Diagnostic Accuracy Statistics for Seven Neuropsychological Assessment Battery (NAB) Test Variables in the Diagnosis of Alzheimer's Disease

Brandon E. Gavett[a], **Katherine R. Lou**[b], **Daniel H. Daneshvar**[b], **Robert C. Green**[b], **Angela L. Jefferson**[b], and **Robert A. Stern**[b]

[a]Department of Psychology, University of Colorado at Colorado Springs, Colorado Springs, Colorado

[b]Alzheimer's Disease Center, Boston University School of Medicine, Boston, Massachusetts

Abstract

Neuropsychological tests are useful for diagnosing Alzheimer's disease (AD), yet for many tests, diagnostic accuracy statistics are unavailable. We present diagnostic accuracy statistics for seven variables from the Neuropsychological Assessment Battery (NAB) that were administered to a large sample of elderly adults (*n* = 276) participating in a longitudinal research study at a national AD Center. Tests included Driving Scenes, Bill Payment, Daily Living Memory, Screening Visual Discrimination, Screening Design Construction, and Judgment. Clinical diagnosis was made independent of these tests, and for the current study, participants were categorized as AD (*n* = 65) or non-AD (*n* = 211). Receiver operating characteristics curve analysis was used to determine each test's sensitivity and specificity at multiple cut points, which were subsequently used to calculate positive and negative predictive values at a variety of base rates. Of the tests analyzed, the Daily Living Memory test provided the greatest accuracy in the identification of AD and the two Screening measures required a significant tradeoff between sensitivity and specificity. Overall, the seven NAB subtests included in the current study are capable of excellent diagnostic accuracy, but appropriate understanding of the context in which the tests are used is crucial for minimizing errors.

Introduction

Neuropsychological assessment is an important component in the assessment and diagnosis of Alzheimer's disease (AD). Neuropsychological assessment instruments are typically used for both descriptive and diagnostic purposes (Busch, Chelune, & Suchy, 2006). When used descriptively, tests contextualize an individual's performance relative to his or her peers. When used diagnostically, tests provide information about the probability that a particular individual has— or will have at some future point—a cognitive disorder, such as AD. The current study focuses on the diagnostic value of several neuropsychological tests from the Neuropsychological Assessment Battery (NAB; Stern & White, 2003) in the assessment of older adults.

Individuals with AD have deficits in memory and at least one other cognitive domain, which interferes with independent functioning (McKhann et al., 1984). Although definite AD cannot be diagnosed until autopsy, the diagnosis can be made clinically, with varying

degrees of certainty, as either possible AD or probable AD. Neuropsychological test data may be useful in determining whether an individual's cognitive functioning is consistent with what is typically seen in clinically diagnosed AD. Cognitive tests may be used for various reasons, such as for screening purposes or as part of a more comprehensive evaluation designed to formally diagnose AD. For instance, a clinician evaluating a patient with subtle memory complaints may prefer tests that reduce the possibility of false-negative errors (i.e., failing to detect cognitive impairment in a person with AD). On the other hand, in a randomized controlled trial for a novel AD treatment, cognitive tests that reduce false-positive errors may be preferred (i.e., to ensure that cognitively healthy individuals are not enrolled in the trial). Sensitivity and specificity data can be used to calculate statistics such as likelihood ratios and predictive values (Ivnik et al., 2001), which offer more diagnostically useful information (O'Bryant & Lucas, 2006) because they account for the prevalence of the condition in the population of interest (Nugent, 2005). Therefore, sensitivity and specificity values must be converted into positive and negative predictive values, using appropriate base rates, to better understand the diagnostic utility of a test.

Many studies have highlighted the excellent sensitivity and specificity of episodic memory measures—especially list learning tests—to AD (Derrer et al., 2001; Kuslansky et al., 2004; Salmon et al., 2002; Schoenberg et al., 2006). The diagnostic utility of the List Learning test included in the NAB has been previously demonstrated in an aging cohort (Gavett et al., 2009, 2010), but other NAB subtests have yet to be examined. One of the major strengths of the NAB is that all of its tests are co-normed on a large, representative, and diverse normative sample, which makes it an excellent tool for contextualizing test performance. However, it is important to quantify the diagnostic utility of these tests as well.

The goal of the current study was to examine the diagnostic utility of a number of tests from the NAB to detect AD in a sample of older adults. For more than 4 years, a select group of NAB tests has been administered to all participants in a longitudinal research registry on cognitive aging at the Boston University (BU) Alzheimer's Disease Center (ADC). These tests from the NAB have not been used to assign clinical diagnoses to participants, a procedure that was implemented at the outset to allow the NAB subtests to be studied independent of the information used to diagnose participants. We examined the diagnostic utility of seven NAB test variables, spanning the domains of attention (Driving Scenes test), language (Bill Payment test), memory (Daily Living Memory–Immediate and Delayed tests), visuospatial ability (Screening Design Construction and Screening Visual Discrimination tests), and executive function (Judgment test), in detecting individuals with possible or probable AD.

Methods

Participants

Archival data were extracted from a longitudinal research registry at the BU ADC. Additional registry details have been described more fully elsewhere (Jefferson et al., 2006). The BU ADC Research Registry data collection procedures were approved by the BU Medical Center Institutional Review Board. All participants provided written informed consent to participate in the study. For the current study, we identified 276 participants, who at their most recent annual registry visit, were administered the selected NAB tests as part of a larger comprehensive evaluation designed for the assessment and diagnosis of AD, including those tests that make up the ADC's Uniform Data Set (Weintraub et al., 2009). The participants included 164 women (59.4%) and 112 men (40.6%); 212 were Caucasian (76.8%), 62 were Black or African American (22.5%), 1 was Asian (0.4%), and 1 identified as another unspecified race (0.4%). The entire sample ranged in age from 51 to 94 years

(M=75.1 years,SD=8.2) and ranged in education from 6 to 21 years (M=15.2 years, SD=2.9 years).

Diagnosis

After each participant's most recent comprehensive annual assessment, which included clinical interview and history taking with the participant and a study partner, neuropsychological assessment, a neurological exam, and a review of medical, social, occupational, and family history, participants were assigned a consensus diagnosis by a multidisciplinary team of experts, including neuropsychologists, neurologists, and a nurse practitioner.

AD was diagnosed using the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Assoication (NINCDS-ADRDA) criteria (McKhann et al., 1984). A diagnosis of probable AD is met when there is evidence of dementia, impairment in two or more areas of cognitive functioning, and progressive memory and other cognitive decline that begins between ages 40 and 90 years old and is not solely attributable to disruptions in consciousness or another neurologic or systemic disease. Criteria for possible AD are met when the onset, course, or presentation of the dementing disorder are atypical for probable AD; when a second potential cause of dementia (e.g., vascular dementia) cannot be ruled out; or when there is progressive decline in only one area of cognition. Mild cognitive impairment (MCI) was diagnosed using the criteria set forth by Winblad et al. (2004), which requires either objective evidence of cognitive decline or a complaint made by the participant, an informant, or a clinician, along with evidence of impairment in one or more cognitive domains, but in the absence of global cognitive impairment and dementia (i.e., little to no loss of independent functioning). When participants met all of the Winblad et al. criteria for MCI except for the presence of a complaint, they were considered to have "possible" MCI; this is a diagnostic label unique to our center that is meant to identify individuals with objective cognitive impairment but no complaint of cognitive difficulties.

Following the consensus diagnosis conference, participant diagnostic status was recoded to indicate the presence or absence of AD (both possible and probable). Those diagnosed with AD were assigned to the clinical AD group (n=65), while those with any other diagnosis were assigned to the not clinical AD group (n=211). For simplicity, we refer to the clinical AD group as the "AD group" and the not clinical AD group as the "non-AD" group. The resulting diagnostic breakdown of both the AD and non-AD groups, as well as the specific diagnostic criteria utilized, is presented in Table 1. Because statistics such as sensitivity and specificity are used to differentiate between two groups, we chose to focus the current analyses on the identification of AD versus a heterogeneous sample of individuals without AD. If we had focused the analyses on discriminating between AD and a homogeneous comparison group (e.g., cognitively healthy adults), this would likely inflate the sensitivity and specificity values of the test but would provide less clinically relevant information, as clinicians are rarely faced with cognitively healthy and AD as the only two diagnostic possibilities.

Data analysis

We conducted independent-samples t-tests to determine whether the two groups differed in performance on the NAB tests, as well as in age and years of education. Receiver operating characteristics curve analyses were then conducted separately for each NAB test variable raw score using the Statistical Package for the Social Sciences Version 16.0. Again, using raw scores, sensitivity and specificity values were calculated for all of the possible cutoff scores for each test. We then identified the cutoff scores that met the following criteria: (a)

maximized Youden's index (Youden, 1950), which identifies the optimal combination of sensitivity and specificity; (b) maximized sensitivity at a value of <1.0; and (c) maximized specificity at a value of <1.0. When multiple cutoff scores resulted in maximum sensitivity or specificity values, we chose the score that yielded the largest Youden's index. We present different cutoff scores for each test to provide a more comprehensive understanding of the diagnostic utility of these tests, depending on whether greater sensitivity, greater specificity, or a balance between sensitivity and specificity is sought. Based on these data, we then calculated positive and negative predictive values across a variety of base rates. In many cases, tests were omitted due to time constraints or participant unwillingness, or because cognitive impairments prevented the participant from understanding the test instructions. Therefore, sample sizes differed for each test; these are listed in Table 2.

Because receiver operating characteristics curves are used to differentiate individuals with a disease from those without a disease (Shapiro, 1999), we chose to focus the current analyses on detecting the presence of AD against an unselected sample of individuals without AD. We believe that this approach more closely matches the situations faced by clinicians when using a test for diagnostic purposes—that is, to rule in or rule out a potential diagnosis when considered in the context of all other potential differential diagnoses. If we had focused the analyses on discriminating between AD and a homogeneous comparison group (e.g., cognitively healthy adults), this would likely inflate the sensitivity and specificity values of the test, but would provide less clinically relevant information, as clinicians are rarely faced with cognitively healthy and AD as the only two diagnostic possibilities. However, because some clinicians may prefer to interpret the diagnostic utility of these tests for differentiating between two specific diagnostic groups, we also calculated sensitivity and specificity values at cutoff scores that maximize sensitivity and specificity for differentiating MCI from cognitively healthy individuals and those with AD using receiver-operating characteristics curve analyses.

Results

Demographic information and neuropsychological test performance, broken down by the AD and non-AD groups, is presented in Table 2. As expected, the AD group produced significantly lower scores on all NAB measures; the AD group was older and there was no group difference in education.

In Tables 3 through 5, we present the primary results of the current study. Each table presents the NAB test, a given cutoff score, and the corresponding sensitivity and specificity values, followed by the positive and negative predictive powers at four clinically relevant base rates. Table 3 provides this information at cutoff scores that jointly maximize sensitivity and specificity, Table 4 at cutoff scores that maximize sensitivity, and Table 5 at cutoff scores that maximize specificity. Depending on the purpose of the evaluation and the base rate of AD in a given population, Tables 3 through 5 allow for these select NAB measures to be evaluated within the appropriate context. Table 6 provides sensitivity and specificity data for differentiating MCI from cognitively healthy individuals and those with AD.

Discussion

We present an analysis of the diagnostic utility of tests from the NAB that have been used in a large-scale study of cognitive aging and AD at a national ADC. The NAB tests under investigation were not used during the formulation of a clinical consensus diagnosis. As such, the current results are based on diagnoses that were made independent from the data used in the current study, which avoids any tautological error. Because, as addressed above,

the diagnostic utility of a test is dependent upon the goals of the assessment (e.g., to rule in or rule out a diagnosis) and the base rate of the condition of interest in the relevant population, we present diagnostic utility statistics that can be applied under a variety of circumstances. When using cutoff scores that jointly maximize sensitivity and specificity, the majority of the tests used in the current study are more specific than sensitive; in settings where the base rate of AD is 33% or less, the negative predictive power of these tests is superior to the positive predictive power. However, at a base rate of 50%, five of these seven tests provide superior positive predictive power (Table 3). In most clinical situations, outside of memory diagnostic clinics for the elderly, the prevalence of AD is likely to be less than 50%. In these circumstances, the seven NAB tests are more appropriate for ruling out AD given a score above the cutoff as opposed to ruling in AD given a score below the cutoff. When the base rate of AD approximates 50%, the Driving Scenes, Bill Payment, and Judgment tests provide positive predictive power greater than .80, and the two Daily Living Memory tests provide excellent positive and negative predictive power.

If the circumstances of the evaluation dictate choosing a cutoff score that maximizes the sensitivity of the test to AD, the results indicate that the cutoff scores should be set higher. This will capture more individuals with AD and thus minimize the number of false-negative errors, but at the expense of yielding a larger number of false-positive errors. The findings presented in Table 4 indicate that when an appropriate cutoff score is chosen, all seven NAB tests are extremely sensitive to AD, with no sensitivity value falling below .95. However, for most tests, this excellent sensitivity comes at the expense of very poor specificity, with only the Delayed Daily Living Memory trial achieving specificity greater than .80. The cutoff scores that maximize sensitivity led to better negative than positive predictive values across all of the base rates that were analyzed.

In other clinical and research conditions, specificity may be preferred; the results presented in Table 5 reveal that the seven NAB tests are also capable of an excellent degree of specificity when the appropriate cutoff scores are applied. These cutoff scores tend to be quite low and therefore result in very few false-positive errors, but at the expense of additional false-negative errors. When specificity is very high, the sensitivity of all seven instruments suffers, with no test achieving a sensitivity greater than .62. At low base rates (5% and 10%), negative predictive values are excellent, while at higher base rates (33% and 50%), nearly all tests provide superior positive predictive values.

Across most circumstances, both the Immediate and Delayed Recall trials from the NAB Daily Living Memory test appear to be useful for the diagnosis of AD. This is to be expected, given the fact that episodic memory is usually the most impaired area of cognition in AD (Backman, Jones, Berger, Laukka, & Small, 2005) and that episodic memory measures often possess excellent diagnostic utility (Rabin et al., 2009). Depending on the cutoff score that is chosen, the Daily Living Memory variables can yield positive and negative predictive values that exceed .90 at all of the base rates presented.

In addition to comprehensive modules that assess attention, language, memory, visuospatial skills, and executive functioning, the NAB also contains a Screening Module that includes abbreviated versions of some of the complete tests in each main module. In our current study, we administered two Screening visuospatial tasks to all participants: Visual Discrimination and Design Construction. The results indicate that these two screening measures are capable of being used with either high sensitivity or high specificity, depending on the cutoff score, but they are not both highly sensitive and specific. As such, their utility is more limited when used to diagnose AD. Other studies have also reported that visuospatial measures are less accurate in detecting AD, relative to tests measuring other cognitive abilities (De Jager, Hogervorst, Combrinck, & Budge, 2003). The two Screening

visuospatial tests reported herein may be most useful for ruling out AD in clinical settings with lower base rates (5% and 10%), when the negative predictive values exceed .90. However, under most circumstances, these screening instruments yield low positive predictive values, indicating that they are not particularly useful for ruling in AD, even when scores are very low.

In addition to the memory and screening tests, we also analyzed the diagnostic utility of three additional measures from the NAB: Driving Scenes, Bill Payment, and Judgment, all of which were designed as "Daily Living" tests. Like the Daily Living Memory test, these three measures are intended to replicate a real-world activity that most people perform on a regular basis for some functional purpose. Because AD is associated with a loss of independent functioning, these measures may be diagnostically useful. The results suggest that, under many circumstances, these Daily Living tests can accurately rule in or rule out AD. Each of these three tests appears to be the most useful when ruling out AD in situations where the base rate is relatively low (i.e., below 10%). However, at higher base rates, negative predictive values greater than .80 can be achieved by using higher cutoff scores, and positive predictive values greater than .80 can be achieved using lower cutoff scores. The sensitivity and specificity values of the NAB Daily Living tests are similar to those reported for other performance-based functional measures (Goldberg et al., 2010).

The current results clearly indicate that the ability of a test to contribute to diagnostic accuracy is a function of the cutoff score chosen and the relevant base rates. Neuropsychologists seeking diagnostic accuracy must be aware of these factors and their influence over positive and negative predictive values, in both research and clinical settings.

These results may be limited by the fact that the primary analyses in the current study focused on differentiating the AD group from a heterogeneous non-AD group that includes individuals with MCI, other non-AD dementias, and other non-dementia diagnoses that did not completely fulfill contemporary criteria for MCI (Winblad et al., 2004). Thus, the results presented in Tables 2 through 5 cannot be used to discriminate individuals with AD from any one particular group; rather, they only indicate the posterior probability of a person having or not having clinical AD given the test results. While we believe that the information provided by positive and negative predictive values is most accurate when used to rule one diagnosis in or out relative to an unselected group of competing diagnoses, we realize that some clinicians may prefer data for more specific group comparisons. Therefore, the data presented in Table 6 can be used to derive positive and negative predictive values at desired base rates by using Bayes's theorem (Nugent, 2005).

Our study is also limited in that the majority of the AD group was diagnosed clinically with possible AD, which is the least degree of confidence that can be assigned to a clinical AD diagnosis. Many of these participants diagnosed with possible AD may have had an alternate cause of dementia, such as cerebrovascular disease, or their degree of functional impairment may have been mild relative to expectations. The current study may generalize to White and Black individuals, but it is not representative of individuals of other races. In addition, our sample, which was recruited from the greater Boston area, may not be representative of the general population in terms of factors such as education and acculturation. Not all participants completed each of the seven NAB tests, resulting in missing data for many of the tests. These data were not missing at random, due to the fact that tests may have been excluded due to confusion or poor frustration tolerance, especially among those participants diagnosed with AD. If full data had been available, the current results may have been more accurate, because those who produced missing data may have been more likely to produce low test scores. Finally, the gold standard for diagnostic accuracy in the present study was a clinical consensus diagnosis of AD; because neuropathological evidence of AD was not

available, the diagnostic accuracy data presented are limited by the accuracy of the clinical consensus diagnosis.

Our results provide diagnostic accuracy statistics for seven test variables from the NAB. In addition, we have previously reported on the diagnostic utility of another NAB test, the List Learning test (Gavett et al., 2009). However, the NAB is composed of 33 distinct tests, some of which provide multiple outcome variables. As only a minority of the NAB test variables has been studied, future research should explore the diagnostic utility of the entire NAB, including its index scores for each major cognitive domain, and for the entire screening module. As mentioned previously, the NAB has a large, modern, and demographically broad set of normative data that make it useful for contextualizing performance relative to an appropriate reference standard. Although there is now research to suggest that select NAB subtests are diagnostically useful, more research is needed on this front. Presently, it is clear that depending on the context of the evaluation, the NAB variables presented in the current study can provide diagnostically useful information.

Acknowledgments

References

Backman L, Jones S, Berger A-K, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. Neuropsychology. 2005; 19:520–531. [PubMed: 16060827]

Busch, RM.; Chelune, GJ.; Suchy, Y. Using norms in neuropsychological assessment of the elderly. In: Attix, DK.; Bohmer, K. A. Welsh, editors. Geriatric neuropsychology: Assessment and intervention. Guilford Press; New York, NY: 2006. p. 133-157.

De Jager CA, Hogervorst E, Combrinck M, Budge MM. Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. Psychological Medicine. 2003; 33:1039–1050. [PubMed: 12946088]

Derrer DS, Howieson DB, Mueller EA, Camicioli RM, Sexton G, Kaye JA. Memory testing in dementia: How much is enough? Journal of Geriatric Psychiatry and Neurology. 2001; 14:1–6. [PubMed: 11281309]

Gavett BE, Ozonoff A, Doktor V, Palmisano J, Nair AK, Green RC, Stern RA. Predicting cognitive decline and conversion to Alzheimer's disease in older adults using the NAB List-Learning test. Journal of the International Neuropsychological Society. 2010; 16:651–660. [PubMed: 20374677]

Gavett B, Poon SJ, Ozonoff A, Jefferson AL, Nair AK, Green RC, Stern RA. Diagnostic utility of the NAB List-Learning test in Alzheimer's disease and amnestic mild cognitive impairment. Journal of the International Neuropsychological Society. 2009; 15:121–129. [PubMed: 19128535]

Goldberg TE, Koppel J, Keehlisen L, Christen E, DresesWerringloer U, Conejero-Goldberg C, Davies P. Performance-based measures of everyday function in mild cognitive impairment. The American Journal of Psychiatry. 2010; 167:845–853. [PubMed: 20360320]

Ivnik RJ, Smith GE, Cerhan JH, Boeve BF, Tangalos EG, Petersen RC. Understanding the diagnostic capabilities of cognitive tests. The Clinical Neuropsychologist. 2001; 15:114–124. [PubMed: 11778572]

Jefferson AL, Wong S, Bolen E, Ozonoff A, Green RC, Stern RA. Cognitive correlates of HVOT performance differ between individuals with mild cognitive impairment and normal controls. Archives of Clinical Neuropsychology. 2006; 21:405–412. [PubMed: 16893623]

Kuslansky G, Katz M, Verghese J, Hall CB, Lapuerta P, LaRuffa G, Lipton RB. Detecting dementia with the Hopkins Verbal Learning Test and the Mini-Mental State Examination. Archives of Clinical Neuropsychology. 2004; 19:89–104. [PubMed: 14670382]

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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. 1984; 34:939–944. [PubMed: 6610841]

Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, Benson DF. Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. Neurology. 1998; 51:1546–1554. [PubMed: 9855500]

Nugent WR. The role of prevalence rates, sensitivity, and specificity in assessment accuracy: Rolling the dice in social work process. Journal of Social Service Research. 2005; 31:51–75.

O'Bryant SE, Lucas JA. Estimating the predictive value of the Test of Memory Malingering: An illustrative example for clinicians. The Clinical Neuropsychologist. 2006; 20:533–540. [PubMed: 16895864]

Rabin LA, Pare' N, Saykin AJ, Brown MJ, Wishart HA, Flashman LA, Santulli RB. Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. Neuropsychology, Development, and Cognition. Section B: Aging, Neuropsychology and Cognition. 2009; 16:357–376.

Roma'n GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, Scheinberg P. Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology. 1993; 43:250–260. [PubMed: 8094895]

Salmon DP, Thomas RG, Pay MM, Booth A, Hofstetter CR, Thal LJ, Katzman R. Alzheimer's disease can be accurately diagnosed in very mildly impaired individuals. Neurology. 2002; 59:1022–1028. [PubMed: 12370456]

Schoenberg MR, Dawson KA, Duff K, Patton D, Scott JG, Adams RL. Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. Archives of Clinical Neuropsychology. 2006; 21:693–703. [PubMed: 16987634]

Shapiro DE. The interpretation of diagnostic tests. Statistical Methods in Medical Research. 1999; 8:113–134. [PubMed: 10501649]

Stern, RA.; White, T. Neuropsychological Assessment Battery. Psychological Assessment Resources; Lutz, FL: 2003.

Weintraub S, Salmon D, Mercaldo N, Ferris S, Graff-Radford NR, Chui H, Morris JC. The Alzheimer's Disease Centers' Uniform Data Set (UDS): The neuropsychologic test battery. Alzheimer Disease and Associated Disorders. 2009; 23:91–101. [PubMed: 19474567]

Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund L-O, Petersen RC. Mild cognitive impairment—beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. Journal of Internal Medicine. 2004; 256:240–246. [PubMed: 15324367]

Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3:32–35. [PubMed: 15405679]

TABLE 1

Diagnostic Makeup of AD and Non-AD Groups

Clinical Consensus Diagnosis	AD Group n (% of total)	Non-AD Group n (% of total)
Cognitively Healthy	—	75 (27.2%)
Without complaint	—	63
With self- or informant complaint	—	12
Ambiguous, non-MCI	—	16 (5.8%)
MCI	—	113 (40.9%)
"Possible" MCI[a]	—	70
"Probable" MCI[b]	—	43
Alzheimer's Disease[c]	65 (23.6%)	—
Possible AD	54	—
Probable AD	11	—
Other dementia etiologies	—	7 (2.5%)
Vascular Dementia[d]	—	1
Frontotemporal Dementia[e]	—	1
Dementia of unknown etiology	—	4
Total	65 (23.6%)	211 (76.4%)

AD = Alzheimer's disease; MCI = Mild cognitive impairment.

[a] Met (diagnostic criteria for MCI (Winblad et al., 2004) except for cognitive complaint.

[b] Winblad et al., 2004.

[c] McKhann et al., 1984.

[d] Román et al., 1993.

[e] Neary et al., 1998.

TABLE 2

Participant Demographics and NAB Test Performance

	Non-AD Group Mean (SD)	AD Group Mean (SD)	Effect Size d	Significance Test	
				t	P
Age	74.4 (8.0)	77.4 (8.5)	0.4	−2.6[*]	.012
Education	15.3 (2.9)	14.9 (3.1)	−0.1	1.0	.327
Driving Scenes[a]	38.3 (8.0)	25.9 (8.5)	−1.5	9.5[**]	<.001
Bill Payment[b]	17.3 (2.2)	13.7 (4.0)	−1.1	5.6[**]	<.001
Daily Living Memory– Immediate Recall[c]	40.4 (5.6)	25.6 (7.4)	−2.3	10.3[**]	<.001
Daily Living Memory– Delayed Recall[c]	12.2 (3.5)	2.3 (3.5)	−2.8	14.0[**]	<.001
Screening Visual Discrimination[d]	4.8 (1.1)	3.7 (1.4)	−0.9	5.6[**]	<.001
Screening Design Construction[e]	6.7 (4.1)	4.0 (3.5)	−0.7	4.8[**]	<.001
Judgment[f]	15.0 (2.5)	11.6 (4.4)	−1.0	5.6[**]	<.001

AD = Alzheimer's disease; d = Cohen's d.

[a]Non-AD group, n = 202; AD Group, n = 51.

[b]Non-AD group, n = 197; AD Group, n = 42.

[c]Non-AD group, n = 163; AD Group, n = 29.

[d]Non-AD group, n = 208; AD Group, n = 58.

[e]Non-AD group, n = 203; AD Group, n = 55.

[f]Non-AD group, n = 185; AD Group, n = 54.

[*]p<.05.

[**]p<.01.

TABLE 3

Sensitivity, Specificity, and Positive and Negative Predictive Values for NAB Subtests at Cutoff Scores That Jointly Maximize Sensitivity and Specificity

| | | | | Base Rate | | | | | | | | |
| | | | | .05 | | .10 | | .33 | | .50 | |
Test	Cutoff Score ( )	Sensitivity	Specificity	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
Driving Scenes	29	.65	.90	.25	.98	.42	.96	.76	.84	.87	.72
Bill Payment	15	.67	.87	.21	.98	.36	.96	.72	.84	.84	.73
Daily Living Memory–Immediate Recall	32	.86	.90	.31	.99	.49	.98	.81	.93	.90	.87
Daily Living Memory–Delayed Recall	8	.97	.88	.30	1.00	.47	1.00	.80	.98	.89	.97
Screening Visual Discrimination	3	.45	.88	.16	.97	.29	.94	.65	.76	.79	.62
Screening Design Construction	5	.76	.54	.08	.97	.16	.95	.45	.82	.62	.69
Judgment	12	.61	.88	.21	.98	.36	.95	.71	.82	.84	.69

PPV = Positive predictive value; NPV = Negative predictive value.

TABLE 4

Sensitivity, Specificity, and Positive and Negative Predictive Values fat NAB Subtests at Cutoff Scores That Maximize Sensitivity

Test	Cutoff Score ( )	Sensitivity	Specificity	Base Rate							
				.05		.10		.33		.50	
				PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
Driving Scenes	41	.98	.35	.07	1.00	.14	.99	.43	.97	.60	.95
Bill Payment	18	.98	.33	.07	1.00	.14	.99	.42	.97	.59	.94
Daily Living Memory–Immediate Recall	38	.97	.66	.13	1.00	.24	.99	.58	.98	.74	.96
Daily Living Memory–Delayed Recall	8	.97	.88	.30	1.00	.47	1.00	.80	.98	.89	.97
Screening Visual Discrimination	5	.95	.29	.07	.99	.13	.98	.40	.97	.57	.85
Screening Design Construction	14	.98	.03	.05	.97	.10	.93	.33	.75	.50	.60
Judgment	18	.98	.05	.05	.98	.10	.96	.34	.84	.51	.71

PPV = Positive predictive value; NPV = Negative predictive value.

TABLE 5

Sensitivity, Specificity, and Positive and Negative Predictive Values for NAB Subtests at Cutoff Scores That Maximize Specificity

Test	Cutoff Score ( )	Sensitivity	Specificity	Base Rate							
				.05		.10		.33		.50	
				PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
Driving Scenes	15	.10	1.00[a]	.51	.95	.69	.91	.91	.69	.95	.53
Bill Payment	6	.07	1.00[a]	.42	.95	.61	.91	.87	.68	.93	.52
Daily Living Memory–Immediate Recall	25	.45	.99	.70	.97	.83	.94	.96	.79	.98	.64
Daily Living Memory–Delayed Recall	1	.62	.99	.77	.98	.87	.96	.97	.84	.98	.72
Screening Visual Discrimination	2	.17	.97	.23	.96	.39	.91	.74	.70	.85	.54
Screening Design Construction	1	.27	.90	.12	.96	.23	.92	.57	.71	.73	.55
Judgment	6	.11	1.00[a]	.54	.96	.71	.91	.92	.69	.96	.51

PPV = Positive predictive value; NPV = Negative predictive value.

[a] Rounded from 0.995.

TABLE 6

Sensitivity and Specificity Values for Distinguishing MCI From Cognitively Healthy and AD Groups

| Test | MCI vs. Cognitively Healthy | | | | | | AD vs. MCI | | | | | |
| | Sensitivity Maximized | | | Specificity Maximized | | | Sensitivity Maximized | | | Specificity Maximized | | |
	Cutoff ( )	Sn	Sp	Cutoff ( )	Sn	Sp	Cutoff ( )	Sn	Sp	Cutoff ( )	Sn	Sp
Driving Scenes	52	.97	.10	30	.21	.98	41	.98	.31	22	.39	.95
Bill Payment	18	.78	.50	14	.15	.98	18	.98	.22	12	.29	.98
Daily Living Memory–Immediate Recall	48	.97	.10	34	.10	.98	38	.97	.58	29	.69	.97
Daily Living Memory–Delayed Recall	15	.97	.29	10	.48	.90	13	.97	.23	2	.66	.97
Screening Visual Discrimination	5	.86	.45	3	.26	.95	5	.95	.14	2	.17	.95
Screening Design Construction	14	.98	.02	1	.10	.95	14	.98	.02	1	.27	.90
Judgment	18	.94	.10	7	.00	.98	18	.98	.06	12	.61	.89

MCI = Mild cognitive impairment (Winblad et al., 1994); AD = Alzheimer's disease (McKhann et al., 1984): Sn = Sensitivity; Sp= Specificity.