Journal of the International Neuropsychological Society (2009), **15**, 121–129.

doi:10.1017/S1355617708090176

Diagnostic utility of the NAB List Learning test in Alzheimer's disease and amnestic mild cognitive impairment

BRANDON E. GAVETT,[1] SABRINA J. POON,[1] AL OZONOFF,[2] ANGELA L. JEFFERSON,[1]
ANIL K. NAIR,[1] ROBERT C. GREEN,[1,3,4] AND ROBERT A. STERN[1]

[1]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts
[2]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts
[3]Department of Medicine (Genetics Program), Boston University School of Medicine, Boston, Massachusetts
[4]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts

(Received June 11, 2008; Final Revision October 8, 2008; Accepted October 10, 2008)

Abstract

Measures of episodic memory are often used to identify Alzheimer's disease (AD) and mild cognitive impairment (MCI). The Neuropsychological Assessment Battery (NAB) List Learning test is a promising tool for the memory assessment of older adults due to its simplicity of administration, good psychometric properties, equivalent forms, and extensive normative data. This study examined the diagnostic utility of the NAB List Learning test for differentiating cognitively healthy, MCI, and AD groups. One hundred fifty-three participants (age: range, 57–94 years; *M* = 74 years; *SD,* 8 years; sex: 61% women) were diagnosed by a multidisciplinary consensus team as cognitively normal, amnestic MCI (aMCI; single and multiple domain), or AD, independent of NAB List Learning performance. In univariate analyses, receiver operating characteristics curve analyses were conducted for four demographically-corrected NAB List Learning variables. Additionally, multivariate ordinal logistic regression and fivefold cross-validation was used to create and validate a predictive model based on demographic variables and NAB List Learning test raw scores. At optimal cutoff scores, univariate sensitivity values ranged from .58 to .92 and univariate specificity values ranged from .52 to .97. Multivariate ordinal regression produced a model that classified individuals with 80% accuracy and good predictive power. (*JINS*, 2009, *15*, 121–129.)

Keywords: Dementia, Sensitivity and specificity, Differential diagnosis, Neuropsychology, Neuropsychological tests, Memory

INTRODUCTION

Alzheimer's disease (AD) compromises episodic memory systems, resulting in the earliest symptoms of the disease (Budson & Price, 2005). Measures of anterograde episodic memory are useful in quantifying memory impairment and identifying performance patterns consistent with AD or its prodromal phase, mild cognitive impairment (MCI; Blacker et al., 2007; Salmon et al., 2002).

List learning tests are commonly used measures of episodic memory that offer a means of evaluating a multitude of variables relevant to learning and memory. Some of the more common verbal list learning tasks are the California Verbal Learning Test (CVLT; Delis et al., 1987), Auditory Verbal Learning Test (AVLT; Rey, 1941, 1964), Hopkins Verbal Learning Test (HVLT; Brandt & Benedict, 2001), and the Word List Recall test from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD; Morris et al., 1989). List learning tests have been shown to possess adequate sensitivity and specificity in differentiating participants with MCI (*Mdn*: sensitivity = .67, specificity = .86) (Ivanoiu et al., 2005; Karrasch et al., 2005; Schrijnemaekers et al., 2006; Woodard et al., 2005) and AD (*Mdn*: sensitivity = .80, specificity = .89) from controls (Bertolucci et al., 2001; Derrer et al., 2001; Ivanoiu et al., 2005; Karrasch et al., 2005; Kuslansky et al., 2004; Salmon et al., 2002; Schoenberg et al., 2006), as well as AD from MCI (*Mdn*: sensitivity = .85, specificity = .83) (de Jager et al., 2003).

Correspondence and reprint requests to: Robert A. Stern, Alzheimer's Disease Clinical and Research Program, Boston University School of Medicine, Robinson 7800, 72 East Concord Street, Boston, MA 02118-2526. E-mail: bobstern@bu.edu

The current study was undertaken to evaluate the diagnostic utility of a new list learning test in a sample of older adults seen as part of a prospective study on aging and dementia. The Neuropsychological Assessment Battery (NAB; Stern & White, 2003a, b) is a recently-developed comprehensive neuropsychological battery that has been standardized for use with individuals ages 18 to 97. It contains several measures of episodic memory, including a List Learning test similar to other commonly used verbal list learning tests. The NAB List Learning test was developed to "create a three trial learning test to avoid the potential difficulties that five trial tasks represent for impaired individuals, include three semantic categories to allow for examination of the use of semantic clustering as a learning strategy, avoid sex, education, and other potential biases, and include both free recall and forced-choice recognition paradigms" (White & Stern, 2003, p. 24).

One major benefit of the NAB includes the fact that all of its 33 subtests, together encompassing the major domains of neuropsychological functioning, are co-normed on the same large sample of individuals (n = 1448), with demographic adjustments available for age, sex, and education. This normative group contains a large proportion of individuals ages 60 to 97 (n = 841), making it particularly well suited for use in dementia evaluations. Despite psychometric validation (White & Stern, 2003), its diagnostic utility has yet to be evaluated.

For the last several years, several NAB measures have been included in the standard research battery in the Boston University (BU) Alzheimer's Disease Core Center (ADCC) Research Registry. The BU ADCC recruits both healthy and cognitively impaired older adults for comprehensive yearly neurological and neuropsychological assessments. After each individual is assessed, a multidisciplinary consensus diagnostic conference is held to diagnose each individual based on accepted diagnostic criteria. Importantly, the NAB measures have yet to be included for consideration when the consensus team meets to diagnose study participants. Therefore, the current study setting offers optimal clinical conditions (i.e., without neuropathological confirmation) for evaluating the diagnostic utility of the NAB List Learning test. In other words, NAB performance can be judged against current clinical diagnostic criteria without the tautological error that occurs when the reference standard is based on the test under investigation. Samples of participants from the BU ADCC Registry were used to evaluate the utility of the NAB List Learning test in the diagnosis of amnestic (a)MCI and AD. As the diagnostic utility of the NAB List Learning test has yet to be examined empirically, the present study was considered exploratory.

METHOD

Participants

Participant data were drawn from an existing database—the BU ADCC Research Registry—and retrospectively analyzed. Participants were recruited from the greater Boston area through a variety of methods, including newspaper advertisements, physician referrals, community lectures, and referrals from other studies. Participants diagnosed as cognitively healthy controls consisted of community-dwelling older adults, many of whom have neither expressed concern about nor been evaluated clinically for cognitive difficulties. Data collection and diagnostic procedures have been described in detail elsewhere (see Ashendorf et al., 2008 and Jefferson et al., 2006). Briefly, after undergoing a comprehensive participant and informant interview, clinical history taking (i.e., psychosocial, medical), and assessment (i.e., neurological, neuropsychological), participants were diagnosed by a multidisciplinary consensus group that included at least two board certified neurologists, two neuropsychologists, and a nurse practitioner. Of an initial pool of 490 participants, 18 were excluded from the present study because English was not their primary language. An additional 172 were excluded because they were not diagnosed as control, aMCI, or AD. Of the remaining 300 participants, 153 completed all relevant portions of the NAB List Learning test. These 153 participants comprised the current sample, from which three groups were established: controls (i.e., cognitively normal older adults), participants diagnosed with single or multiple domain aMCI (based on Winblad et al., 2004), and participants diagnosed with possible or probable AD (based on NINCDS-ADRDA criteria; McKhann et al., 1984).

The sample consisted of 93 women (60.8%) and 60 men (39.2%), ranging in age from 57 to 94 (M = 73.9; SD = 8.1). There were 128 (83.7%) non-Hispanic Caucasian participants and 25 (16.3%) African American participants. The data used in the current study were collected between 2005 and 2007 at each participant's most recent assessment, which ranged from the first to ninth visit (Mdn = 4.0) of their longitudinal participation.

Measures

NAB List Learning Test

Administration of the NAB List Learning test begins by telling the examinee to try to remember a list of 12 words that he or she is read three times (List A), followed by testing for free recall of the list after a short delay (during which the examinee is asked to recall a distractor list, List B). After a longer delay of approximately 15 min (during which other cognitive tasks are administered), free recall is again tested, as well as forced-choice recognition (see the NAB Administration, Scoring, and Interpretation Manual, Stern & White, 2003a, for more detail). Four variables were extracted from the NAB List Learning test for the current study: List A Immediate Recall, List B Immediate Recall, List A Short Delay Recall, and List A Long Delay Recall. These four variables were chosen because they are demographically corrected for age, sex, and education, are psychometrically sound, and evaluate several different aspects of learning and recall (White & Stern, 2003). Form 1 of the NAB was administered to 75 of the current participants and Form 2 (developed and

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press

shown to be equivalent to Form 1, White & Stern, 2003) was administered to 78 participants.

Procedure

The BU ADCC Research Registry data collection procedures were approved by the Boston University Medical Center Institutional Review Board. All participants provided written informed consent to participate in the study. Participants were administered a comprehensive neuropsychological test battery designed for the assessment of individuals with known or suspected dementia, including all tests that make up the Uniform Data Set (UDS) of the National Alzheimer's Coordinating Center (Beekly et al., 2007; Morris et al., 2006). Neuropsychological assessment was carried out by a trained psychometrist in a single session. The identification of cognitive impairment in each of the domains assessed (language, memory, attention, visuospatial functioning, and executive functioning) was based on BU ADCC Research Registry procedures, which defined psychometric impairment *a priori* as a standardized score (e.g., *Z*-score, T-score) of greater than or equal to 1.5 standard deviation units below appropriate normative means on one or more "primary" variables. Primary variables in the memory domain include Trial 3 and Delayed Recall from the CERAD Word List, and both Immediate and Delayed portions of the Logical Memory and Visual Reproduction subtests from the Wechsler Memory Scales-Revised (WMS-R; Wechsler, 1987). WMS-R subtests were administered according to UDS procedures (e.g., only Story A from Logical Memory is administered) and no other WMS-R subtests were used.

In addition to neuropsychological testing, participant information was also obtained *via* clinical interview with the participant and a close informant, neurological evaluation, review of medical history, and informant questionnaires.

Diagnosis

The results from the "primary" neuropsychological variables were used by the multidisciplinary consensus team, along with social and medical history, neurological examination results, and self/informant report (i.e., interviews and questionnaires), to arrive at a diagnosis for each participant. Diagnoses were made based only on information obtained during the participant's most recent visit. The NAB List Learning test was not a "primary" neuropsychological variable, and thus, was not considered for diagnostic purposes by the multidisciplinary consensus team.

Data Analysis

Univariate analyses

To examine the diagnostic utility of the individual NAB List Learning variables, we calculated test sensitivity and specificity (along with 95% confidence intervals) by conducting receiver operating characteristics (ROC) curve analyses using the demographically-corrected T-scores (which correct for age, sex, and education) for each relevant group comparison (i.e., control *vs.* impaired [aMCI or AD], control *vs.* aMCI, control *vs.* AD, and aMCI *vs.* AD). For the purposes of identifying an optimal cutoff score, we used Youden's index (Youden, 1950), which identifies the cutoff score that jointly maximizes sensitivity and specificity. ROC curve analyses were conducted for each individual NAB List Learning variable to discriminate between the various groups. After the optimal cutoff score was selected, sensitivity and specificity values were used to calculate positive likelihood ratios (PLR) and negative likelihood ratios (NLR).

Multivariate analyses

To examine the diagnostic utility of the four NAB List Learning variables when considered together, we used multivariate ordinal logistic regression, with a negative log-log link function, using the PLUM procedure in SPSS (version 15.0; SPSS, Chicago, IL). All four variables were force-entered into the regression model, with the dependent variable coded ordinally (control = 0, aMCI = 1, AD = 2).

Cross-validation

The resultant model was cross-validated using a fivefold cross-validation procedure (Efron & Tibshirani, 1994). The data set was randomly divided into five groups of roughly equal size (for three groups, $n = 51$; for two groups, $n = 50$). Four of the five groups were used to estimate model parameters and classification accuracy was evaluated on the remainder of the sample. This procedure was repeated five times, leaving each group out of the model exactly once. The resulting classification accuracy statistics are an average of the results from the five cross-validation steps.

RESULTS

A breakdown of the participant demographics among the three diagnostic groups is provided in Table 1. Table 1 also depicts the level of global impairment for each group, based on both Clinical Dementia Rating (CDR; Morris, 1993) Global Score and Mini-Mental State Exam (MMSE; Folstein et al., 1975) scores. Significant group differences were found on age (control < aMCI = AD), education (control > aMCI = AD), CDR Global Score, and average MMSE score (control > aMCI > AD).

Univariate Analyses

Independent samples *t* tests demonstrated significant group differences on each of the four NAB List Learning tests (Table 1). ROC curve analyses for the NAB List Learning variables are presented in Table 2. The cutoff scores presented in Table 2 were chosen to maximize sensitivity and specificity, with equal emphasis on both (Youden, 1950).

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press

Table 1. Participant demographics and test results

	Control	aMCI	AD
N	98	29	26[a]
Visit Number (*Mdn*)	4.0	4.0	4.0
Age			
M (*SD*)	71.5 (7.8)	76.1 (6.4)	80.6 (6.6)
Education			
M (*SD*)	16.5 (2.4)	14.7 (2.5)	14.7 (2.9)
Sex			
Male (*n*)	32	13	15
Female (*n*)	66	16	11
Race			
Caucasian (*n*)	81	24	23
Black/AA (*n*)	17	5	3
CDR Global Score			
0.0 (*n*)	97	12	0
0.5 (*n*)	1	17	5
1.0 (*n*)	0	0	13
2.0 (*n*)	0	0	8
MMSE			
M (*SD*)	29.6 (0.6)	28.0 (1.9)	23.1 (4.6)
NAB List Learning T-Scores			
List A Immediate Recall	52.3 (9.0)	40.4 (10.9)	30.2 (10.0)
List B Immediate Recall	51.4 (7.6)	44.3 (8.8)	39.7 (8.8)
List A Short Delay Recall	53.1 (8.5)	38.8 (10.7)	28.0 (7.4)
List A Long Delay Recall	53.2 (9.2)	38.9 (11.5)	31.0 (6.2)

Note. aMCI = amnestic mild cognitive impairment; AD = Alzheimer's disease; AA = African American; CDR = Clinical Dementia Rating; MMSE = Mini-Mental Status Examination; NAB = Neuropsychological Assessment Battery.
[a] Possible AD: *n* = 6; Probable AD: *n* = 20.

The individual NAB List Learning test variables were able to differentiate aMCI from controls (*Mdn*: sensitivity = .73; specificity = .71), AD from controls (*Mdn*: sensitivity = .89; specificity = .94), and AD from aMCI (*Mdn*: sensitivity = .69; specificity = .78). Additional prevalence-free classification accuracy statistics (i.e., those independent of base rates, such as sensitivity, specificity, PLR, and NLR) for conventional cutoff scores are provided in Table 3.

Multivariate Analyses

Likelihood ratio and goodness-of-fit tests revealed that the multiple ordinal logistic regression model explained a significant portion of outcome variance and fit the data well, −2 Log Likelihood χ^2 (4, n = 153) = 127.80; $p < .01$; Pearson Goodness of Fit χ^2 (298, n = 153) = 216.86; p = 1.00. Of the four independent variables, List B Immediate Recall (parameter estimate = −0.05; 95% CI = −0.09 to −0.01; Wald (1, n = 153) = 5.30; p = .02) and List A Short Delay Recall (parameter estimate = −0.10; 95% CI = −0.15 to −0.05; Wald (1, n = 153) = 14.5; $p < .001$) were found to be the two that contributed significantly to the model. List A Immediate Recall (parameter estimate = −0.02; 95% CI = −0.07 to 0.02; Wald (1, n = 153) = 1.33; p = .25) and List A Long Delay Recall (parameter estimate = 0.03; 95% CI = −0.08 to 0.02, Wald (1, n = 153) = 1.11; p = .29) were not significant contributors to the model.

Cross-validation

The estimated classification accuracy of the model using cross-validation was 80% (95% CI = 72–88%). In identifying aMCI, the model yielded a sensitivity of .47 (95% CI = .17–.77) and a specificity of .91 (95% CI = .83–.99; PLR = 4.96; NLR = .59). In identifying AD, the model yielded a sensitivity of .65 (95% CI = .41–.89) and a specificity of .97 (95% CI = .94–.99; PLR = 21.18; NLR = .36). A frequency table of predicted by actual diagnosis is presented in Table 4. Table 5 presents the positive predictive power (PPP) and negative predictive power (NPP) of the ordinal model across a range of clinically relevant base rates.

DISCUSSION

The results of this study show that the NAB List Learning test can differentiate between cognitively normal older adults and those with aMCI and AD. Univariate analyses showed that each of the four variables was able to make dichotomous classifications with sensitivity values ranging from .58 to .92 and specificity values ranging from .52 to .97. For instance, AD was differentiated from controls with over 90% sensitivity and specificity using a cutoff score of T ≤ 37 on List A Short Delay Recall or T ≤ 40 on List A Long Delay Recall. In addition, AD was differentiated from aMCI with over 70% sensitivity and 80% specificity using

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press

Table 2. Prevalence-free classification accuracy statistics for NAB List Learning variables at optimal cutoff scores

Variable	Optimal cutoff	Sensitivity (95% CI)	Specificity (95% CI)	PLR	NLR
All Impaired (aMCI and AD) *vs.* Control					
List A Immediate Recall	T ≤ 44	.78 (.65–.88)	.80 (.70–.87)	3.83	0.27
List B Immediate Recall	T ≤ 44	.66 (.51–.78)	.79 (.69–.86)	3.05	0.44
List A Short Delay Recall	T ≤ 38	.75 (.61–.85)	.97 (.91–.99)	24.35	0.26
List A Long Delay Recall	T ≤ 40	.76 (.63–.87)	.97 (.91–.99)	24.95	0.24
aMCI *vs.* Control					
List A Immediate Recall	T ≤ 46	.76 (.57–.90)	.69 (.59–.78)	2.48	0.35
List B Immediate Recall	T ≤ 47	.69 (.49–.85)	.62 (.52–.72)	1.83	0.50
List A Short Delay Recall	T ≤ 48	.86 (.68–.96)	.72 (.63–.81)	3.13	0.19
List A Long Delay Recall	T ≤ 40	.62 (.42–.79)	.97 (.91–.99)	20.28	0.39
AD *vs.* Control					
List A Immediate Recall	T ≤ 40	.85 (.65–.96)	.90 (.82–.95)	8.29	0.17
List B Immediate Recall	T ≤ 44	.85 (.65–.96)	.79 (.69–.86)	3.95	0.20
List A Short Delay Recall	T ≤ 37	.92 (.75–.99)	.97 (.91–.99)	30.15	0.08
List A Long Delay Recall	T ≤ 40	.92 (.75–.99)	.97 (.91–.99)	30.15	0.08
AD *vs.* aMCI					
List A Immediate Recall	T ≤ 30	.58 (.37–.77)	.86 (.68–.96)	4.18	0.49
List B Immediate Recall	T ≤ 41	.65 (.44–.83)	.72 (.53–.87)	2.37	0.48
List A Short Delay Recall	T ≤ 30	.73 (.52–.88)	.83 (.64–.94)	4.24	0.33
List A Long Delay Recall	T ≤ 36	.89 (.70–.98)	.52 (.33–.71)	1.83	0.22

Note. PLR = Positive Likelihood Ratio; NLR = Negative Likelihood Ratio; aMCI = Amnestic mild cognitive impairment; AD = Alzheimer's disease.

a cutoff score of T ≤ 30 on List A Short Delay Recall (see Table 2).

The multivariate ordinal logistic regression model, which incorporated four NAB List Learning variables, yielded an overall accuracy estimate of 80% based on fivefold cross-validation. In particular, the model was able to identify participants diagnosed with aMCI and AD with high specificity (.91 for aMCI and .97 for AD), but lower sensitivity (.47 for aMCI and .65 for AD). Taking prevalence into account, the ordinal logistic regression model was found to perform best when ruling out aMCI or AD (i.e., higher NPP) at lower base rates and when ruling in aMCI or AD (i.e., higher PPP) at higher base rates (Table 5). More specifically, in settings with clinical base rates of aMCI and AD at 20% or below, good performance on the NAB List Learning test can yield high confidence (i.e., NPP ≥ .87) that the patient would not be diagnosed as aMCI or AD by our consensus team. Similarly, in a setting with base rates of aMCI or AD around 50% or greater, as may be seen in a memory disorders clinic, poor performance on the NAB List Learning test can provide a high degree of confidence (i.e., PPP ≥ .72) that the patient would be given a diagnosis of aMCI or AD by our consensus team.

It should be noted that the current sample excluded individuals who did not complete the entire NAB List Learning

Table 3. Prevalence-free diagnostic accuracy statistics for NAB List Learning variables at conventional cutoff scores

	Cutoff	Impaired[a] *vs.* Control				aMCI *vs.* Control				AD *vs.* Control				AD *vs.* aMCI			
Variable	(T ≤)	Sn	Sp	PLR	NLR	Sn	Sp	PLR	NLR	Sn	Sp	PLR	NLR	Sn	Sp	PLR	NLR
List A Immediate Recall	40	.66	.90	6.41	0.38	.48	.90	4.73	0.58	.85	.90	8.29	0.17	.85	.52	1.75	0.30
	35	.51	1.00	—	0.49	.35	1.00	—	0.66	.69	1.00	—	0.31	.69	.66	2.01	0.47
	30	.35	1.00	—	0.65	.17	1.00	—	0.86	.58	1.00	—	0.42	.58	.86	4.18	0.49
List B Immediate Recall	40	.40	.94	6.53	0.64	.28	.94	4.51	0.77	.54	.94	8.79	0.49	.54	.72	1.95	0.64
	35	.22	1.00	—	0.78	.10	1.00	—	0.89	.35	1.00	—	0.65	.35	.90	3.35	0.73
	30	.11	1.00	—	0.89	.07	1.00	—	0.93	.15	1.00	—	0.85	.15	.93	2.23	0.91
List A Short Delay Recall	40	.75	.96	18.26	0.27	.59	.96	14.36	0.43	.92	.96	22.62	0.08	.92	.41	1.57	0.19
	35	.60	.97	19.60	0.41	.38	.97	12.39	0.64	.85	.97	27.64	0.16	.85	.62	2.23	0.25
	30	.44	1.00	—	0.56	.17	1.00	—	0.83	.73	1.00	—	0.27	.73	.83	4.24	0.33
List A Long Delay Recall	40	.76	.97	24.95	0.24	.62	.97	20.28	0.39	.92	.97	30.15	0.08	.92	.38	1.49	0.20
	35	.64	.98	31.18	0.37	.48	.98	23.66	0.53	.81	.98	39.58	0.20	.81	.52	1.67	0.37
	30	.35	.99	33.85	0.66	.24	.99	23.66	0.77	.46	.99	45.23	0.54	.46	.76	1.91	0.71

Note. aMCI = Mild Cognitive Impairment; AD = Alzheimer's disease; Sn = Sensitivity; Sp = Specificity; PLR = Positive Likelihood Ratio; NLR = Negative Likelihood Ratio. Dashes represent a value of positive infinity due to a specificity of 1.00.
[a]Includes both aMCI and AD.

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press

Table 4. Frequency of predicted diagnosis by actual consensus diagnosis

		Actual consensus diagnosis		
		Control	aMCI	AD
Predicted diagnosis (ordinal model)	Control	94	12	3
	aMCI	4	13	8
	AD	0	4	15

Note. aMCI = Amnestic mild cognitive impairment; AD = Alzheimer's disease.

test, which, for some participants, was due to excessive cognitive impairment. In addition, the participants with AD in the current study were predominantly in the very mild (CDR = 0.5, $n = 5$ [19%]) to mild (CDR = 1.0; $n = 13$ [50%]) stages. Because the current sample is generally free from severe impairment, it may be a valid representation of the types of patients that clinicians are asked to evaluate for early diagnosis.

Of the four variables entered into the ordinal logistic regression model, only two were found to contribute significantly: List B Immediate Recall and List A Short Delay Recall. Despite these findings, the results do not necessarily suggest that the nonsignificant variables lack value in differentiating healthy controls from individuals with aMCI from those with AD; in fact, both List A Immediate Recall and List A Long Delay Recall, in isolation, can differentiate control, aMCI, and AD groups with sensitivity values ranging from .58 to .92 and specificity values ranging from .52 to .97 (see Table 2). However, the results do suggest that these nonsignificant variables do not lead to a significant increase in explanatory power beyond what can be attained after considering List B Immediate Recall and List A Short Delay Recall performance.

Despite the fact that the MCI and AD groups were older and less educated than the control group, these demographic differences are unlikely to be contributing to the current results. Although age and education differ across groups, the use of demographically-corrected normative data protects against their potential confounding influence. In other words, the use of demographically-corrected norms prevents age and education from being associated with the independent variables. In fact, the NAB Psychometric and Technical Manual (White & Stern, 2003) illustrates that age accounts for 0.0% of the variance and education accounts for 0.0% to 0.2% of the variance in scores on the independent variables that were used in the current study.

Table 5. Positive and negative predictive power of the ordinal NAB List Learning model at various base rates

Predicted diagnosis (ordinal model)	Base rate					
	1%	5%	10%	20%	33%	50%
aMCI						
PPP	.05	.22	.37	.57	.72	.84
NPP	.99	.97	.94	.87	.78	.63
AD						
PPP	.18	.53	.71	.84	.91	.96
NPP	1.00	.98	.96	.92	.85	.73

Note. aMCI = Amnestic mild cognitive impairment; PPP = Positive Predictive Power; NPP = Negative Predictive Power; AD = Alzheimer's disease.

The classification accuracy of the NAB List Learning test compares favorably to published data on other list learning tests. For instance, the median sensitivity and specificity values of the individual NAB List Learning variables are generally on par with those seen in tests such as the CVLT, AVLT, HVLT, and the CERAD Word List (Table 6). More specifically, for example, a recent study found that long delay free recall on the CVLT differentiated AD from controls with a sensitivity of .98 and a specificity of .88 (Salmon et al., 2002), similar to the reported values of NAB List A Long Delay Recall in the current study (sensitivity = .92, specificity = .97). However, a major strength of the current study is that it validates a single model, developed using multiple ordinal logistic regression, that combines several list learning variables simultaneously to discriminate between three diagnostic groups (i.e., control, aMCI, and AD). One advantage of this ordinal logistic regression model is that it combines the NAB List Learning variables quantitatively, yielding results that can be integrated with applicable base rates to estimate diagnostic likelihood. The use of this model allows for an empirically-validated, quantitative method of combining important variables, as opposed to using clinical judgment for "profile" analysis, which may be susceptible to limitations in human cognitive processing, such as interpreting patterns among multiple neuropsychological test variables (Wedding & Faust, 1989).

Although the current findings support the diagnostic utility of the NAB List Learning test, the generalizability of the current results is limited. For instance, the sample is highly educated; data were collected in a research setting where many individuals volunteered due to self-awareness of memory difficulties; and the specifics of the reference standard, such as the clinicians participating in the consensus team and the assessment protocol used, are unique to our setting. Although the sample contains a fair number of African American participants (16%), representation of other minority groups is lacking. An additional limitation is the fact that the NAB List Learning test was not directly compared with other list learning tests in the same sample, precluding more definitive statements about its diagnostic accuracy in relationship to alternate tests. Finally, the results are limited by the reference standard that was used to establish a diagnosis. Despite the documented advantages of actuarial approaches over subjective approaches to clinical decision making (Dawes et al., 1989; Grove et al., 2000), it is important to emphasize that the reference standard used in the current study is a multidisciplinary consensus diagnosis based on contemporary clinical diagnostic criteria, not neuropathological diagnosis. At the present time, diagnosis of definite AD requires neuropathological confirmation

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press

Table 6. Comparison of sensitivity and specificity between individual variables from the NAB and the AVLT, CERAD, CVLT, and HVLT

Test	Variable	Sensitivity	Specificity
MCI *vs.* Control			
CERAD	IR[a]	.33–.73	.80–.93
	IR[b]	.67	1.00
	DR[c]	.83	.60
	DR[a]	.26	1.00
	DR[b]	.81	.86
	%Ret[c]	.89	.55
	%Ret[a]	.33	.66
	%Ret[b]	.62	.90
	Recognition[a]	.33–.70	.93–1.00
	Recognition[c]	.94	.35
HVLT	IR[d]	.82	.79
NAB	List A IR[e]	.76	.69
	List B IR[e]	.69	.62
	List A SDR[e]	.86	.72
	List A LDR[e]	.62	.97
AD *vs.* Control			
CERAD	IR[f]	.86	.87
	IR[a]	.60–.86	.80–.93
	IR[g]	.95	.89
	IR[b]	.89	1.00
	DR[f]	.74	.82
	DR[a]	.86	1.00
	DR[g]	.92	.89
	DR[b]	1.00	.86
	%Ret[a]	.80	.66
	%Ret[b]	.79	.90
	Recognition[f]	.76	.87
	Recognition[a]	.60–.80	.93–1.00
CVLT	IR[h]	.95	.89
	LDR[h]	.98	.88
HVLT	IR[i]	.75	.92
AVLT	Trial 1[j]	.63	.90
	Trial 5[j]	.80	.43
	SDR[j]	.79	.81
	LDR[j]	.83	.83
NAB	List A IR[e]	.85	.90
	List B IR[e]	.85	.79
	List A SDR[e]	.92	.97
	List A LDR[e]	.92	.97
AD *vs.* MCI			
HVLT	IR[d]	.79	.96
	IR[k]	.91	.69
NAB	List A IR[e]	.58	.86
	List B IR[e]	.65	.72
	List A SDR[e]	.73	.83
	List A LDR[e]	.89	.52

Note. MCI = Mild cognitive impairment; CERAD = Consortium to Establish a Registry for Alzheimer's Disease; IR = Immediate Recall; DR = Delayed Recall; %Ret = Percent Retention; HVLT = Hopkins Verbal Learning Test; NAB = Neuropsychological Assessment Battery; SDR = Short Delay Recall; LDR = Long Delay Recall; AD = Alzheimer's disease; CVLT = California Verbal Learning Test; AVLT = Auditory Verbal Learning Test.
[a] Karrasch et al. (2005).
[b] Ivanoiu et al. (2005).
[c] Woodard et al. (2005).
[d] Schrijnemaekers et al. (2006).
[e] Current study (see Table 2).
[f] Bertolucci et al. (2001).
[g] Derrer et al. (2001).
[h] Salmon et al. (2002).
[i] Kuslansky et al. (2004).
[j] Schoenberg et al. (2006).
[k] de Jager et al. (2003).

(McKhann et al., 1984). Consequently, the classification accuracy statistics reported herein cannot be interpreted to reflect the likelihood that a patient actually has AD; instead, they indicate the likelihood that this specific consensus diagnostic team would make a particular diagnosis when using the assessment methods described above. It should also be noted that the consensus diagnosis was made, in part, on the basis of other neuropsychological tests, some of which are methodologically and psychometrically similar to the NAB List Learning test. This may have introduced an inherent and unavoidable source of bias. However, the diagnoses was based on consensus after consideration of a wide range of information, thus reducing the likelihood that shared method variance between the NAB List Learning test and other episodic memory measures would have caused significant tautological concerns.

From a methodological standpoint, there are other limitations that require future study. The data were analyzed retrospectively and at various points in the longitudinal assessment of participants. An important line of future research would be to longitudinally follow individuals diagnosed with aMCI to prospectively examine whether NAB List Learning test performance is associated with AD progression. Because the current study does not include other dementia subtypes, future studies should also examine non-AD dementias. Finally, to limit the number of predictor variables in the ordinal logistic regression model, the NAB List Learning variables that are considered "secondary" or "descriptive" (White & Stern, 2003) were excluded. However, these additional variables may add additional diagnostic utility to the List Learning test, and future study is warranted.

Despite its limitations, the current study has several strengths. For instance, diagnostic accuracy statistics are provided for a large number of cutoff scores, providing users of the test considerable flexibility in interpreting test results. For example, depending on the desired purpose of the examination, users may wish to choose cutoff scores that place a higher value on sensitivity (e.g., clinical settings, where false positive errors may preferable to false negative errors) or specificity (e.g., research settings, where false negative errors may be preferable to false positive errors). Users of the test may choose to interpret results using traditional cutoff scores (e.g., Z-scores ≤ 1.5 or 2.0), or to use the empirically-derived cutoff scores presented herein to emphasize sensitivity and specificity equally. In addition, test users may choose to examine each test variable individually, or to interpret the overall pattern of test scores using the multiple ordinal logistic regression model, which accounts for performance on the four primary NAB List Learning variables simultaneously. For the latter approach, positive and negative predictive values are provided for a range of base rates, allowing for a more individually tailored approach to test interpretation. An additional strength of the study was the lack of tautological error, as the NAB List Learning test was not used in diagnostic formulations. Instead, NAB List Learning performance was examined independently against the clinical "gold standard," a multidisciplinary consensus diagnostic conference.

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press

The cross-validation of the ordinal logistic regression model allows for examination of the degree of precision in estimates of sensitivity, specificity, and overall accuracy. Based on the reported confidence intervals, there is a good degree of precision in the ordinal model's overall accuracy (accuracy = 80%; 95% CI = 72–88%) and in the model's specificity to the diagnosis of both aMCI (specificity = .91; 95% CI = .83–.99) and AD (specificity = .97; 95% CI = .94–.99). However, in examining the 95% confidence intervals surrounding the sensitivity estimates for both aMCI and AD, it is apparent that the sensitivity of the ordinal model is considerably lower and lacking precision. This may be due in part to the relatively small sizes of the clinical sample and in part due to the negative log-log link function that was used in the multiple ordinal logistic regression model. This link function makes an *a priori* assumption that the underlying distribution of the data is skewed toward "normality." In other words, the model was chosen based on the assumption that the prevalence of healthy controls is greater than the prevalence of individuals with aMCI and AD. As a result, the ordinal logistic regression model may be more prone to false negative errors (i.e., reduced sensitivity) than to false positive errors (i.e., reduced specificity). This decreased sensitivity to aMCI and AD may also reflect the fact that individuals with aMCI and AD perform similarly on measures of episodic memory, and that functional measures may be necessary to improve diagnostic sensitivity once a certain degree of cognitive decline has occurred in an individual. Although the current results present diagnostic accuracy statistics for the NAB List Learning test, it should be emphasized that a diagnosis of aMCI or AD cannot be made on the basis of a single neuropsychological instrument.

The current results demonstrate that the NAB List Learning test was able to classify older adults into cognitively normal, AD, and aMCI groups with accuracy levels similar to other published list learning tests (Bertolucci et al., 2001; de Jager et al., 2003; Derrer et al., 2001; Ivanoiu et al., 2005; Karrasch et al., 2005; Kuslansky et al., 2004; Salmon et al., 2002; Schoenberg et al., 2006; Schrijnemaekers et al., 2006; Woodard et al., 2005). The NAB List Learning test possesses a large and up-to-date set of demographically-corrected normative data ($n = 1441$) and it was co-normed as part of a comprehensive neuropsychological test battery. In addition, it was developed to include two equivalent forms; in fact, in the NAB standardization sample ($n = 1448$), test form accounted for less than 1.5% of the total variance seen in List Learning performance (White & Stern, 2003), making it suitable for clinical re-evaluation and longitudinal research applications. The findings from the current study, along with the overall strengths of the NAB, suggest that the NAB List Learning test is an appropriate and clinically useful tool for the evaluation of older adults with known or suspected Alzheimer's disease. Although the current study did not directly compare the diagnostic utility of the NAB List Learning test to other list learning measures, the classification accuracy data presented herein are similar to those reported in the literature investigating the diagnostic utility of other list learning tests in control, MCI, and AD samples (see Table 6). Future research is warranted to make direct comparisons of diagnostic utility to other list learning instruments.

ACKNOWLEDGMENTS

The project described was supported by Grant Number M01 RR000533 and the CTSA Grant Number 1UL1RR025771 from the National Center for Research Resources (NCRR), a component of the National Institute of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. This research was also supported by P30-AG13846 (Boston University Alzheimer's Disease Core Center), R03-AG026610 (ALJ), R03-AG027480 (ALJ), K12-HD043444 (ALJ), K23-AG030962 (ALJ), RO1-HG02213 (RCG), RO1-AG09029 (RCG), R01-MH080295 (RAS), and K24-AG027841 (RCG). Robert A. Stern is one of the developers of the NAB and receives royalties from its publisher, Psychological Assessment Resources Inc. Portions of this manuscript were presented at the International Conference on Alzheimer's Disease, Chicago, July 2008. With this exception, the information in this manuscript and the manuscript itself has never been published either electronically or in print. A scoring program based on the multiple ordinal regression model reported in this manuscript is available as an electronic addendum to this manuscript.

REFERENCES

Ashendorf, L., Jefferson, A.L., O'Connor, M.K., Chaisson, C., Green, R.C., & Stern, R.A. (2008). Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology*, *23*, 129–17.

Beekly, D.L., Ramos, E.M., Lee, W.W., Deitrich, W.D., Jacka, M.E., Wu, J., Hubbard, J.L., Koepsell, T.D., Morris, J.C., Kukull, W.A., and the NIA ADCs (2007). The National Alzheimer's Coordinating Center (NACC) Database: The Uniform Data Set. *Alzheimer Disease and Associated Disorders*, *21*, 249–258.

Bertolucci, P.H.F., Okamoto, I.H., Brucki, S.M.D., Siviero, M.O., Neto, J.T., & Ramos, L.R. (2001). Applicability of the CERAD neuropsychological battery to Brazilian elderly. *Arquivos de Neuro-Psiquiatria*, *59*, 532–536.

Blacker, D., Lee, H., Muzikansky, A., Martin, E.C., Tanzi, R., McArdle, J.J., Moss, M., & Albert, M. (2007). Neuropsychological measures in normal individuals that predict subsequent cognitive decline. *Archives of Neurology*, *64*, 862–871.

Brandt, J. & Benedict, R.H.B. (2001). *Hopkins Verbal Learning Test-Revised. Professional manual*. Lutz, FL: Psychological Assessment Resources.

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

Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668–1673.

de Jager, C.A., Hogervorst, E., Combrinck, M., & Budge, M.M. (2003). Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine*, *33*, 1039–1050.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (1987). *The California Verbal Learning Test*. New York: Psychological Corporation.

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press

Derrer, D.S., Howieson, D.B., Mueller, E.A., Camicioli, R.M., Sexton, G., & Kaye, J.A. (2001). Memory testing in dementia: How much is enough? *Journal of Geriatric Psychiatry and Neurology*, *14*, 1–6.

Efron, B. & Tibshirani, R.J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.

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.

Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19–30.

Ivanoiu, A., Adam, S., Van der Linden, M., Salmon, E., Juillerat, A., Mulligan, R., & Seron, X. (2005). Memory evaluation with a new cued recall test in patients with mild cognitive impairment and Alzheimer's disease. *Journal of Neurology*, *252*, 47–55.

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.

Karrasch, M., Sinervä, E., Grönholm, P., Rinne, J., & Laine, M. (2005). CERAD test performances in amnestic mild cognitive impairment and Alzheimer's disease. *Acta Neurologica Scandinavica*, *111*, 172–179.

Kuslansky, G., Katz, M., Verghese, J. Hall, C.B., Lapuerta, P., LaRuffa, G., & Lipton, R.B. (2004). Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination. *Archives of Clinical Neuropsychology*, *19*, 89–104.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDS 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. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, *43*, 2412–2414.

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., Mellits, E.D., & Clark, C. (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.

Morris, J.C., Weintraub, S., Chui, H.C., Cummings, J., Decarli, C., Ferris, S., Foster, N.L., Galasko, D., Graff-Radford, N., Peskind, E.R., Beekly, D., Ramos, E.M., & Kukull, W.A. (2006). The Uniform Data Set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer disease centers. *Alzheimer Disease and Associated Disorders*, *20*, 210–216.

Rey, A. (1941). L'examen psychologie dans les cas d'encephalopathie traumatique. *Archives de Psychologie*, *28*, 286–340.

Rey, A. (1964). *L'examen clinique en psychologie.* Paris: Presses Universitaires de France.

Salmon, D.P., Thomas, R.G., Pay, M.M., Booth, A., Hofstetter, C.R., Thal, L.J., & Katzman, R. (2002). Alzheimer's disease can be diagnosed in very mildly impaired individuals. *Neurology*, *59*, 1022–1028.

Schoenberg, M.R., Dawson, K.A., Duff, K., Patton, D., Scott, J.G., & Adams, R.L. (2006). Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Archives of Clinical Neuropsychology*, *21*, 693–703.

Schrijnemaekers, A.M.C., de Jager, C.A., Hogervorst, E., & Budge, M.M. (2006). Cases with mild cognitive impairment and Alzheimer's disease fail to benefit from repeated exposure to episodic memory tests as compared with controls, *Journal of Clinical and Experimental Neuropsychology*, *28*, 438–455.

Stern, R.A. & White, T. (2003a). *NAB Administration, Scoring, and Interpretation Manual*. Lutz, FL: Psychological Assessment Resources.

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

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

Wedding, D. & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology*, *4*, 233–265.

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

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.O., Nordberg, A., Backman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., van Duijn, C., Visser, P., & Petersen, R.C. (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.

Woodard, J.L., Dorsett, E.S.W., Cooper, J.G., Hermann, B.P., & Sager, M.A. (2005). Development of a brief cognitive screen for mild cognitive impairment and neurocognitive disorder. *Aging, Neuropsychology, and Cognition*, *12*, 299–315.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32–35.

https://doi.org/10.1017/S1355617708090176 Published online by Cambridge University Press