Predicting cognitive decline and conversion to Alzheimer's disease in older adults using the NAB List Learning test

BRANDON E. GAVETT,[1,2,3] AL OZONOFF,[2,4] VLADA DOKTOR,[2] JOSEPH PALMISANO,[2,5]
ANIL K. NAIR,[1,2] ROBERT C. GREEN,[1,2,6,7] ANGELA L. JEFFERSON,[1,2,8,9] AND ROBERT A. STERN[1,2,3]

[1]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts
[2]Boston University Alzheimer's Disease Center, Boston, Massachusetts
[3]Center for the Study of Traumatic Encephalopathy, Boston University, Boston, Massachusetts
[4]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts
[5]Data Coordinating Center, Boston University School of Public Health, Boston, Massachusetts
[6]Department of Medicine (Genetics), Boston University School of Medicine, Boston, Massachusetts
[7]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts
[8]Department of Medicine (Geriatrics), Boston University School of Medicine, Boston, Massachusetts
[9]Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, Massachusetts

Abstract

To validate the Neuropsychological Assessment Battery (NAB) List Learning test as a predictor of future multi-domain cognitive decline and conversion to Alzheimer's disease (AD), participants from a longitudinal research registry at a national AD Center were, at baseline, assigned to one of three groups (control, mild cognitive impairment [MCI], or AD), based solely on a diagnostic algorithm for the NAB List Learning test (Gavett et al., 2009), and followed for 1–3 years. Rate of change on common neuropsychological tests and time to convert to a consensus diagnosis of AD were evaluated to test the hypothesis that these outcomes would differ between groups (AD>MCI>control). Hypotheses were tested using linear regression models ($n = 251$) and Cox proportional hazards models ($n = 265$). The AD group declined significantly more rapidly than controls on Mini-Mental Status Examination (MMSE), animal fluency, and Digit Symbol; and more rapidly than the MCI group on MMSE and Hooper Visual Organization Test. The MCI group declined more rapidly than controls on animal fluency and CERAD Trial 3. The MCI and AD groups had significantly shorter time to conversion to a consensus diagnosis of AD than controls. The predictive validity of the NAB List Learning algorithm makes it a clinically useful tool for the assessment of older adults. (*JINS*, 2010, *16,* 651–660.)

Keywords: Memory, Dementia, Differential diagnosis, Aging, Neuropsychology, Neuropsychological tests

INTRODUCTION

Alzheimer's disease (AD) and its prodromal state, mild cognitive impairment (MCI), cause early episodic memory deficits (Budson & Price, 2005). Later in its course, AD causes abnormally rapid decline in non-memory cognitive domains and eventually leads to deterioration of most or all aspects of cognitive functioning (Yaari & Corey-Bloom, 2007). In the past decade, progress has been made toward the development of therapeutic agents with the potential to modify the disease course of AD (Salloway, Mintzer, Weiner, & Cummings, 2008). With the goal of enhancing treatment outcomes in this manner, a greater focus has been placed on identifying AD as early as possible, using a combination of biomarkers and cognitive test results (Dubois et al., 2007). When used to aid in the early diagnosis of AD, the most useful neuropsychological measures of episodic memory will be those that can (a) detect early changes in episodic memory, (b) predict future cognitive decline in extra-memory cognitive domains, and (c) estimate future risk of AD. Because factors such as age and education play a role in determining AD risk and rate of cognitive decline, neuropsychological tests should also be capable of accounting for these individual differences (Andel, Vigen, Mack, Clark, & Gatz, 2006; Hall, Derby, LeValley, Katz, Verghese, & Lipton, 2007; Jacobs et al., 1994; Mungas, Reed, Ellis, & Jagust, 2001).

In a previous study by our group (Gavett et al., 2009), the Neuropsychological Assessment Battery (NAB; Stern & White, 2003) List Learning test was shown to be sensitive to

---

Correspondence and reprint requests to: Brandon E. Gavett, Instructor of Neurology, Boston University School of Medicine, 72 East Concord Street B-7800, Boston, MA 02118-2526. E-mail: begavett@bu.edu

the episodic memory difficulties seen in AD and to accurately discriminate between cognitively normal, MCI, and AD groups. In our prior study, we used ordinal logistic regression to develop an algorithm that assigns probability values to the suspected diagnoses of control, MCI, and AD, on the basis of four components of NAB List Learning performance (immediate recall across three trials, immediate recall of a distractor trial, and short- and long-delay recall). All data from our previous study were collected at participants' *most recent* annual registry visit. Overall, the algorithm achieved 80% classification accuracy across the three groups when judged against a clinical consensus diagnosis, also based on data from the participants' *most recent* annual registry visit. However, the prior study was performed using cross-sectional data, and therefore neither addressed whether the three groups followed an expected pattern of cognitive aging over time, nor whether (in non-demented samples) the three groups were associated with different levels of future AD risk. List learning tests have previously been shown to be good predictors of future cognitive decline in older adults (Andersson, Lindau, Almkvist, Engfeldt, Johansson, & Jonhagen, 2006). The current study was performed with the goal of examining the predictive validity of the NAB List Learning regression algorithm developed by our group, in terms of predicting cognitive decline and conversion to AD.

We examined the cognitive aging trajectories, both globally and within specific cognitive domains, of individuals given a baseline classification of control, MCI, or AD, solely by the NAB List Learning algorithm, *independent of any data contributing to the clinical consensus diagnosis*. Our previous study (Gavett et al., 2009) developed and cross-validated the ordinal model based on participants' *current* status; we then retroactively applied the model to participant data that were collected at an earlier ("baseline") visit. In other words, this is not a longitudinal follow-up of participants from our previous study; rather, it is a retrospective examination of longitudinal data collected before the development of the NAB List Learning ordinal regression algorithm. Data from the baseline visits were used to establish three groups that could be "followed" to the present time using retrospective data.

It was hypothesized that the three groups would experience differing rates of change across several cognitive domains (as measured by commonly used neuropsychological instruments), with the greatest decline observed in the AD group, moderate decline in the MCI group, and the least decline in the control group. To test this hypothesis, we used linear longitudinal regression models using generalized estimating equations (GEE) to test for significant group differences in linear trend over three annual time points. We also hypothesized that the NAB List Learning algorithm would predict time to convert to a consensus diagnosis of AD in a non-demented sample, such that the AD group would have the shortest time to conversion and the control group would have the longest time to conversion over one to 3 years of follow-up. This hypothesis was tested using Cox proportional hazards models to estimate relative risk. Support of

these hypotheses would further validate the NAB List Learning test as a valuable measure in the memory assessment of older adults, for both cross-sectional and longitudinal applications.

METHOD

Prediction of Cognitive Decline

Participants

Participants were drawn from a longitudinal research registry on aging and dementia at the Boston University (BU) Alzheimer's Disease Core Center (ADCC). A description of this registry, including participant recruitment, has been provided elsewhere (Jefferson, Wong, Bolen, Ozonoff, Green, & Stern, 2006). Participants were eligible for inclusion in the present study if, between 2005 and 2009, they had attended at least three annual registry visits as part of the Uniform Data Set (UDS; Weintraub et al., 2009), as set forth by the National Alzheimer's Coordinating Center. An additional criterion for inclusion was the availability of NAB List Learning data from the first of these three visits (herein referred to as the "baseline" visit or Time 0). Because many participants' very first registry visit took place prior to the implementation of the UDS, not all participants were administered the NAB List Learning test at their very first registry visit; the mean number of visits before a baseline NAB List Learning test score was available was 3.6 annual visits ($SD = 1.5$; range = 1–9). Because the current study is focused on AD-related cognitive decline, individuals with non-AD dementias at baseline were excluded. Individuals with MCI and AD were not excluded, resulting in a sample of 251 individuals with consensus diagnoses of control (without self- or informant-complaint $n = 80$; with self- or informant-complaint, $n = 24$), MCI ($n = 59$), AD (Probable, $n = 10$; Possible, $n = 4$), and other classifications reflecting non-dementia diagnostic ambiguity (e.g., meeting criteria for MCI with the exception of no cognitive complaint, $n = 74$). See Table 1 for a further breakdown of the participant characteristics for this sample.

Although data from many of the same participants as the Gavett et al. (2009) study were used, the current study classified participants based on their *baseline* visit (i.e., Time 0), while the Gavett et al. (2009) study classified participants based on their *most recent* visit (e.g., Time 2). The retroactive group assignment used for the current study was independent of the participants' test performance and consensus diagnosis at Time 2. Overall, the current sample of 251 participants contains data from 98 individuals who participated in the Gavett et al. (2009), study, a 39.0% overlap.

Procedure

The BU ADCC Research Registry data collection procedures were approved by the BU Institutional Review Board. Baseline visit data were used retrospectively to divide participants into three groups on the basis of their NAB List Learning test

Table 1. Sample and demographic characteristics in the present study

	Study sample	
	Prediction of cognitive decline	Prediction of conversion to AD
N	251	265
Sample characteristics	All non-demented participants and those with AD, based on consensus diagnosis.	All non-demented participants, based on consensus diagnosis.
Evaluation period	Three annual visits	Two to four annual visits
	M = 913.7 days, SD = 183.8	M = 848.3 days, SD = 239.4
Age (years)		
Range	55–93	55–93
Mean (SD)	73.2 (7.6)	73.1 (7.7)
Education (years)		
Range	7–21	7–21
Mean (SD)	15.9 (2.9)	15.6 (2.9)
Sex (% Women)	60%	62%
Race		
% White	76%	74%
% Black/African American	23%	26%

Note. AD = Alzheimer's disease.

performance. Demographically corrected (for age, sex, and education) T-scores from the NAB List Learning test were entered into the ordinal regression algorithm, which yields probability values for the suspected diagnoses of control, MCI, and AD (see Appendix 1 for details). Participants were assigned to the group with the largest probability value, and group assignment was made independently of other visit data, including the participants' clinical consensus diagnoses. Similarly, clinicians in the consensus conference were unaware of the results of the NAB List Learning ordinal regression algorithm groupings. This procedure assigned 167 participants to the control group, 58 to the MCI group, and 26 to the AD group. The concordance between the NAB List Learning groupings at Time 0 with actual consensus diagnosis at Time 0 was 89.4% for controls, 50.1% for MCI, and 85.7% for AD.

We evaluated group differences in rate of change on several commonly used neuropsychological test variables, representing one global and five specific cognitive domains, over the course of three visits. Variables of interest included the Mini-Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975) as an indicator of global cognitive status, animal fluency as a measure of language, Word List Recall Trial 3 from the Consortium to Establish a Battery for Alzheimer's Disease (CERAD; Morris et al., 1989) Neuropsychological Test Battery as a measure of episodic memory, Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) Digit Symbol Coding as a measure of attention and processing speed, Hooper Visual Organization Test (VOT; Hooper, 1983) as a measure of visuospatial functioning, and Trail Making Test-Part B (Reitan & Wolfson, 1993) as a measure of executive functioning. Although the UDS contains several additional variables, these six variables were chosen because they provide a valid representation of each cognitive domain and they are some of the most

commonly used tests in cognitive aging research and clinical practice. While other NAB variables were also available for examination (i.e., they had been administered to the participants, but had not been used to establish a consensus diagnosis), our goal was to investigate the utility of this particular diagnostic algorithm, which does not exist for other tests or variables in the NAB.

Statistical analysis

For each outcome variable, we used longitudinal regression models to determine whether differential rate of change was observed over three annual time points (Time 0, 1, and 2) between the three groups (control, MCI, and AD), with an *a priori* significance level of .05 for each test. Six regression models were fit in SPSS (version 15.0; SPSS, Chicago, IL) using GEE, which has advantages over traditional regression methods when modeling correlated longitudinal data (Hamilton et al., 2008). Each model included group, time, and (group × time) interaction terms as main effects. Fixed effects terms for age and education level were also included to control for the potential confounding effects of these covariates on rate of change over time. Because not all participants were administered all tests at each visit, available sample sizes for each model differed (see Table 2).

Prediction of Conversion to AD

Participants

For the survival analysis, participants were selected from the longitudinal research registry described above, with the following inclusion and exclusion criteria. Participants administered the NAB List Learning test at UDS baseline and with at least 1 additional year of UDS follow-up data were included. Because we were interested in determining

Table 2. Baseline cognitive test scores by NAB List Learning algorithm-defined group

Test Variable	Control			MCI			AD		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
MMSE	135	29.1*	1.1	51	28.6	1.6	21	25.5*	5.7
Animal fluency	134	20.5*	6.0	51	17.9*	4.8	21	13.9*	5.5
CERAD WLR Trial 3	114	8.5*	1.4	41	7.9*	1.6	17	5.8*	1.5
WAIS-R Digit Symbol	133	46.9*	11.8	49	44.2	10.2	18	34.8*	12.0
Hooper VOT	65	24.5	4.6	21	24.8	2.1	11	22.7	4.3
Trails B	129	78.0*	43.5	47	112.4*	61.2	17	137.5*	85.1

Note. NAB = Neuropsychological Assessment Battery; MCI = Mild cognitive impairment; AD = Alzheimer's disease; MMSE = Mini-Mental Status Examination; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; WLR = Word List Recall; WAIS-R = Wechsler Adult Intelligence Scale-Revised; VOT = Visual Organization Test.
 * = Significant group difference within test ($p < .05$).

rates of conversion to AD, participants with a baseline clinical consensus diagnosis of AD or any other dementia were excluded, resulting in an overall sample size of 265 non-demented individuals. The breakdown of baseline consensus diagnosis groups in this sample is as follows: control (without self- or informant-complaint, $n = 87$; with self- or informant-complaint, $n = 26$), MCI ($n = 77$), and other diagnoses reflecting nondementia diagnostic ambiguity (i.e., meeting criteria for MCI with the exception of a cognitive complaint, $n = 75$). See Table 1 for more information about the participant characteristics in this sample.

Procedure

As described above, the NAB List Learning algorithm was retroactively applied to baseline data to classify participants into one of three groups: control, MCI, or AD. This assignment classified 182 participants as control, 64 as MCI, and 19 as AD. The concordance between the NAB List Learning groupings at Time 0 and the consensus diagnosis at Time 0 was 89.4% for control and 32.5% for MCI (participants with a baseline consensus diagnosis of AD were excluded). After their most recent study visit, which occurred 1 to 3 years post-baseline, participants were given a diagnosis by a consensus team of experts, including at least one board certified neurologist, neuropsychologist, and nurse practitioner, taking into account neuropsychological test results, neurological examination findings, self- and study partner-report, and medical, social, and occupational histories. NINCDS-ADRDA criteria (McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984) were used by the consensus team to diagnose AD and Petersen/Working Group criteria (Winblad et al., 2004), were used to diagnose MCI. The consensus team relied on a battery of neuropsychological test scores for diagnosis, including the tests used as outcome variables in the current study, the tests that make up the UDS (Weintraub et al., 2009), letter fluency (FAS), and WMS-R Visual Reproduction (Wechsler, 1987). Diagnoses were made *independent* of NAB List Learning performance, and the consensus team was blinded to the NAB List Learning algorithm classifications used in the present study. Follow-up status was categorized as either "converted to AD" or "not

converted to AD" based on one or more follow-up evaluations. Individuals who, at follow-up, were or were not diagnosed with AD were assigned the status "converted to AD" or "not converted to AD" respectively.

Of the 265 participants in this sample, 103 were included in the Gavett et al. (2009) study, which represents 38.9% of the current sample. Because of this overlap, we repeated the Cox survival analysis after removing the 103 participants whose data were used to create the original ordinal model. This was done to ensure that our final criterion (a follow-up clinical consensus diagnosis of AD) was not biased by overlap between participants in this study and our previous study.

Statistical Analysis

Time to conversion was defined as the days from baseline visit to conversion (the date of the visit that led to the consensus diagnosis of AD). Times to conversion for subjects that never received a diagnosis of AD were treated as right-censored at the time in days from baseline to last visit. Survival curves for time to conversion were plotted using the Kaplan-Meier estimator, and the log rank test was used to compare survival times between groups. Cox proportional hazards models were used to estimate hazards ratios (HR) for AD and MCI groups relative to controls, controlling for the effects of age and education on time to conversion to AD.

RESULTS

Prediction of Cognitive Decline

Baseline cognitive test scores are presented in Table 2. For all variables other than the Hooper VOT, significant differences were found between the control and AD groups at baseline; in addition, significant differences were found between the MCI group and the other two groups on animal fluency, CERAD Trial 3, and Trails B at baseline.

The sample sizes reported in Table 2 correspond to the sample sizes used in the longitudinal regression models. Differences in rates of change between the three groups, before adjusting for age and education, are shown in Figure 1. Table 3
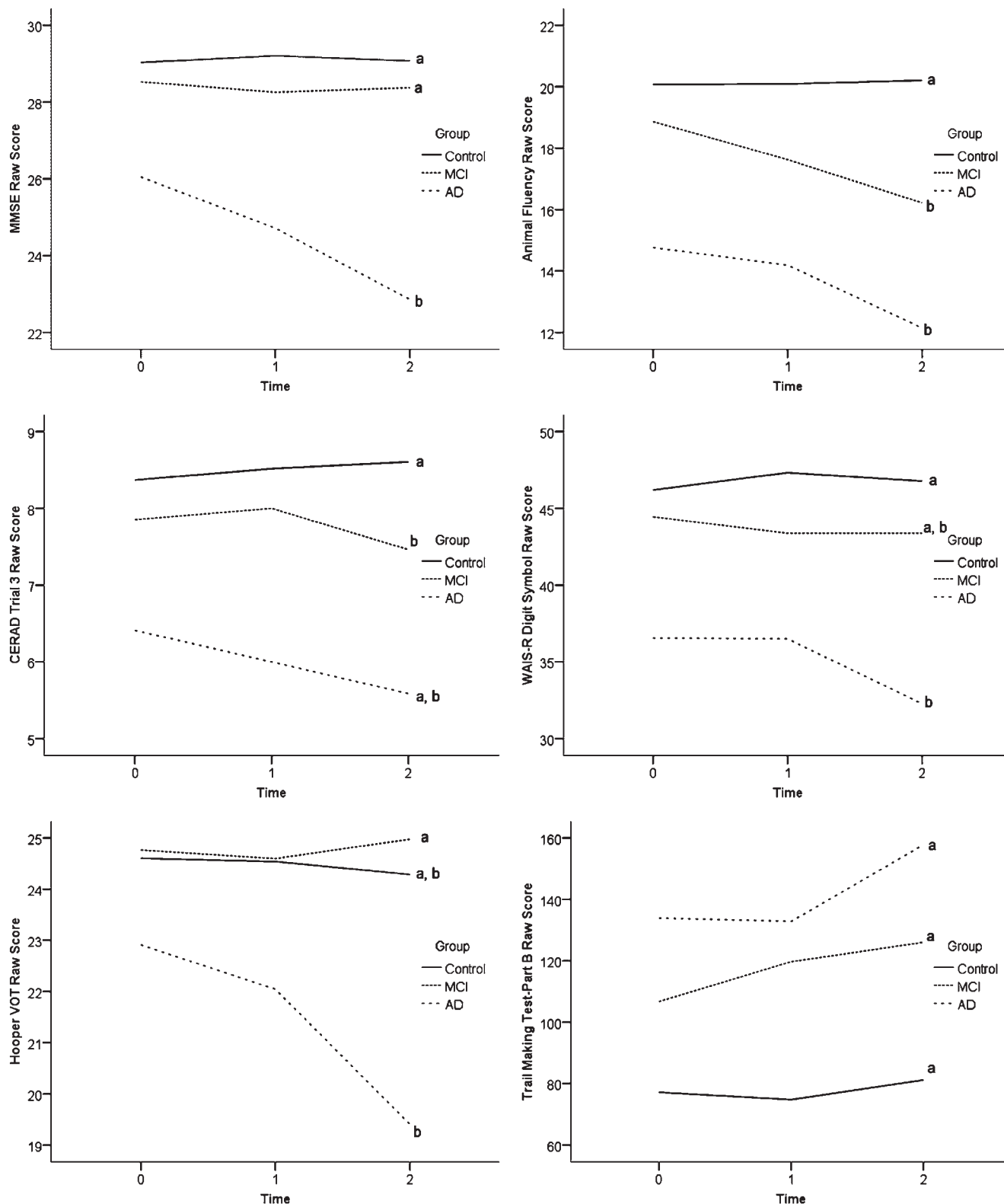
Fig. 1. Neuropsychological Assessment Battery (NAB) List Learning algorithm-defined Group (control, mild cognitive impairment [MCI], and Alzheimer's disease [AD]) performance over time on six cognitive tests measuring global cognitive functioning (Mini-Mental Status Examination [MMSE]), language (animal fluency), episodic memory (CERAD WLR Trial 3), attention/processing speed (Digit Symbol), visuospatial functioning (Hooper Visual Organization Test [VOT]), and executive functioning (Trails B). Lines represent data uncorrected for age and education. Lines marked with the same letter ("a" or "b") do not differ significantly in linear slope after correcting for age and education in the longitudinal models.

Table 3. Overall regression model effects for the six cognitive outcome measures

Test variable	Age (years)		Education (years)		Time		Group		Group x Time	
	Wald χ^2 (df)	p	Wald χ^2 (df)	p	Wald χ^2 (df)	p	Wald χ^2 (df)	p	Wald χ^2 (df)	p
MMSE	5.4* (1)	.02	30.1** (1)	< .01	11.4** (1)	< .01	16.2** (2)	< .01	12.1** (2)	< .01
Animal fluency	20.3** (1)	< .01	39.4** (1)	<.01	24.4** (1)	< .01	21.3** (2)	< .01	22.3** (2)	< .01
CERAD WLR Trial 3	30.1** (1)	< .01	7.2** (1)	.01	2.6 (1)	.11	31.9** (2)	< .01	8.0* (2)	.02
WAIS-R Digit Symbol	37.6** (1)	< .01	36.3** (1)	< .01	5.1* (1)	.02	13.4** (2)	< .01	8.5* (2)	.01
Hooper VOT	32.2** (1)	< .01	3.5 (1)	.06	3.8 (1)	.05	1.6 (2)	.44	4.2 (2)	.12
Trails B	23.9** (1)	< .01	13.1** (1)	< .01	7.7** (1)	.01	24.8** (2)	< .01	4.6 (2)	.10

Note. MMSE = Mini-Mental Status Examination; CERAD = Consortium to Establish a Registry for Alzheimer's disease; WLR = Word List Recall; WAIS-R = Wechsler Adult Intelligence Scale-Revised; VOT = Visual Organization Test.
* = $p < .05$.
** = $p < .01$.

summarizes the results of the overall regression models, which covaried for the effects of age and education on rate of change. These results indicate significant group differences in the linear rate of cognitive decline over time on four of the six variables (MMSE, animal fluency, CERAD Trial 3, and Digit Symbol). The parameter estimates in Table 4 reveal pairwise group differences in linear rate of cognitive decline. Significant differences in linear trend between control and AD groups can be seen on the MMSE, animal fluency, and WAIS-R Digit Symbol, with the AD group declining more rapidly on each test. The linear rate of cognitive decline was significantly greater in the MCI group compared with the control group on animal fluency and CERAD Trial 3. The AD group declined more rapidly than the MCI group on the MMSE and the Hooper VOT.

Prediction of Conversion to AD

Participants were followed for an average of 848 days (*SD* = 239.4 days). There was no significant difference in length of follow up between groups, $F(2,262) = 1.79$, not significant. The Kaplan-Meier survival curve shown in Figure 2 illustrates the cumulative survival fractions of each group over a 4-year period, without adjusting for age and education. There was a significant between-group difference in survival time, log rank χ^2 (df = 2) = 33; p < .01. Survival rates for each group are provided in Table 5; the proportion converting to a consensus diagnosis of AD was largest in the AD group (21.1%), followed by the MCI group (15.6%) and the control group (0.5%).

The results of the Cox proportional hazards model are reported in Table 6. Unless otherwise noted, those classified as controls at baseline assessment were used as the comparison group for all hazards ratios. The overall Cox model explained a significant proportion of variation in survival time, χ^2 (df = 4) = 48.2; p < .01. Age and group were both found to significantly decrease the time to consensus AD diagnosis; in contrast, education did not have a significant impact (HR = 0.9; 95% CI = 0.7–1.1). Both AD and MCI groups were

associated with shorter times to receive a consensus diagnosis of AD. In fact, being classified as AD by the NAB List Learning algorithm was associated with a 73-fold (95% CI = 7.5–704.4) increased hazard of AD diagnosis. For the group classified as MCI by the NAB List Learning algorithm, the estimated hazards ratio was 45.2 (95% CI = 5.6–362.4). There was no significant difference in conversion times between the MCI and AD groups (HR = 1.6; 95% CI = 0.5–5.3; reference group = MCI).

These survival analyses were based, in part, on participant data that were also used to create the original ordinal model (Gavett et al., 2009). Broken down by baseline diagnosis, the percentage of sample overlap was 62.8% for the control group and 21.1% for the MCI group (participants with a baseline diagnosis of AD were excluded). Because of the high proportion of sample overlap in the control group, we repeated the above Cox survival analysis methods using only those participants who were not part of the original study.

At baseline, this new independent sample contained 162 individuals. This sample (M = 15.3 years; SD = 3.1) was slightly less educated than the participants who were excluded due to overlap (M = 16.1 years; SD = 2.5); t(263) = −2.1; p < .05, and, at baseline, contained a smaller proportion of control participants (26%) than the excluded sample (69%) and a larger proportion of participants with MCI (74% *vs.* 31%). There were no between-groups differences in age or sex. At baseline, the NAB List Learning algorithm classified the 162 members of the independent sample into the three experimental subgroups as follows: control, n = 97 (60%); MCI, n = 49 (30%); and AD, n = 16 (10%), which differed from the excluded sample in terms of the overall frequency distribution (control, n = 85 [82.5%]; MCI, n = 15 [14.6%]; and AD, n = 3 [2.9%]); χ^2 (df = 2) = 15.4; p < .01. This new independent sample had an overall concordance rate of 48.6% between the baseline NAB List Learning algorithm classifications and the baseline clinical consensus diagnosis.

At follow-up, 0 of 97 (0%) controls converted to AD; 6 of 49 (12.2%) of the MCI group converted to AD, and 4 of 16

Table 4. Pairwise Group × Time linear slope differences

Test Variable	MCI vs. Control				AD vs. Control				AD vs. MCI			
	B	95% CI for B	Wald χ^2 (df)	p	B	95% CI for B	Wald χ^2 (df)	p	B	95% CI for B	Wald χ^2 (df)	p
MMSE	−0.1	−0.4 to 0.02	0.6 (1)	.43	−1.6	−2.5 to −0.7	11.7** (1)	<.01	−1.5	−2.5 to −0.6	9.8** (1)	<.01
Animal fluency	−1.4	−2.1 to −0.7	17.1** (1)	<.01	−1.4	−2.2 to −0.5	10.0** (1)	<.01	0.1	−0.9 to 1.0	0.0 (1)	.98
CERAD WLR Trial 3	−0.3	−0.6 to −0.04	5.1* (1)	.03	−0.5	−1.1 to 0.01	3.7 (1)	.06	−0.2	−0.8 to 0.4	0.5 (1)	.47
WAIS-R Digit Symbol	−0.8	−1.7 to 0.05	3.5 (1)	.06	−2.4	−4.4 to −0.5	6.1* (1)	.01	−1.6	−3.6 to 0.4	2.5 (1)	.12
Hooper VOT	0.3	−0.4 to 0.9	0.7 (1)	.40	−1.6	−3.3 to 0.1	3.3 (1)	.07	−1.9	−3.6 to −0.1	4.1* (1)	.04
Trails B	7.6	−0.7 to 15.9	3.2 (1)	.07	9.9	−4.8 to 24.7	1.7 (1)	.19	2.3	−14.0 to 18.7	0.1 (1)	.78

Note. MCI = Mild cognitive impairment; AD = Alzheimer's disease; CI = confidence interval; MMSE = Mini-Mental Status Examination; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; WLR = Word List Recall; WAIS-R = Wechsler Adult Intelligence Scale-Revised; VOT = Visual Organization Test.
* = $p < .05$.
** = $p < .01$.

Table 5. Rate of conversion to AD by NAB List Learning algorithm-defined Group over a 1- to 3-year follow-up period

Group	n	Conversion to AD		Survival Time (days)	
		n	%	M	95% CI
Control	182	1	0.5%	1284.0	1274.4–1293.7
MCI	64	10	15.6%	1097.4	1026.1–1168.7
AD	19	4	21.1%	1048.2	918.1–1178.3

Note. AD = Alzheimer's disease, NAB = Neuropsychological Assessment Battery; CI = confidence interval; MCI = mild cognitive impairment.

(25%) of the AD group converted to AD, which is similar to the results obtained in the larger sample (0.5%, 15.6%, and 21.1% for control, MCI, and AD groups, respectively). The overall Cox model fit the survival data well, χ^2 (df = 4) = 33.5; $p < .01$, but this χ^2 statistic was significantly smaller than the results of the Cox model that used the entire sample, χ^2 (df = 1) = 14.7; $p < .01$. Because of the reduced sample size in this secondary analysis, the hazards ratio estimates for the NAB List Learning MCI and AD classifications are large but extremely imprecise; for this reason they are not presented.

DISCUSSION

The NAB List Learning test and its classification algorithm have been previously shown to accurately distinguish between individuals diagnosed as control, amnestic MCI, and AD by a clinical consensus diagnostic conference (Gavett
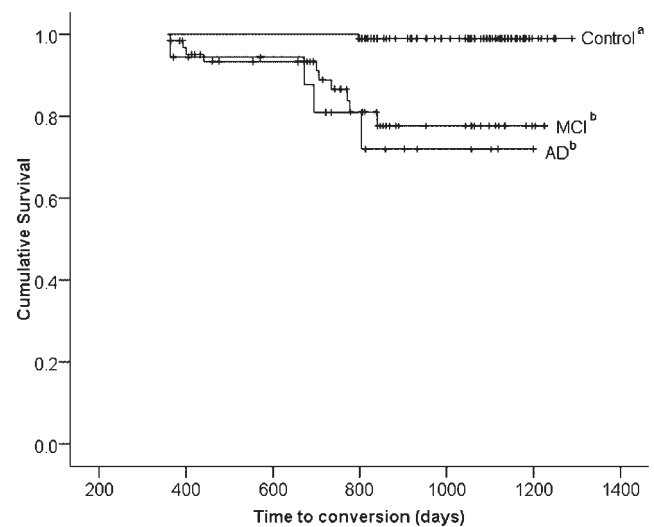
Fig. 2. Kaplan-Meier survival curve depicting time to conversion to a consensus diagnosis of Alzheimer's disease (AD), broken down by the three Neuropsychological Assessment Battery (NAB) List Learning algorithm-defined groups (control, mild cognitive impairment [MCI], and AD). Lines represent data uncorrected for age and education. Lines marked with the same letter ("a" or "b") do not differ significantly in hazards ratio after correcting for age and education in the Cox proportional hazards model. Vertical hash marks represent censored data.

Table 6. Cox proportional hazards model

Variable	B	Wald χ²	df	p	HR	95% CI for HR
Age (years)	0.13	13.0**	1	< .01	1.1	1.1–1.2
Education (years)	-0.16	2.0	1	.16	0.9	0.7–1.1
Group		14.4**	2	< .01		
MCI vs. Control	3.8	12.9**	1	< .01	45.2	5.6–362.4
AD vs. Control	4.3	13.7**	1	< .01	72.9	7.5–704.4
AD vs. MCI	0.5	0.6	1	.43	1.6	0.5–5.3

Note. HR = hazards ratio; CI = confidence interval; MCI = mild cognitive impairment; AD = Alzheimer's disease.
** = $p < .01$.

et al., 2009). The current study extends these findings to demonstrate a meaningful association between NAB List Learning algorithm classifications and future clinical course and outcome. The principal findings from the current study suggest that the NAB List Learning algorithm possesses clinically useful predictive validity. The results suggest that a NAB List Learning algorithm classification of AD was predictive of more rapid decline in language (animal fluency), attention/processing speed (WAIS-R Digit Symbol), and overall cognitive functioning (MMSE), as well as a significantly reduced time to reach a consensus diagnosis of AD, relative to those classified by the algorithm as controls. A NAB List Learning algorithm classification of MCI was predictive of more rapid decline in language (animal fluency) and episodic memory (CERAD WLR Trial 3), as well as a significantly shorter time to consensus diagnosis of AD, relative to controls. Fewer differences between groups classified as AD and MCI were apparent; the AD group declined more rapidly in overall cognitive functioning (MMSE) and visuospatial functioning (Hooper VOT) and had a higher proportion convert to AD compared with those with MCI, but there was not a significant difference in time to future AD diagnosis relative to the MCI group. Although robust and statistically significant, the hazards ratios from the Cox models for the AD and MCI groups are imprecise estimates with wide confidence intervals. However, even the lower limits of these ranges (HR = 7.5 and 5.6 for AD and MCI, respectively), suggest a clinically meaningful decrease in time to conversion relative to the control group. Surprisingly, no difference in the trajectory of cognitive decline was observed between the control group and the AD group on the CERAD, a measure of episodic memory. This may be due to floor and/or ceiling effects, with the AD group having little room to decline below the level of their attention span and the control group having little room to improve before reaching the maximum score for the test. In contrast, based on the results presented in Figure 1, it is surprising that no significant difference in slope was observed on the Hooper VOT between the control and AD groups. The lack of significance can likely be attributed to (a) the linear trends observed in Figure 1 may be misleading because they are not corrected for age and education (although the tests of significance are corrected) or (b) the smaller sample size for the Hooper VOT (see Table 2) reduced the power to detect a significant difference in trend.

In addition to establishing predictive validity, the results also appear to provide support for the face validity and convergent validity of the NAB List Learning test. The clinical manifestations of AD are strongly linked to neurofibrillary tangle (NFT) burden (Wilcock & Esiri, 1982). In the early stages of the disease, NFTs aggregate in the medial temporal lobes; as the disease progresses, neuropathological changes expand considerably beyond the medial temporal lobes into the neocortex (Arriagada, Growdon, Hedley-Whyte, & Hyman, 1992; Guillozet, Weintraub, Mash, & Mesulam, 2003). As expected based on this pattern of neuropathology, individuals with clinically diagnosed AD often show rapid decline in overall cognitive functioning and in several cognitive domains in addition to episodic memory (Yaari & Corey-Bloom, 2007). Individuals with MCI are often believed to be showing the earliest clinical manifestations of AD (Morris, 2006), and are thought to undergo more subtle changes to episodic memory and language, especially verbal category fluency (Murphy, Rich, & Troyer, 2006). The fact that the AD and MCI groups in the current study experienced cognitive changes consistent with these expectations suggests that the NAB List learning algorithm groupings are clinically meaningful.

One limitation of this study is the fact that participants with missing cognitive data were excluded from the longitudinal regression models of cognitive decline. Participants with missing data were unlikely to be excluded at random; in other words, individuals with more severe cognitive impairment were more likely to be excluded due to an inability to participate in cognitive testing. Had data from severely impaired participants been available, the current findings may have been bolstered by increasing both rate of decline and conversion to AD in the NAB List Learning algorithm-defined AD group. Given the geographic region in which the study was conducted, our samples may have higher socioeconomic status and educational experience (M > 15 years) than the general population. Our samples are representative of black and white individuals, but not of individuals from other racial and ethnic backgrounds. The outcome variable in the survival analyses was the clinical diagnosis of probable or possible AD by a consensus diagnosis team. Ideally, the gold standard would be neuropathologically confirmed AD, and as such, the accuracy of the current study is limited by the accuracy of clinical consensus diagnosis. Finally, because the results were obtained from the same longitudinal

research registry used by Gavett et al. (2009), there is overlap between the current sample and the sample that was used to develop and cross-validate the ordinal regression algorithm. However, the algorithm was developed based on participants' *most recent* annual visit and, for the current study, was retroactively applied to data from a previous visit. In addition, we redid our survival analysis using a sample that was independent from the original sample used to create the algorithm. Results revealed a significant, albeit less robust, overall model. We believe that these procedures mitigate the potential for tautological error and criterion contamination. Nevertheless, replication and extension of these findings in independent samples is necessary.

Future attempts should be made to evaluate the ability of the NAB List Learning test to predict neuropathologically confirmed AD. Because the AD and MCI groups in the current study did not differ as greatly as expected, longer follow-up periods, perhaps 10 years or more, may be necessary to determine whether these two groups, as defined by the NAB List Learning algorithm, age differently. Future research should also focus on examining whether a combination of NAB List Learning test results and an AD biomarker, such as cerebrospinal fluid tau/Aβ_{42} ratio, provides better diagnostic utility and predictive validity than either marker alone.

Although an algorithmic classification of MCI was predictive of future cognitive decline, these results were based on a subset of the entire MCI group due to missing data, and are therefore limited. This speaks to a more general point about the NAB List Learning algorithm as it pertains to a diagnosis of MCI. The algorithm appears to be considerably more accurate when participants are classified as either controls or AD. A classification of MCI should engender less confidence, both in terms of predicting future outcomes and for classification accuracy; for example, Gavett et al. (2009) reported a sensitivity of .47 for correctly identifying MCI. This conclusion appears to be consistent with the fact that MCI is better described as a state of diagnostic ambiguity, rather than a specific disease entity (Royall, 2006). Nevertheless, the results suggest that the NAB List Learning algorithm should not be relied upon as a primary tool for the diagnosis of MCI, and caution is warranted in interpreting an algorithmic MCI classification as a risk factor for future cognitive decline.

It was hypothesized that the three groups (control, MCI, and AD), derived solely on the basis of NAB List Learning performance, would undergo differential rates of cognitive decline and conversion to consensus diagnosis of AD. Because these hypotheses were supported for several outcome measures, the NAB List Learning test algorithm appears to be a valid predictor of changes in cognitive functioning and time to reach a consensus diagnosis of AD. Recently, compelling arguments have been made for eliminating the requirement for "dementia" in the clinical diagnosis of AD, instead focusing on making the diagnosis prior to full-blown dementia using a combination of biomarkers and cognitive indices (Dubois et al., 2007). This approach to early diagnosis is especially important considering the many potential disease-modifying AD drugs currently in clinical trials (Salloway et al., 2008). With the caveats about the algorithmic classification of MCI mentioned above, the NAB List Learning test can be diagnostically useful and predictive of multiple cognitive domain decline and time to conversion to AD. Therefore, it fits well within the Dubois et al. (2007) framework for progress in AD research and clinical care.

ACKNOWLEDGMENTS

REFERENCES

Andel, R., Vigen, C., Mack, W., Clark, L.J., & Gatz, M. (2006). The effect of education and occupational complexity on rate of cognitive decline in Alzheimer's patients. *Journal of the International Neuropsychological Society*, *12*, 147–152.

Andersson, C., Lindau, M., Almkvist, O., Engfeldt, P., Johansson, S.-E., & Jonhagen, M. (2006). Identifying patients at high and low risk of cognitive decline using Rey Auditory Verbal Learning Test among middle-aged memory clinic outpatients. *Dementia and Geriatric Cognitive Disorders*, *21*, 251–259.

Arriagada, P.V., Growdon, J.H., Hedley-Whyte, E.T., & Hyman, B.T. (1992). Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology*, *42*, 631–639.

Budson, A.E., & Price, B.H. (2005). Memory dysfunction. *New England Journal of Medicine*, *352*, 692–699.

Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurology*, *6*, 734–746.

Folstein, M., Folstein, S.E., & McHugh, P.R. (1975). "Mini-Mental State." A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Gavett, B.E., Poon, S.J., Ozonoff, A., Jefferson, A.L., Nair, A.K., Green, R.C., et al. (2009). Diagnostic utility of the NAB List Learning test in Alzheimer's disease and amnestic mild cognitive impairment. *Journal of the International Neuropsychological Society*, *15*, 121–129.

Guillozet, A.L., Weintraub, S., Mash, D.C., & Mesulam, M.-M. (2003). Neurofibrillary tangles, amyloid, and memory in aging and mild cognitive impairment. *Archives of Neurology*, *60*, 729–736.

Hall, C.B., Derby, C., LeValley, A., Katz, M.J., Verghese, J., & Lipton, R.B. (2007). Education delays accelerated decline on a memory test in persons who develop dementia. *Neurology*, *69*, 1657–1664.

Hamilton, J.M., Salmon, D.P., Galasko, D., Raman, R., Emond, J., Hansen, L.A., et al. (2008). Visuospatial deficits predict rate of cognitive decline in autopsy-verified dementia with Lewy bodies. *Neuropsychology*, *22*, 729–737.

Hooper, H. (1983). *Hooper Visual Organization Test (HVOT)*. Los Angeles: Western Psychological Services.

Jacobs, D., Sano, M., Marder, K., Bell, K., Bylsma, F., Lafleche, G., et al. (1994). Age at onset of Alzheimer's disease: Relation to pattern of cognitive dysfunction and rate of decline. *Neurology*, *44*, 1215–1220.

Jefferson, A.L., Wong, S., Bolen, E., Ozonoff, A., Green, R.C., & Stern, R.A. (2006). Cognitive correlates of HVOT performance differ between individuals with mild cognitive impairment and normal controls. *Archives of Clinical Neuropsychology*, *21*, 405–412.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

Morris, J.C. (2006). Mild cognitive impairment is early-stage Alzheimer disease: Time to revise diagnostic criteria. *Archives of Neurology*, *63*, 15–16.

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., et al. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*, 1159–1165.

Mungas, D., Reed, B.R., Ellis, W.G., & Jagust, W.J. (2001). The effects of age on rate of progression of Alzheimer disease and dementia with associated cerebrovascular disease. *Archives of Neurology*, *58*, 1243–1247.

Murphy, K.J., Rich, J.B., & Troyer, A.K. (2006). Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society*, *12*, 570–574.

Reitan, R.M., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). Tuscon, AZ: Neuropsychology Press.

Royall, D. (2006). Mild cognitive impairment and functional status. *Journal of the American Geriatrics Society*, *54*, 163–165.

Salloway, S., Mintzer, J., Weiner, M.F., & Cummings, J.L. (2008). Disease-modifying therapies in Alzheimer's disease. *Alzheimer's & Dementia*, *4*, 65–79.

Stern, R.A., & White, T. (2003) *Neuropsychological Assessment Battery*. Lutz, FL: Psychological Assessment Resources.

Wechsler, D. (1981). *WAIS-R manual*. New York: The Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised*. San Antonio, TX: The Psychological Corporation.

Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N.R., Chui, H., et al. (2009). The Alzheimer's Disease Centers' Uniform Data Set (UDS): The neuropsychologic test battery. *Alzheimer Disease and Associated Disorders*, *23*, 91–101.

Wilcock, G.K., & Esiri, M.M. (1982). Plaques, tangles and dementia. A quantitative study. *Journal of the Neurological Sciences*, *56*, 343–356.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.O., et al. (2004). Mild cognitive impairment – beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, *256*, 240–246.

Yaari, R., & Corey-Bloom, J. (2007). Alzheimer's disease. *Seminars in Neurology*, *27*, 32–41.

APPENDIX 1. NAB LIST LEARNING ALGORITHM

$$P_C = e^{-e^{(8.01-((T_{LLA}\times 0.024)-(T_{LLB}\times 0.046)-(T_{LLA-SD:drc}\times 0.1)-(T_{LLA-LD:drc}\times 0.027)))}}$$

$$P_{MCI} = e^{-e^{(6.08-((T_{LLA}\times 0.024)-(T_{LLB}\times 0.046)-(T_{LLA-SD:drc}\times 0.1)-(T_{LLA-LD:drc}\times 0.027)))}} - P_C$$

$$P_{AD} = 1 - P_C + P_{MCI}$$

Where

P_C = probability of a consensus diagnosis of control;

P_{MCI} = probability of a consensus diagnosis of mild cognitive impairment;

P_{AD} = probability of a consensus diagnosis of Alzheimer's disease;

T_{LLA} = List A Immediate Recall T-Score;

T_{LLB} = List B Immediate Recall T-Score;

$T_{LLA-SD:drc}$ = List A Short Delay Recall T-Score; and

$T_{LLA-LD:drc}$ = List A Long Delay Recall T-Score.