLSA, $N = 147$; 22)–a neuroimaging substudy, and the Wisconsin Registry for Alzheimer's Prevention (WRAP, $N = 171$; 23). These are subsamples from the larger cohorts that were selected on the basis of having comparable imaging and biofluid AD biomarkers available. The overall sample size was $N = 1746$, all of whom were cognitively normal at enrollment. At the time of these analyses, participants had been followed for up to 28 years.

Specific details about each data set are available elsewhere. Briefly, each study began in the 1990s or early 2000s and recruited cognitively normal participants, most of whom were middle aged. With the exception of BLSA, most participants in each study had a first degree relative with dementia. The overall goal of each of these studies was to identify early cognitive, imaging, and biospecimen markers of progression to MCI. The ACS began in 2002 at the Charles F. and Joanne Knight Alzheimer's Disease Research Center at Washington University in St. Louis School of Medicine and consists largely of adult children of persons with AD dementia [19]. The AIBL study, begun in 2006, recruited older adults with and

Table 1
Cognitive tests in studies across five studies in the preclinical AD consortium ($N = 1746$)

Cognitive test	ACS	AIBL	BIOCARD	BLSA	WRAP
Memory					
Logical Memory IA–immediate	36.9	100.0	100.0		100.0
Logical Memory IIA–delayed	36.9	100.0	100.0		100.0
Logical Memory IB–immediate			96.0		100.0
Logical Memory IIB–delayed			96.0		100.0
Buschke Selective Reminding Test	100.0				
CVLT immediate recall		100.0	96.0	100.0	
CVLT short-delay recall		100.0	96.0	100.0	
CVLT long-delay recall		100.0	96.0	100.0	
AVLT total recall					100.0
AVLT delayed recall					100.0
Nonmemory					
MMSE	99.7	100.0	99.3	97.3	100.0
Boston Naming Test, percent correct	56.7	99.9	99.7	95.2	99.4
WAIS-R Block Design			99.7		93.6
WAIS-R Digit Symbol	36.9		99.7	93.2	100.0
Trail Making, part A	100.0		95.0	98.0	100.0
Trail Making, part B	100.0		95.0	98.0	99.4
Animal fluency, 60 seconds	100.0	94.0	95.0	98.0	93.0
Vegetable fluency, 60 seconds	36.9		91.0	98.0	
Fruits fluency, 60 seconds		92.7			
Phonemic fluency–S words	36.9	92.8	92.0	98.0	
Phonemic fluency–errors					93.0
Digits Forward, Trials correct		94.3	100.0	100.0	100.0
Digits Backward, Trials correct		94.3	100.0	100.0	100.0
Digit Symbol Copy		99.9			
Clock drawing		100.0			
Rey Complex Figure Draw, immediate		99.9	99.7		
Rey Complex Figure Draw, delayed		99.9	99.7		

Abbreviations: ACS, Adult Children Study; AIBL, Australian Imaging, Biomarker, and Lifestyle; AVLT, Auditory Verbal Learning Test; BLSA, Baltimore Longitudinal Study of Aging; CVLT, California Verbal Learning Test; MMSE, Mini–Mental State Examination; WAIS, Wechsler Adult Intelligence Scale; WRAP, Wisconsin Registry for Alzheimer's Prevention.

NOTE. Numbers are percentages of observations in a study with data on the test.

without dementia at baseline to study neuropsychological, lifestyle, and mood predictors of AD dementia. In analyses presented here, only those AIBL participants who were cognitively normal at baseline were included [20]. The BIOCARD study began in 1995 at the Geriatric Psychiatry branch of the Intramural Program of the National Institute of Mental Health; the study was stopped in 2005 and reestablished in 2009 by a research team at the Johns Hopkins School of Medicine [21]. The BLSA is an NIA Intramural Research Study; it began in 1958 [22] and, starting in 1994, a subset of participants were invited to the neuroimaging substudy in which they received longitudinal neuroimaging, cognitive testing, and clinical assessments [23]. The WRAP study began in 2001, recruiting cognitively normal adults of which 72% were adult children of late-onset AD dementia patients [24].

2.2. Consensus diagnosis approach of each study

Each of the studies included in these analyses completed a consensus diagnosis at baseline and annually or as needed thereafter. The teams at each site consisted of a mix of clinicians, including physicians, research nurses, research assistants, and neuropsychologists. Each team met together on a regular basis (usually monthly) and adjudicated the cases. Each study used the Clinical Dementia Rating (CDR) [25] in its clinical assessment, as one piece of information to establish a consensus diagnosis for each subject at each visit. Additional information examined in consensus diagnosis procedures included cognitive test performance (single time point or longitudinal, depending on study), medical, neurologic and psychiatric assessments, and informant interviews that are part of the CDR.

In addition, each preclinical AD study assessed their participants with a comprehensive neuropsychological battery that spanned domains including episodic memory, executive function, language, spatial ability, attention, and psychomotor speed, but the batteries were not identical between any two studies. Individual tests in each battery are listed in Table 1 and described in Supplementary Information 1.

2.3. Analysis plan

The analysis was performed in two major steps. First, cognitive factors were constructed in the pooled data using available tests in each study. Second, the algorithmic classification was derived and validated.

2.3.1. Harmonization of cognitive tests

To facilitate comparison and combination of cognitive performance across the studies in pooled data, three cognitive factors were derived using confirmatory factor analysis methods described elsewhere [15–18]. These were labeled: (1) general cognitive performance, (2) episodic memory, and (3) executive function based on the content of the tests that loaded on each factor. Briefly, for each cognitive domain factor analysis, models were estimated consistent with two-parameter logistic item response models using a Bayesian estimator that permitted use of tests common across studies as well as noncommon tests in Mplus software, version 7.3 [26]. Item-level fit of data in the models was evaluated using normalized residuals [27,28]. Factor scores were estimated in the pooled data for each model based on averages of 30 plausible values from the posterior distribution generated by the models [26]. Before combining cognitive testing data across the different studies, test versions and characteristics were reviewed with study-specific codebooks and documentation, and by comparing means and ranges of the variables. To further verify equivalence of common tests, measurement noninvariance across data sets was evaluated using alignment analysis in Mplus, version 7.3 [26].

The episodic memory factor included the Auditory Verbal Learning Test immediate and delayed recall; California Verbal Learning Test (CVLT) immediate, short-, and long-delay recall; Wechsler Memory Scale Logical Memory immediate and delayed recall for stories A and B, and the Free and Cued Selective Reminding Test (see Supplementary Information 1). The executive function factor included fluency for animals, fruits, and vegetables; phonemic fluency; Digit Span Backwards; Digit Symbol Substitution; and Trail Making tests A and B. The general cognitive factor included all available cognitive test variables in each data set.

2.3.2. Algorithmic classification of MCI

To determine the optimal method for constructing an algorithmic classification of progression from normal cognition to MCI, three alternative approaches were developed and their relationship with the study-assigned diagnoses in each cohort computed. The three algorithms were based on (1) cognitive test performance only, (2) CDR ≥0.5 only, and (3) both cognitive test performance and CDR ≥0.5. These criteria made it possible to classify participants at each study visit for which cognitive and CDR data were available.

Internally derived age-adjusted means were established for each test in the pooled data (see Supplementary Table 7). Several analyses were conducted to determine the optimal cut-point that would identify those with a decline in cognitive performance. Then, several potential cut-points (0.5 standard deviation [SD], 1 SD, and 1.5 SD) below the age-adjusted mean of each test were examined. The cut-point 1 SD was selected to balance specificity and sensitivity to study-assigned diagnoses in each study data set (see Supplementary Information 2 for results comparing the different cut-points to one another) [13]. Poor cognitive performance was operationalized as having either two or more memory test scores or two or more nonmemory test scores ≥1 SD below the age-adjusted mean for the test.

A participant was defined using the algorithmic classification as progressing from cognitively normal to MCI if they did not meet criteria at their baseline study visit, but later met algorithmic criteria during at least two consecutive follow-up study visits. Two consecutive follow-up visits were required for this definition of progression because previous studies have demonstrated that data from a single time point can misclassify low-performing participants who may never progress to impairment or overestimation of one's status if there are retest effects [29,30].

2.3.3. Validation procedures

First, each of the three alternative algorithms was evaluated within each data set separately, to examine the correspondence between each algorithmic classification approach and the study-assigned clinical diagnosis on follow-up, using receiving operating characteristic (ROC) curve analysis to calculate the area under the curve (AUC), sensitivity, and specificity. The ROC represents a combined function of the sensitivity (true positive rate) and the specificity (true negative rate) of prediction, and the AUC is widely considered a highly informative reflection of a measure(s)' overall accuracy for predicting a disease-related outcome.

Then, the pooled data were used to examine the predictive criterion validity of associations between baseline cognitive factor scores and the algorithmic classification status 5 years post baseline, using ROC analyses. Finally, the pooled data set was used to examine the associations of the rate of change in the cognitive factors with time to progress to a future algorithmic classification of MCI, using joint survival/growth models in a structural equations modeling framework [31]. The same rate of change analyses were conducted using individual cognitive tests administered in at least three of the five studies (five memory tests and 10 nonmemory tests) in place of the cognitive factors.

2.3.4. Sensitivity analyses

These approaches to validating the algorithmic classifications introduce circularity, given that cognitive tests are, in some instances, used in both the construction of the algorithmic classifications and in the outcomes used to validate the classifications. To address this problem, for each cognitive test, algorithmic classifications were reconstructed by excluding the test and repeating all analyses using the reconstructed algorithmic classifications.

Table 2
Sample demographic characteristics and longitudinal follow-up (N = 1746)

Characteristic	Overall sample	ACS	AIBL	BIOCARD	BLSA	WRAP
Sample size	1746	360	767	301	147	171
Number of visits, mean (SD)	4.4 (2.8)	3.5 (2.3)	3.5 (0.9)	7.0 (2.9)	7.7 (4.9)	3.4 (1.1)
Years of follow-up, mean (SD)	6.6 (4.9)	4.4 (3.3)	3.9 (1.5)	11.3 (4.3)	14.4 (6.7)	8.2 (1.7)
Age at recruitment, mean (SD)	64.0 (9.6)	60.4 (8.1)	70.0 (7.0)	58.1 (8.6)	63.5 (10.1)	55.3 (5.9)
Years of education, mean (SD)	16.2 (2.5)	16.0 (2.5)	14.7 (1.6)	17.0 (2.4)	16.9 (2.2)	15.6 (2.6)
Female sex, N (%)	1033 (59.2)	231 (64.2)	439 (57.3)	177 (58.8)	72 (49.0)	114 (66.7)
White race, N (%)	1655 (94.8)	313 (86.9)	767 (100)	310 (96.6)	117 (79.6)	168 (98.2)

Abbreviations: ACS, Adult Children Study; AIBL, Australian Imaging, Biomarker, and Lifestyle; BLSA, Baltimore Longitudinal Study of Aging; SD, standard deviation; WRAP, Wisconsin Registry for Alzheimer's Prevention.

3. Results

The mean age at baseline in each study was between 55 years (WRAP) and 70 years (AIBL), with participants as young as 40 years in several cohorts. Most participants were female and non–Hispanic white (Table 2). Longitudinal cognitive data follow-up spanned an average of 3.9 years (AIBL) to 14.4 years (BLSA). The majority of the 1746 participants had data for every cognitive test administered in a study (Table 1) (of note, the ACS had less data available for Logical Memory, Boston Naming, Digit Symbol Substitution, and fluency measures because, by design, those tests were only administered to the subset of participants older than 65 years at baseline). Each study administered list learning or story recall tests to assess episodic memory, for a total of three to seven episodic memory test variables per study. Each study administered eight to 11 nonmemory tests.

3.1. Concurrent criterion validity

Table 3 summarizes the number of cases with MCI on follow-up in each study using the study-specific consensus diagnoses ("Study cases" in Table 3) and the three algorithmic classifications ("Algorithm cases" in Table 3).

Compared to the study-assigned diagnosis (N = 187), the algorithmic classification based on "CDR only" or "cognition only" classified more participants as MCI (N = 245 and N = 830, respectively). The algorithm based on combination of CDR and cognition yielded a similar number of cases (N = 183) compared to the study-assigned diagnoses. The incidence of MCI based on the CDR and cognition algorithm was 15.3 cases per 1000 person-years (Table 3).

To complement Table 3, Fig. 1 illustrates the time course to algorithmic classifications of MCI using Kaplan-Meier survival curves for each formulation of the algorithmic classification. Based on these curves, requiring both CDR and cognition is more conservative than requiring either alone. Using this formulation, those who progressed to MCI did so evenly over the first 20 years of study follow-up.

For a closer inspection of agreement between algorithmic classifications and study-assigned diagnoses at the unit of individual visits, Table 4 provides results from ROC analyses by data set for each formulation of the algorithmic classification. The AUC for the algorithmic classification based on cognitive tests only was lower in each data set compared to other approaches; it provided poor specificity (47%–68%) in each data set due to larger numbers of false positive cases. The AUC of the algorithmic classification based on CDR

Table 3
Comparison of study-assigned diagnoses to algorithmic classifications of mild cognitive impairment by data set (N = 1746)

			CDR + Cognition			CDR only			Cognition only		
Data set	N	Study cases	Algorithm cases	Progressors based on algorithmic classification	Incidence (cases per 1000 person-years)	Algorithm cases	Progressors based on algorithmic classification	Incidence (cases per 1000 person-years)	Algorithm cases	Progressors based on algorithmic classification	Incidence (cases per 1000 person-years)
BIOCARD	301	74	89	70	20.4	114	95	27.7	257	148	43.2
ACS	360	28	19	14	8.7	28	19	11.8	193	124	77.3
WRAP	171	5	16	13	8.7	21	21	14.0	110	83	55.5
BLSA	147	15	18	11	4.8	36	33	14.3	117	75	32.6
AIBL	767	65	103	75	24.2	107	77	24.8	607	400	129.1
Total	1746	187	245	183	15.3	306	245	20.5	1284	830	69.6

Abbreviations: ACS, Adult Children Study; AIBL, Australian Imaging, Biomarker, and Lifestyle; BLSA, Baltimore Longitudinal Study of Aging; CDR, Clinical Dementia Rating; WRAP, Wisconsin Registry for Alzheimer's Prevention.

NOTE. The study diagnosis is the number of individuals diagnosed as having MCI in each study, determined via study-specific adjudication procedures. An algorithm case is a participant with at least one study visit meeting criteria for algorithmic classification. A "progressor based on algorithmic classification" is a participant who was cognitively normal at baseline and had two consecutive visits with impaired performance based on the algorithm.
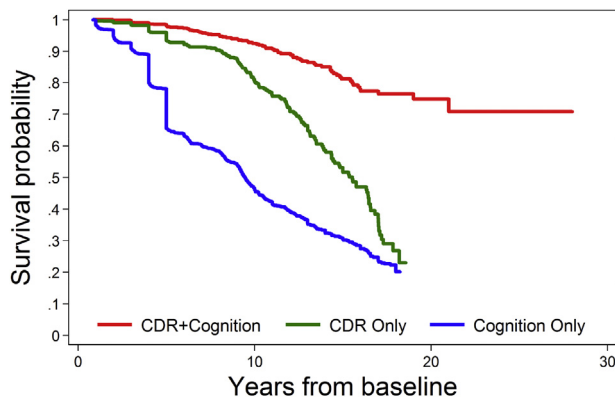
Fig. 1. Kaplan-Meier survival curves of time to algorithmic classifications of mild cognitive impairment (MCI) (N = 1746). These plots show time to progression from normal cognition until progression to MCI using algorithmic classifications based on Clinical Dementia Rating (CDR) + cognition (red), CDR only (green), and cognition only (blue). A progressor is a participant who was cognitively normal at baseline and had two consecutive visits with impaired performance based on the algorithm.

only ranged from 78% to 99%; for the majority of the sites, it was lower in comparison to that for the combination of the cognitive tests and the CDR. The AUC for the algorithm based on the cognitive tests and the CDR combined ranged from 85% to 95%. In general, using both cognitive tests and the CDR together considerably reduced false positive cases at the expense of a somewhat higher number of false negative cases.

3.2. Predictive criterion validity

Table 5 provides associations of baseline cognitive factors with algorithmic classification 5 years later for participants with at least 5 years of follow-up data available (N = 1300). The algorithmic classification using cognitive data and CDR combined tended to outperform other approaches in terms of AUC, sensitivity, and specificity for each cognitive factor score, indicating that combining cognitive data and CDR yielded better predictive criterion validity. At the level of individual cognitive tests, two tests of episodic memory (Logical Memory and CVLT) performed the best, as they produced the highest AUC, sensitivity, and specificity values compared to the other individual tests (Supplementary Table 5).

Table 6 provides longitudinal associations of cognitive trajectories with progression to each algorithmic-based classification, using joint survival/growth curve models. Steeper cognitive decline in all factor scores was associated with elevated risk of progression to MCI during follow-up. As with earlier results, the combined algorithm based on cognitive data and the CDR outperformed other approaches. Using this algorithm, steeper cognitive decline in all factor scores was associated most strongly with elevated risk of progression to MCI during follow-up. The strongest associations were seen for individual tests of episodic memory and executive function, including Logical Memory (immediate and delayed recall), CVLT (immediate, short-, and long-delay recall), Trails A and B, and animal fluency (Supplementary Table 6).

Table 4
Criterion validity of formulations of an algorithmic classification of mild cognitive impairment by data set (N = 1746)

	AUC	Sensitivity	Specificity	Kappa	N	TP	FP	FN	TN
ACS									
Cognition and CDR	0.86	0.72	1.00	0.83	1249	39	0	15	1195
CDR only	0.99	0.98	1.00	0.99	1249	53	0	1	1195
Cognition only	0.71	0.74	0.68	0.10	1249	40	388	14	807
AIBL									
Cognition and CDR	0.95	0.92	0.97	0.64	2671	83	80	7	2501
CDR only	0.94	0.92	0.96	0.60	2671	83	95	7	2486
Cognition only	0.74	1.00	0.47	0.06	2671	90	1356	0	1225
BIOCARD									
Cognition and CDR	0.92	0.85	0.99	0.87	1954	176	17	31	1730
CDR only	0.97	0.96	0.98	0.88	1954	199	41	8	1706
Cognition only	0.73	0.89	0.56	0.18	1954	184	760	23	987
BLSA									
Cognition and CDR	0.85	0.80	0.90	0.18	317	4	31	1	281
CDR only	0.78	0.80	0.76	0.07	317	4	74	1	238
Cognition only	0.82	1.00	0.63	0.05	317	5	115	0	197
WRAP									
Cognition and CDR	0.86	0.80	0.92	0.14	575	4	43	1	527
CDR only	0.83	0.80	0.86	0.08	575	4	78	1	492
Cognition only	0.80	1.00	0.60	0.03	575	5	226	0	344

Abbreviations: ACS, Adult Children Study; AIBL, Australian Imaging, Biomarker, and Lifestyle; AUC, area under the curve; BLSA, Baltimore Longitudinal Study of Aging; CDR, Clinical Dementia Rating; FN, false negative; FP, false positive; ROC, receiving operating characteristic; TN, true negative; TP, true positive; WRAP, Wisconsin Registry for Alzheimer's Prevention.

NOTE. Each visit for each participant was a different data point used in ROC analyses.

Table 5
Relationship of baseline cognitive factors to likelihood of progression to algorithmic classification of mild cognitive impairment 5 years later (N = 1300)

Algorithm and cognitive factor	AUC	Sensitivity	Specificity
Cognition and CDR			
General cognitive performance	0.79	0.74	0.71
Memory	0.77	0.67	0.74
Executive factor	0.70	0.67	0.66
CDR only			
General cognitive performance	0.71	0.74	0.61
Memory	0.69	0.62	0.68
Executive factor	0.66	0.67	0.61
Cognition only			
General cognitive performance	0.74	0.73	0.62
Memory	0.72	0.68	0.64
Executive factor	0.65	0.57	0.66

Abbreviations: AUC, area under the curve; CDR, Clinical Dementia Rating; ROC, receiving operating characteristic.

NOTE. These ROC analyses show the algorithmic classification based on both cognitive factors and CDR 5 years after baseline and are marginally more associated with baseline cognitive test performance than other algorithmic classifications.

3.3. Sensitivity analyses

Sensitivity analyses in which algorithmic classifications were tested by omitting one test and comparing the results to the remaining tests yielded no change in inferences.

Table 6
Relationship of rate of change in cognitive factors to likelihood of progression to algorithmic classification of mild cognitive impairment (N = 1746)

Factor	Hazard ratio (95% confidence interval)	Z-statistic
Cognition and CDR		
General cognitive performance	0.30* (0.23, 0.40)	−8.57
Memory	0.31* (0.24, 0.41)	−8.29
Executive functioning	0.30* (0.19, 0.46)	−5.30
CDR only		
General cognitive performance	0.36* (0.29, 0.45)	−9.18
Memory	0.40* (0.32, 0.50)	−8.27
Executive functioning	0.39* (0.28, 0.54)	−5.59
Cognition only		
General cognitive performance	0.45* (0.38, 0.54)	−8.89
Memory	0.51* (0.43, 0.59)	−8.50
Executive functioning	0.48* (0.37, 0.62)	−5.62

Abbreviation: CDR, Clinical Dementia Rating.

NOTE. Results are from joint survival/growth models of the association between rate of change in cognitive factors and time to algorithmic classification. As indicated by hazard ratios less than 1.0, shallower rates of cognitive decline are associated with less risk of progression to mild cognitive impairment based on each algorithmic classification. The algorithmic classification based on CDR and the general cognitive factor shows the strongest associations because the hazard ratios are furthest away from 1.0 and the z-statistics of the tests are larger than for other approaches.

*P < .05.

4. Discussion

A classification algorithm was developed to identify individuals who were cognitively normal at baseline but who subsequently progressed to MCI, using data from five individual studies. Results suggest the combination of the CDR and age-adjusted cut-points on cognitive tests derives a more reliable classification of MCI, as compared with using either data alone. A 1 SD cutoff for each of the cognitive tests was chosen after evaluating several alternatives. Overall, compared to a 1 SD cut-point, the benefit to sensitivity was only negligibly better using a 0.5 SD but was adversely affected using a 1.5 SD cut-point. Specificity of the algorithm in each data set was largely unaffected by the cut-point used on individual cognitive tests. Additionally, longitudinal information was required to define a case of progression, which has been demonstrated in other studies to better differentiate cognitively normal individuals who progressed over time to the symptomatic phase of disease [29,30].

The AUC for the algorithm using cognitive tests only ranged from 71% to 82%, whereas for the combination of the CDR and cognition, the AUC was 85% to 95%. It is noteworthy that the AUC for the algorithm using only the CDR had the broadest range across the sites (78%–99%), suggesting there may be site differences in the way in which the CDR was implemented during the diagnostic process.

The validity of this approach was further established by the significant relationship between the classification of MCI based on this algorithmic combination and baseline levels of, and rates of change in, factor scores for general cognitive performance, memory, and executive function. This latter finding is consistent with several previous single-site studies in which the outcome was progression to MCI [21,32–35], supporting the validity of the algorithmic classification developed here for application across multiple studies.

The algorithm described here provides a method of combining clinical and cognitive data across individual longitudinal studies of cognitively normal individuals to generate a valid, reliable outcome classification of MCI for those with preclinical AD. This approach may also be useful in clinical trials aimed at those thought to have preclinical AD in which an outcome is progression to MCI, since it provides an alternative to implementation of a time-consuming adjudication panel. Because the approach requires serial cognitive testing and CDR ratings, it may also be applicable to other cohort studies of initially cognitively normal adults with diverse ages, genetic dispositions, and other characteristics. Clinical decision-making relies on many factors including cognitive testing, proxy reports, and differential diagnosis using a medical record. Although our algorithm does not take all these factors into consideration, it appears to reflect current approaches to clinical classification applied in research settings. Caution should be exercised in applying these or similar criteria to patients in clinical settings.

Strengths of the study are inclusion of well-characterized samples with long follow-up, detailed neuropsychological test batteries, and experienced clinicians. Additional strengths include data harmonization approaches based on state-of-the-art methods. A limitation of the study is the heterogeneity in cognitive tests administered across the studies, necessitating an assumption that low performance on one test is as indicative of impairment as low performance on another test. Specific assignments of impairment may have been different if data on different tests had been available. Another major limitation is the lack of autopsy data on the subjects classified as having MCI with which to truly evaluate criterion validity; this study relied on study-assigned diagnoses which themselves are imperfect and susceptible to information biases [35]. A further limitation is that while this study's approach for defining progression takes into account baseline cognitive status and uses cutoff scores that differ by age, the algorithmic classification's assignment at any single time point does not consider change from a previous point under observation. We believe that endeavoring to determine how much change from baseline is clinically significant would place an additional criterion on the data that are not recommended in either the Petersen criteria or the NIA-AA criteria. Taken from another perspective, the algorithmic approach presented here might be considered as a strength, since one could argue that we achieved high validity for the algorithmic approach without adding an additional dimension, which might be seen by some as problematic. A final study limitation is generalizability of the pooled sample: all studies included biomarker procedures, so participants in them may not be representative of the general population.

In conclusion, this study showed that serial cognitive and clinical data from cognitively normal individuals that differ across data sets can be leveraged to derive a common classification algorithm for progression to MCI across five longitudinal studies. These classifications were associated with poorer cognitive performance at baseline and greater rate of cognitive decline over time.

Establishing a method for diagnostic harmonization across current cohorts and applying methods for combining the cognitive data are the first essential steps in providing statistical power necessary to examine the relationships between biomarker levels at baseline and their relationship to progression to MCI, understanding who progresses and when people are changing on what variables, as well as for analyzing lifestyle factors that influence rates of cognitive decline over time. It is anticipated that these analyses will uncover key genetic, social, and biological factors that influence progression during preclinical AD.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.dadm.2017.05.003.

RESEARCH IN CONTEXT

1. Systematic review: There are few existing studies that comprehensively assess the clinical, cognitive, and biomarker profiles of preclinical Alzheimer's disease (AD) in cognitively normal middle-aged and older adults. This limitation fostered the establishment of the preclinical AD consortium, designed to combine data across existing longitudinal studies that have followed cognitively normal individuals over time.

2. Interpretation: We derived and validated a generalizable classification algorithm for mild cognitive impairment to implement across multiple studies. Results suggest both cognitive tests and Clinical Dementia Rating can be combined to obtain a reliable mild cognitive impairment classification with high specificity and sensitivity. Classifications based on this algorithm are associated with baseline cognitive performance and cognitive decline.

3. Future directions: Establishing a method for cross-cohort diagnostic harmonization will facilitate opportunities to capitalize on combined clinical, cognitive, and biomarker data, providing statistical power necessary to uncover key genetic, social, and biological factors that influence progression during preclinical AD.

References

[1] Villemagne VL, Fodero-Tavoletti MT, Masters CL, Rowe CC. Tau imaging: early progress and future directions. Lancet Neurol 2015; 14:114–24.

[2] Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28.

[3] Morris JC, Aisen PS, Bateman RJ, Benzinger TL, Cairns NJ, Fagan AM, et al. Developing an international network for Alzheimer

research: The Dominantly Inherited Alzheimer Network. Clin Investig (Lond) 2012;2:975–84.

[4] Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med 2014;6:228fs13.

[5] Sperling RA, Aisen P, 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.

[6] Sperling RA, Jack CR, Aisen PS. Testing the right target and right drug at the right stage. Sci Transl Med 2011;3:111cm33.

[7] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–8.

[8] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:270–9.

[9] Bennett DA. Editorial comment on 'Prevalence of dementia in the United States: the aging, demographics, and memory study' by Plassman et al. Neuroepidemiology 2007;29:133–5.

[10] Erkinjuntti T, Ostbye T, Steenhuis R, Hachinski V. The effect of different diagnostic criteria on the prevalence of dementia. N Engl J Med 1997;337:1667–74.

[11] Wilson RS, Weir DR, Leurgans SE, Evans DA, Hebert LE, Langa KM, et al. Sources of variability in estimates of the prevalence of Alzheimer's disease in the United States. Alzheimers Dement 2011;7:74–9.

[12] Petersen RC, Thomas R, Aisen P, Mohs R, Carrillo M, Albert M. Randomized controlled trials in mild cognitive impairment: sources of variability. Neurology 2017;88:1751–8.

[13] Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. J Alzheimers Dis 2014;42:275–89.

[14] Chan KS, Gross AL, Pezzin LE, Brandt J, Kasper JD. Harmonizing Measures of Cognitive Performance Across International Surveys of Aging Using Item Response Theory. J Aging Health 2015;27:1392–414.

[15] Gross AL, Sherva R, Mukherjee S, Newhouse S, Kauwe JSK, Munsie LM, et al, for the Alzheimer's Disease Neuroimaging Initiative, GENAROAD Consortium, and AD Genetics Consortium. Calibrating longitudinal cognition in Alzheimer's disease across diverse test batteries and datasets. Neuroepidemiology 2014;43:194–205.

[16] Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK. Calibration and validation of an innovative approach for estimating general cognitive performance. Neuroepidemiology 2014;42:144–53.

[17] Gross AL, Mungas DM, Crane PK, Gibbons LE, MacKay-Brandt A, Manly JJ, et al. Effect of education and race on cognitive decline: An integrative study of generalizability versus study-specific results. Psychol Aging 2015;30:863–80.

[18] Gross AL, Power MC, Albert MS, Deal JA, Gottesman RF, Griswold M, et al. Application of latent variable methods to the study of cognitive decline when tests change over time. Epidemiology 2015;26:878–87.

[19] Coats M, Morris JC. Antecedent biomarkers of Alzheimer's disease: The adult children study. J Geriatr Psychiatry Neurol 2005;18:242–4.

[20] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P. 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.

[21] Albert MS, Soldan A, Gottesman R, McKhann G, Sacktor N, Farrington L, et al. Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to ApoE genotype. Curr Alzheimer Res 2014;11:773–84.

[22] Shock NW, Greulich RC, Costa PT, Andres R, Lakatta EG, Arenberg D, et al. Normal Human Aging: the Baltimore Longitudinal Study of Aging. Washington, D.C: National Institutes of Health; 1984.

[23] Resnick SM, Goldszal AF, Davatzikos C, Golski S, Kraut MA, Metter EJ, et al. One-year age changes in MRI brain volumes in older adults. Cereb Cortex 2000;10:464–72.

[24] Sager MA, Hermann B, La Rue A. Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. J Geriatr Psychiatry Neurol 2005;18:245–9.

[25] Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412–4.

[26] Muthen LK, Muthen BO. Mplus User's Guide. 7th ed. Los Angeles, CA: Muthen & Muthen; 1998–2012.

[27] Bollen KA. Structural Equations with Latent Variables. New York, NY: Wiley-Interscience; 1989.

[28] McDonald R. Test theory: A unified Treatment. Mahwah, NJ: Erlbaum Associates; 1999.

[29] Collie A, Maruff P, Currie J. Behavioral characterization of mild cognitive impairment. J Clin Exp Neuropsychol 2002;24:720–33.

[30] Koscik RL, La Rue A, Jonaitis EM, Okonkwo OC, Johnson SC, Bendlin BB, et al. Emergence of mild cognitive impairment in late middle-aged adults in the wisconsin registry for Alzheimer's prevention. Dement Geriatr Cogn Disord 2014;38:16–30.

[31] McArdle JJ, Small BJ, Bäckman L, Fratiglioni L. Longitudinal models of growth and survival applied to the early detection of Alzheimer's disease. J Geriatr Psychiatry Neurol 2005;18:234–41.

[32] Amieva H, Le Goff M, Millet X, Orgogozo JM, Pérès K, Barberger-Gateau P, et al. Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. Ann Neurol 2008;64:492–8.

[33] Johnson DK, Storandt M, Morris JC, Galvin JE. Longitudinal study of the transition from healthy aging to Alzheimer disease. Arch Neurol 2009;66:1254–9.

[34] Tierney MC, Yao C, Kiss A, McDowell I. Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. Neurology 2005;64:1853–9.

[35] Wilson RS, Leurgans SE, Boyle PA, Bennett DA. Cognitive decline in prodromal Alzheimer disease and mild cognitive impairment. Arch Neurol 2011;68:351–6.

Author/s:
Gross, AL; Hassenstab, JJ; Johnson, SC; Clark, LR; Resnick, SM; Kitner-Triolo, M; Masters, CL; Maruff, P; Morris, JC; Soldan, A; Pettigrew, C; Albert, MS

Title:
A classification algorithm for predicting progression from normal cognition to mild cognitive impairment across five cohorts: The preclinical AD consortium.

Date:
2017

Citation:
Gross, A. L., Hassenstab, J. J., Johnson, S. C., Clark, L. R., Resnick, S. M., Kitner-Triolo, M., Masters, C. L., Maruff, P., Morris, J. C., Soldan, A., Pettigrew, C. & Albert, M. S. (2017). A classification algorithm for predicting progression from normal cognition to mild cognitive impairment across five cohorts: The preclinical AD consortium.. Alzheimers Dement (Amst), 8 (1), pp.147-155. https://doi.org/10.1016/j.dadm.2017.05.003.

Persistent Link:
http://hdl.handle.net/11343/256338

File Description:
published version
License:
CC BY-NC-ND