

Confrontation naming does not add incremental diagnostic utility in MCI and Alzheimer's disease

JULIE A. TESTA,¹ ROBERT J. IVNIK,^{1,2} BRADLEY BOEVE,^{2,3} RONALD C. PETERSEN,^{2,3} V. SHANE PANKRATZ,⁴ DAVID KNOPMAN,^{2,3} ERIC TANGALOS,^{2,5} AND GLENN E. SMITH^{1,2}

¹Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota 55901

²Mayo Alzheimer's Disease Research Center, Rochester, Minnesota 55901

³Department of Neurology, Mayo Clinic, Rochester, Minnesota 55901

⁴Division of Biostatistics, Mayo Clinic, Rochester, Minnesota 55901

⁵Division of Community Internal Medicine, Mayo Clinic, Rochester, Minnesota 55901

(RECEIVED September 27, 2002; REVISED October 9, 2003; ACCEPTED November 10, 2003)

Abstract

As the incidence of dementia increases, there is a growing need to determine the diagnostic utility of specific neuropsychological tests in the early diagnosis of Alzheimer's disease (AD). In this study, the relative utility of Boston Naming Test (BNT) in the diagnosis of AD was examined and compared to the diagnostic utility of other neuropsychological measures commonly used in the evaluation of AD. Individuals with AD ($n = 306$), Mild Cognitive Impairment (MCI; $n = 67$), and cognitively normal subjects ($n = 409$) with at least 2 annual evaluations were included. Logistic regression analysis suggested that initial BNT impairment is associated with increased risk of subsequent AD diagnosis. However, this risk is significantly less than that imparted by measures of delayed recall impairments. A multivariate Cox proportional hazards regression analysis suggested that BNT impairment imparted no additional risk for subsequent AD diagnosis after delayed recall impairments were included in the model. Although BNT impairment occurred in all severity groups, it was ubiquitous only in moderate to severe dementia. Collectively these results suggest that although BNT impairments become more common as AD progresses, they are neither necessary for the diagnosis of AD nor particularly useful in identifying early AD. (*JINS*, 2004, *10*, 504–512.)

Keywords: Boston Naming Test, Alzheimer's disease, Mild Cognitive Impairment

INTRODUCTION

The establishment of methods to accurately detect early signs of Alzheimer's disease (AD) has become increasingly important as techniques for prevention or delay of dementia have been sought and developed. Animal models of dementia highlight advancement in this arena as scientists work to develop a vaccine against amyloid deposition, which may prevent neuritic plaque buildup (Schenk, 2002). Possible candidates for early diagnostic indices include neuroimaging, genetics and neuropsychology (Soininen & Scheltens, 1998).

Neuropsychological assessment of cognitive functioning has been shown to be useful in discriminating individuals who later develop AD (Albert et al., 2001; Bondi et al., 1994; Jacobs et al., 1995; Petersen et al., 1994; Rubin et al.,

1998; Tierney et al., 1996). In both retrospective and prospective studies, individuals who later develop AD have poorer initial performances on measures of prepositional naming, verbal memory, and abstract reasoning (Jacobs et al., 1995), delayed recall, verbal fluency, and auditory attention (Masur et al., 1994; Nielsen et al., 1999), and verbal retention (Elias et al., 2000). Bozoki and colleagues (2001) reported that nondemented patients with mild cognitive impairment in several domains including memory were twice as likely to develop AD over a period of 2–5 years when compared to those with only memory impairment. Comparisons among verbal fluency measures found category fluency to have the highest sensitivity and specificity when used to discriminate patients with AD from normal control subjects (Cerhan et al., 2002; Monsch et al., 1992). Similarly, measures of delayed recall, category fluency, and global cognitive status provided high sensitivity (96%) and specificity (93%) for differentiating between very mildly impaired AD patients and normal control subjects (Salmon

Reprint requests to: Julie A. Testa, Ph.D., Department of Psychiatry & Psychology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905.
E-mail: julietesta@hotmail.com

et al., 2002). Across studies, however, the most consistent predictor of whether an individual will be diagnosed with AD is initial performance on verbal memory tests (Albert et al., 2001; Bondi et al., 1994, 1999; Collie & Maruff, 2000; Rubin et al., 1998; Tierney et al., 1996).

The presence of word finding difficulties, i.e. anomia, in AD is well documented (Bayles & Kaszniak, 1987; Bayles et al., 1992; Bowles et al., 1987; Fisher et al., 1999). Deficits on confrontation naming tasks are reported to occur early in the course of dementia (Appell et al., 1982; Kirshner et al., 1984; Williams et al., 1989). Additionally, dementia severity has been reported to correlate strongly and positively with the degree of anomia (Bayles, 1982; Faber-Langendoen et al., 1988; Kaszniak et al., 1986; Kirshner et al., 1984). As such, confrontation naming tests, such as the Boston Naming Test (BNT; Kaplan et al., 1983), are commonly used to assist in the diagnosis of AD.

Despite a strong correlation between increasing dementia severity and word finding difficulties, anomia is not consistently present in patients with AD. Martin et al. (1986) identified a subgroup of AD patients who had preserved naming ability in the face of other significant cognitive deficits. Bayles and Tomoeda (1983) reported naming impairments in moderately demented AD patients but did not find significant deficits in mildly demented AD patients. Aphasia has been documented in 36% of patients with mild AD but was characterized primarily by an early decline in measures of comprehension and written expression with relatively preserved oral naming (Faber-Langendoen et al., 1988). In spite of these findings, many clinicians and researchers still consider naming impairments a requisite deficit in AD.

Some researchers (Albert et al., 2001; Elias et al., 2000; Masur et al., 1990; Small et al., 2000; Tierney et al., 1996), have not found confrontation naming particularly useful in discriminating individuals at baseline who are subsequently diagnosed with AD. In a prospective longitudinal study, Albert et al. (2001) followed 165 individuals classified as cognitively normal or questionable AD for a period of 3 years. A baseline battery of 17 neuropsychological tests in six different domains of cognitive function was administered including measures of memory, executive function, language (including the BNT), spatial ability, sustained attention, and general intelligence. Results suggested that neuropsychological measures of memory and executive function were most useful in discriminating individuals who converted from a cognitively normal state to AD. A retrospective cross-sectional discrimination study reported that patients with AD were best discriminated from normal control subjects by tests of delayed recall of figures and stories, while tests of confrontation naming, semantic fluency and design recognition were better for staging dementia severity and distinguishing mild or moderate AD from severe AD (Locascio et al., 1995). A retrospective study found that initial BNT and Block Design performances were not useful in differentiating a group of subjects with preclinical AD (i.e., subject who developed

AD during the next 12–16 months) from normal controls (Jacobson et al., 2002). In contrast, Jacobs and colleagues (1995) reported high positive predictive values (PPV) for the 15-item version of the Boston Naming Test (PPV = 90%), Immediate Recall on the Selective Reminding Test (PPV = 88%), and WAIS-R Similarities subtest (PPV = 79%) when used to discriminate subjects who were subsequently diagnosed with AD during a period of up to 4 years. The diverse findings regarding the utility of confrontation naming tasks may be related, in part, to study design, as those studies with shorter follow-up periods may be less sensitive to cognitive changes and cross-sectional studies may be confounded by cohort differences. Thus, the utility of confrontation naming tasks such as BNT in the diagnosis of AD remains to be elucidated.

Diagnostic utility refers to the ability of a test to differentiate persons with and without a specific disorder (Ivnik et al., 2001; Smith et al., 2003). Unlike statistics that rely on null-hypothesis testing to identify the impact of brain dysfunction on cognitive tests, analyses using diagnostic utility statistics establish a test's ability to make correct individual predictions (Smith et al., 2003). The diagnostic validity of a test can be reflected in indices such as sensitivity, specificity, hit rate, predictive values and likelihood ratios. For purposes of individual diagnosis these statistics can be translated into probability statements for specific test scores. These values will vary based on the characteristics of the condition of interest (COI).

Positive predictive value is the proportion of people who actually have a COI within the group predicted by the test as having the COI. Negative predictive value (NPV) is the proportion of people who do not have a COI within the group predicted by the test as not having the COI. These values are dependent on the base rate of the COI in the reference population. In settings with high base rates for the COI, it may be more difficult to identify conditions other than the COI. In those instances, NPV may have equal or greater importance than PPV. Once individual test scores have been obtained, likelihood ratios can be used to calculate predictive values for specific test scores to express the probability of disease in a patient. Likelihood ratios allow clinicians to express, based on a patient having earned a score of y on test z , that the patient's chance of having the COI are increased by (likelihood ratio of y) times.

The goal of the current study is to determine the relative utility of impairment on the BNT in the diagnosis of AD. The diagnostic utility of BNT is examined by using sensitivity and specificity to generate likelihood ratios for score values (see Fletcher et al., 1996, for review). These BNT values are compared to likelihood ratios for category fluency and delayed recall measures, which have been found to have high predictive value in previous studies. Additionally, the utility of these measures in predicting conversion to AD from either unimpaired or Mild Cognitive Impairment (MCI; Petersen et al., 1999) status is assessed. Based on previous research, we hypothesize that measures of de-

layed memory will have the strongest diagnostic utility when used in the assessment of suspected AD and will be a significant predictor of subsequently being diagnosed with AD. Due to the considerable variability found in prior research regarding the importance of other cognitive performances, we hypothesize that performance on category fluency and BNT will have weaker diagnostic utility and, as such, present less risk toward a future diagnosis of AD.

METHODS

Research Participants

Participants in either the Mayo Alzheimer's Disease Patient Registry (ADPR; AG 06786) or the Mayo Alzheimer's Disease Research Center (ADRC; AG 16574) were utilized in this analysis. The ADPR recruits research participants from Olmsted County, MN via the Mayo Department of Community Internal Medicine. The ADRC recruits participants from the upper Midwest region through the Mayo Section of Behavioral Neurology. In either case, patients presenting with a cognitive complaint generated by themselves, their family or their primary physician are recruited into the ADPR or ADRC as a potential cognitive impairment case. In contrast, a "cognitively normal" control group is recruited from people presenting to their primary doctor in the Mayo Department of Community Internal Medicine. These individuals are independently functioning, community-dwelling persons who have recently been examined by their personal physician and who have no active neurologic or psychiatric disorder with potential to affect cognition (Malec et al., 1993). All participants receive behavioral neurology evaluations, including mental status testing and extensive medical history review. They also receive neuroimaging and appropriate laboratory studies. To avoid circularity, neuropsychological data are not used in establishing normalcy at initial evaluation. Recruitment for these research projects has been described more extensively elsewhere (Petersen et al., 1990).

A diagnosis of cognitive impairment or normalcy is assigned after each evaluation through consensus meetings that include one or more board-certified behavioral neurologists, one or more board-certified clinical neuropsychologists, nurses, and psychometrists. In this study, a diagnosis of probable or possible AD was made in accordance with the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADSDA) criteria (McKhann et al., 1984). A diagnosis of MCI was made in accordance with criteria established by Petersen et al. (1999). ADPR/ADRC participants with other diagnoses (e.g., Lewy body dementia, vascular dementia) were excluded from this study.

Participants in the ADPR and ADRC receive serial neuropsychological evaluations. Only those with at least two evaluations were included in this study. Diagnosis was reassessed at each evaluation. As such, their diagnosis following their

initial evaluation (e.g., MCI) may have changed or *converted* to a different diagnosis (e.g., AD) at subsequent assessments. The diagnosis assigned to each participant at his or her last evaluation is considered the *criterion diagnosis* for this study. The diagnosis at last evaluation is considered more reliable because it is based on all available information such as laboratory results, neuroimaging, medical and neurologic examinations, behavioral observations, and neuropsychological testing.

As noted, participants were grouped according to the presence or absence of an AD diagnosis at their last evaluation. Since MCI patients do not meet AD diagnostic criteria they were grouped with the normal controls in the calculation of diagnostic utility statistics. Some researchers may choose to suppress MCI patients as neither normal nor AD; however, doing so inflates diagnostic utility statistics. To provide the most conservative assessment of diagnostic utility, we chose to include MCI patients in the Non-AD group. This study included 306 participants with an AD diagnosis at last evaluation and 476 non-AD participants (409 cognitively normal and 67 MCI). Demographic characteristics of the AD patients and non-AD are presented in Table 1. The groups did not differ significantly on age or education.

Procedure

The Clinical Dementia Rating (CDR; Morris, 1993) and the Dementia Rating Scale (DRS; Mattis, 1973) were used to assess dementia severity. Neuropsychological tests included the Boston Naming Test (BNT; Kaplan et al., 1983), a category fluency task (CF; Lucas et al., 1998), Wechsler Memory Scale-Revised, Logical Memory subtest (WMS-R; LM; Wechsler, 1987) and the Rey Auditory Verbal Learning Test (AVLT; Rey, 1964). These measures were included because they have been shown to have diagnostic utility in previous studies and were routinely in use in the ADPR/ADRC. All tests were administered by experienced psychometrists supervised by two ABPP-certified clinical neuropsychologists (G.E.S. and R.J.I.). Scores for BNT were compared with category fluency and two measures of delayed recall, Rey Auditory Verbal Learning Percent Retention (AVLT % Retention) and WMS-R Logical Memory Percent Retention (LM % Retention) scores. MOANS age-adjusted scaled scores (Ivnik et al., 1992; 1996) were used for each variable. Therefore, all four variables have a normative mean of 10 and a standard deviation of 3.

Statistics

Summaries of initial demographic and neuropsychological characteristics were computed for patients, separately for AD patients and non-AD subjects, as determined by the diagnosis at last follow-up. The summarized means, or percentages, were compared between the two diagnosis groups with *t* tests, or chi-square tests.

In addition to the mean comparisons of the MOANS age-corrected scaled scores for BNT, CF, and delayed recall

Table 1. Initial demographic and neuropsychological characteristics for AD patients and non-AD subjects by final diagnosis

	Non-AD subjects		AD patients		<i>p</i>
	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>	
Age	476	79.2 (7.2)	306	77.9 (8.4)	ns
Education	476	13.1 (3.1)	305	12.7 (3.0)	ns
Gender (% Male)		34%		32%	ns
No. of evaluations (range: 2–10)		2.8 (1.7)		2.5 (1.6)	
MMSE	475	28.1 (1.7)	297	23.6 (3.8)	<0.0001
CDR	307	0 (0.2)	210	0.7 (0.4)	<0.0001
DRS	474	134.6 (6.9)	301	117.6 (13.7)	<0.0001
LM % RET	421	9.5 (3.5)	271	4.4 (3.2)	<0.0001
AVLT % RET	440	9.8 (2.8)	287	5.8 (2.3)	<0.0001
BNT	473	10.4 (3.0)	301	6.6 (3.4)	<0.0001
CF	145	9.5 (2.6)	109	5.0 (2.4)	<0.0001

Note. MMSE = Mini-Mental Status Exam; DRS = Mattis Dementia Rating Scale total scores; LM % RET = Wechsler Memory Scale – Revised Logical Memory Percent Retention MOANS score; AVLT % RET = Auditory Verbal Learning Test Percent Retention MOANS score; BNT = Boston Naming Test MOANS score; CF = Category Fluency MOANS score.

data from initial evaluation, diagnostic cut-points for the several tests were selected and evaluated. The cut scores were chosen such that the selected value had the maximal diagnostic accuracy (i.e., highest rates of combined sensitivity and specificity). Along with the cut score, a number of diagnostic validity indices were reported for each of the tests. These included sensitivity, specificity, overall diagnostic accuracy or hit rate, likelihood ratios and odds ratios. Logistic regression was used to obtain estimates of odds ratios, and corresponding 95% confidence intervals. Further logistic regression analyses were conducted to compare odds ratios among the four diagnostic cutoff scores. Because each diagnostic test had been given to each individual, comparisons were made within individuals. This was achieved by using generalized estimating equations to account for repeated measurements within individuals, and testing the null hypothesis that the odds ratios for any two diagnostic cutoff scores were equal.

To further examine the relationship between naming performance and risk of development of AD, a stepwise Cox proportional hazards regression analysis was completed. The dependent measure in this analysis was time from initial evaluation to either AD diagnosis, or last follow-up. Participants with initial diagnoses of AD or other dementia were excluded from this analysis. Since date of “onset” of AD was impossible to calculate, time to conversion in those progressing to AD was identified as: (last nonconverted test date + 1) – initial test date. In the stepwise procedure, alphas to enter and remain in the model were set at $p < .05$. For this analysis cognitive scores were dichotomized into impaired *versus* non-impaired based on the cut scores used above. Three cognitive variables (BNT, LM and AVLT) were permitted in the preliminary model. CF data was available for only 21 converters and 144 nonconverters, significantly reduced the *n* for the analysis. Thus, CF was excluded from HA modelling.

RESULTS

Group Descriptive Statistics

Traditional means comparisons based on scores at initial evaluation are presented in Table 1. All differences are highly significant. For each task, AD patients performed more poorly than non-AD subjects. As expected, tests of significance in this study indicated that each cognitive measure differs, on average, between groups of cognitively normal and cognitively impaired people. Although the possibility exists that some individuals in the non-AD group had evidence of cognitive decline, as a group, their scores at initial evaluation were in the average and non-impaired range compared to available normative data.

Diagnostic Validity Statistics

Analyses of the MOANS age-corrected scaled scores for BNT, CF, and delayed recall data from initial evaluation using diagnostic validity indices including cut score, sensitivity, specificity, overall diagnostic accuracy or hit rate, and odds ratios are shown in Table 2. Odds ratios associated with the cut scores are also presented along with their corresponding 95% confidence intervals. Odds ratios are useful for assessing the diagnostic validity of a specific test in relation to a designated cut score (Smith et al., 2003). The odds ratio allows one to say that a person who scores at or below a scaled score of 7 on the BNT is 8 times more likely to be diagnosed with AD than a person who scores above 7. The confidence interval reflects sampling uncertainty and allows one to say, with 95% certainty, that the estimate of relative risk is between 4.6 and 10.9. The odds ratios for the cut-off scores of the other diagnostic tests are all higher than that of the BNT. The *p* values from the pairwise tests comparing BNT to the other diagnostic exams support that

Table 2. Diagnostic utility statistics for AD diagnosis at last evaluation

Measure	Cut score	Sens.	Spec.	Hit rate	PPV	NPV	Odds ratio	95% CI
BNT	7	62.8%	82.7%	73%	0.70	0.78	8.0	4.6–10.9
CF	7	83.5%	75.9%	80%	0.72	0.86	15.4	7.6–27.0
LM % RET	5	76.3%	86.1%	81%	0.78	0.85	19.6	16.9–25.5
AVLT % RET	7	84.3%	79.3%	82%	0.73	0.89	20.4	13.7–28.0

Note. LM % RET = Wechsler Memory Scale – Revised Logical Memory Percent Retention MOANS score; AVLT % RET = Auditory Verbal Learning Test Percent Retention MOANS score; BNT = Boston Naming Test MOANS score; CF = Category Fluency MOANS score; Sens. = Sensitivity; Spec. = Specificity; PPV = Positive Predictive Value; NPV = Negative Predictive Value; LR = Likelihood Ratio.

BNT is significantly different from AVLT % Retention ($p < .005$) and LM % Retention ($p < .004$) and marginally significant when compared to CF ($p < .087$). Thus, while initial BNT impairment is associated with increased risk of subsequent AD diagnosis, this risk is significantly less than that imparted by delayed recall impairments.

As explained above, likelihood ratios allow us to calculate the probability of a person having a COI for specific test scores. Likelihood ratios reflect the probability that a person obtaining a given test score has AD versus the probability that a person doesn't have AD (Cerhan et al., 2002; Ivnik et al., 2000; Sackett et al., 1991). Table 3 displays the likelihood ratios for BNT, category fluency, LM % Retention and AVLT % Retention at initial evaluation. One can conclude that the risk of being diagnosed with AD for a person who obtains a MOANS scaled score of 5 on BNT is increased 3.9 times over his or her baseline risk of AD. In contrast, a person with a scaled score of 5 on AVLT % Retention has an increased risk of 7.9. Thus, an identical scaled score on AVLT % Retention is associated with a higher likelihood that a person will be diagnosed with AD when compared with BNT.

The relationship between severity of dementia in AD (as defined by CDR) and impaired performance on BNT, category fluency, LM % retention and AVLT % retention at initial evaluation is displayed in Figure 1. This figure displays the percentage of AD patients classified as impaired

based on the identified cut scores grouped by increasing dementia severity. Figure 2 displays the percent of AD patients impaired based on the identified cut scores on one or more measures (BNT, LM % Retention and/or AVLT % Retention, Category Fluency) at the time of their initial evaluation. A majority of patients were impaired on all three measures at initial evaluation.

Next, we examined the characteristics of subjects at their initial evaluation compared to their last evaluation, shown in Tables 4 and 5. Subjects who were initially grouped as non-AD but were subsequently diagnosed with AD were labeled as *converters*. Preliminary comparison of converters and nonconverters at initial evaluation reveals no significant differences in age or education. As expected given the preponderance of initial MCIs in the convert group, CDR scores were higher ($t = -9.4, p < .0001$) and DRS Total scores were lower for the converters ($t = -7.0, p < .0001$).

Time to Event Analysis

To further examine the relationship between naming performance and risk of development of AD, the time to diagnosis of AD was examined using Cox proportional hazards regression analyses (Table 6). There were 86 conversion to AD events and 472 participants who did not receive a diag-

Table 3. Likelihood ratios (LR) for AD patients and non-AD subjects at initial evaluation on BNT, Category Fluency, WMS-R Logical Memory % Retention, and AVLT % Retention

Percentile ranges	MOANS scaled score	LRs			
		BNT	CF	LM % RET	AVLT % RET
≤2	≤4	20.6	26.7	8.4	15.0
3–5	5	3.9	3.5	1.7	7.9
6–10	6	2.3	2.2	0.8	3.8
11–18	7	1.5	1.0	0.5	1.1
19–28	8	0.9	1.0	0.3	0.7
29–40	9	0.8	0.1	0.4	0.1
41–59	10	0.4	0.2	0.2	0.2
>59	>10	<0.4	<0.1	<0.1	<0.1

Table 4. Comparison of diagnosis at initial evaluation to diagnosis at last evaluation

		Diagnosis following initial evaluation				Total
		Controls	MCI	AD	Other	
Diagnosis at last evaluation	Controls	409				409
	MCI	23	40	1	3	67
	AD	26	60	196	24	306
Total		458	100	197	27	782

nosis of AD. These individuals were censored at the time of their last exam in the analysis.

We first examined the relative hazard associated with initial status (i.e. control vs. MCI). As might be expected, initial MCI status was associated with very elevated risk relative to control status (relative risk, RR = 20.3) for conversion to dementia. Since initial hazard for the two groups was not proportional, we performed subsequent stepwise multivariate Cox proportional hazards analysis separate for controls and MCI. The results of these analyses are presented in Table 6. Due to missing data on one or more of the cognitive variables the control group included 400 participants (including 24 converters) while the MCI group included 91 participants (with 55) converters. In controls, only LM and AVLT entered the stepwise hazards modeling. Impairment on BNT did not impart statistically significant additional risk after the delayed recall variables had entered the model. In practical terms this model suggests that delayed recall impairments on AVLT and LM each appeared to increase three to five times the risk to progress to AD diagnosis over a given time interval. The presence or absence of an impaired BNT score imparted no further risk

after accounting for these other two scores. In MCI patients no cognitive impairment variable was significant in the stepwise hazards model. In the MCI sample 83% of participants were impaired on AVLT, 69% were impaired on LM, and 45% were impaired on BNT.

DISCUSSION

The primary purpose of the current study was to investigate the relative utility of impairment on the BNT for establishing the diagnosis of AD. For comparison purposes, the utility of BNT was compared to category fluency and two measures of delayed recall. Additionally, the utility of BNT in predicting the conversion from either an unimpaired or MCI status to AD was assessed. As expected, traditional univariate analysis revealed that AD patients and non-AD subjects (as classified at their last evaluation) differed significantly from each other at their initial evaluation on each test, with AD patients performing more poorly than non-AD subjects on category fluency, BNT and LM and AVLT. These findings validate the use of traditional significance tests to document how groups of people with defined conditions differ from each other. The findings do not, however, validate these measures for use in diagnostic classification (Smith et al., 2003).

To more directly assess the utility of BNT, category fluency, LM % Retention, and AVLT % Retention in diagnosing AD, diagnostic validity indices were calculated. Odds ratios for BNT were clearly significant. However, these odds ratios were also significantly less than the odds ratios for

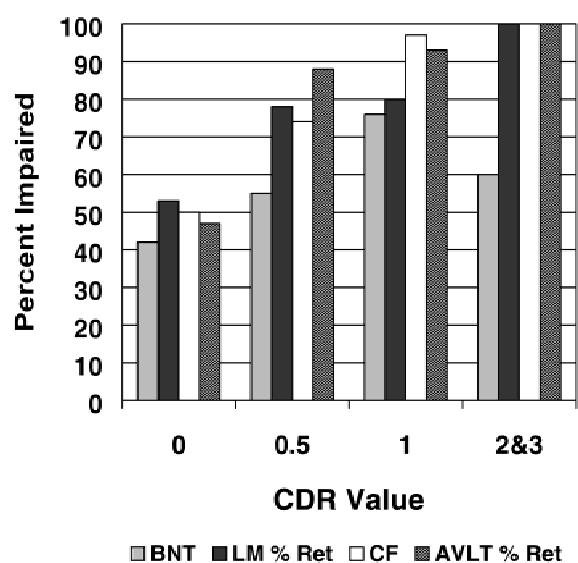


Fig. 1. Percent of AD patients impaired on BNT, Category Fluency, AVLT % Retention, and LM % Retention by CDR at initial evaluation. *Note.* CDR=0 in participants that were clinically normal at initial evaluation but converted to AD at last evaluation.

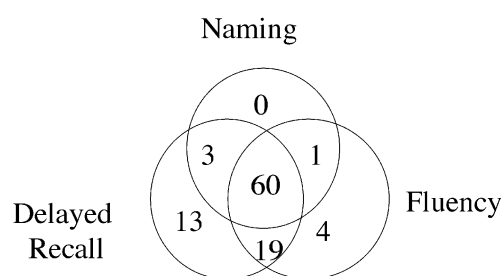


Fig. 2. Percent of AD patients at initial evaluation impaired on BNT, LM % Retention and/or AVLT % Retention, and Category Fluency. *One AD patient had no impairment on any of these measures at initial evaluation.

Table 5. Characteristics of converters and non-converters

	Converters	Non-converters	<i>p</i>
	(<i>N</i> = 86)	(<i>N</i> = 472)	
Age	79.91 (7.65)	79.25 (7.18)	n.s.
Education	13.67 (2.80)	13.13 (3.06)	n.s.
Gender (% male)	30%	34%	n.s.
MMSE	26.62 (2.25)	28.11 (1.72)	<i>p</i> < .0001
CDR	0.36 (0.23)	0.00 (0.16)	<i>p</i> < .0001
DRS	127.15 (9.41)	134.61 (6.89)	<i>p</i> < .0001
Follow-up interval, days*	813.16 (801.35)	1645.63 (929.14)	<i>p</i> < .0001

*For converters, the average interval between initial assessment and subsequent diagnosis of AD is reported. For non-converters, the average interval between initial assessment to the last available follow-up visit is reported.

delayed recall. Inspection of likelihood ratios suggests that at more impaired levels on the BNT, the risk associated with a subsequent diagnosis of AD climbs steeply. Conversely, for scores falling above the cutoff, the amount of “protection” conferred is less for BNT than the other measures.

At each level of dementia severity, impairments on measures of delayed recall and category fluency were more common than BNT impairments. Less than 1 in 5 non-AD patients had moderate or severe BNT impairments while more than 1 in 5 demented AD patients performed normally on the BNT. No AD patients had exclusive naming impairments at initial evaluation. More commonly, patients had either impairments in both delayed recall and naming, or in delayed recall exclusively at their initial evaluation. These findings are consistent with Bayles and Tomoeda (1983) and others (Faber-Langendoen et al., 1988; Martin et al., 1986) who found naming impairments to be significant in moderate but not mild AD. Thus, it appears that BNT impairment is not particularly necessary for detecting early AD.

A stepwise multivariate Cox proportional hazards analysis was conducted to determine the association of impairments on the cognitive variables with time to diagnosis of AD. Although previous research (e.g., Jacobs et al., 1995) has attempted to characterize the cognitive changes in pre-clinical AD, a group of patients with a formal diagnosis of MCI was not included in that analysis. The current study has the benefit of the inclusion of MCI patients, longer average follow-up, and the use of an optimized cut-score

for each particular test rather than a standard level of impairment (e.g., -2 *SD* below the mean). This allowed for maximal sensitivity and specificity for each neuropsychological measure used in the study. Additionally, when the outcome variable (e.g., progression to dementia) is time dependent, use of survival analysis methods is preferred.

Survival analysis results corroborate and extend the well-established view that early impairment in delayed recall is predictive of later development of AD. In some respects this is tautological. The diagnosis of AD requires the presence of memory impairment. So finding that memory impairment is present prior to and imparts risk for the subsequent diagnosis of AD should not be surprising. The question is really are there other early cognitive impairments that impart additional risk for the subsequent diagnosis of AD. In the present proportional hazards modeling with controls, BNT did not impart further risk beyond that present from the delayed recall variables alone. In MCI patients none of the variables entered the model. The failure of delayed recall measures to predict conversion in this group may reflect the ubiquity of memory impairments in MCI subjects. Note however that BNT impairment was also not significant in this model though present in slightly less than half of the MCI group. Verbal fluency measures have been found useful in predicting AD in individuals (Cerhan et al., 2002) and may perform better than BNT in this regard. Unfortunately, the low number of people with CF scores precluded that analysis here.

From an economic standpoint, early detection and treatment of individuals at risk for AD is vitally important. With

Table 6. Cox proportional hazards analysis of time to AD diagnosis

Cohort	Impaired score	Step entered	Parameter	Wald	<i>p</i>	Relative risk	95% CI for RR	
				chi-squared		(RR)		
Controls	LM % RET	1	1.58	13.1	.0003	4.8	2.1–11.3	
	AVLT % RET	2	1.07	5.4	.02	2.9	1.2–7.2	
MCI	No vari	No variables remained in model						

Note. CI = confidence interval.

the average age of death in the United States at 78 years, and the prevalence of AD increasing exponentially each decade after 60 years of age, it is estimated that 14 million individuals will have AD by the year 2050 (Katzman & Fox, 1999). Currently, there are approximately 1.7 to 4 million people in the US with AD (Brookmeyer et al., 1998; Hy & Keller, 2000), at an estimated cost of \$40,000 per year per patient. Delaying disease onset would not only improve the patients' quality of life but also have direct financial benefits for society as a whole.

As the fields of neuropsychology and medicine turn increasingly toward early diagnosis and prevention of AD, neuropsychologists need to develop and employ tests to aid in clinical decision-making. The use of evidence-based approaches in neuropsychological decision-making will allow us to tailor neuropsychological assessments by choosing tests proven to have high diagnostic utility. Additionally, tests with high diagnostic utility may be combined with other diagnostic indices to improve the assessment of individuals at risk for developing dementia. Bondi et al. (1999) reported that measures of delayed recall and ApoE- ϵ 4 allele status were significant and independent predictors of conversion to AD. Combined with neuroimaging markers of early hippocampal atrophy, genetic and neuropsychological data may increase clinicians' ability to identify individuals at high risk for developing dementia (Soininen & Scheltens, 1998).

In conclusion, the current study suggests that confrontation naming deficits are neither necessary nor sufficient findings for the early diagnosis of AD. Although naming impairments are common in moderate to severe AD and impart important functional limitations at any stage, deficits on BNT are not particularly useful for early diagnosing or determining future risk for subsequent diagnosis of AD.

ACKNOWLEDGMENTS

This work was supported by National Institute on Aging Grants (AG 06786 to the Mayo Alzheimer's Disease Patient Registry and AG 11687 to the Mayo Alzheimer's Disease Research Center). Portions of these analyses were presented at the Annual Conference of the International Neuropsychological Society, Hawaii 2003.

REFERENCES

Albert, M., Moss, M.B., Tanzi, R., & Jones, K. (2001). Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychological Society*, 7, 631–639.

Appell, J., Kertesz, A., & Fisman, M. (1982). A study of language functioning in Alzheimer's disease. *Brain and Language*, 17, 73–91.

Bayles, K.A. (1982). Language function in senile dementia. *Brain and Language*, 16, 265–280.

Bayles, K.A. & Tomoeda, C.K. (1983). Confrontation naming impairment in dementia. *Brain and Language*, 19, 98–114.

Bayles, K.A. & Kaszniak, A.W. (1987). *Communication and cognition in normal aging and dementia*. Boston: College Hill/Little, Brown, and Company.

Bayles, K.A., Tomoeda, C.K., & Trosset, M.W. (1992). Relation of linguistic communication abilities of Alzheimer's patients to stage of disease. *Brain and Language*, 42, 454–472.

Bondi, M., Monsch, A., Galasko, D., Butters, N., Salmon, D., & Delis, D. (1994). Preclinical cognitive markers of dementia of the Alzheimer type. *Neuropsychology*, 8, 374–384.

Bondi, M.W., Salmon, D.P., Galasko, D., Thomas, R.G., & Thal, L.J. (1999). Neuropsychological function and apolipoprotein E genotype in the preclinical detection of Alzheimer's disease. *Psychology of Aging*, 14, 295–303.

Bowles, N.L., Obler, L.K., & Albert, M.L. (1987). Naming errors in healthy aging and dementia of the Alzheimer type. *Cortex*, 23, 519–524.

Bozoki, A., Giordani, B., Heidebrink, J.L., Berent, S., & Foster, N.L. (2001). Mild cognitive impairments predict dementia in nondemented elderly patients with memory loss. *Archives of Neurology*, 58, 411–416.

Brookmeyer, R., Gray, S., & Kwas, C. (1998). Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *American Journal of Public Health*, 88, 1337–1342.

Cerhan, J.H., Ivnik, R.J., Smith, G.E., Tangalos, E.C., Petersen, R.C., & Boeve, B.F. (2002). Diagnostic utility of letter fluency, category fluency, and fluency difference scores in Alzheimer's disease. *Clinical Neuropsychology*, 16, 35–42.

Collie, A. & Maruff, P. (2000). The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. *Neuroscience and Biobehavioral Reviews*, 24, 365–374.

Elias, M.F., Beiser, A., Wolf, P.A., Au, R., White, R.F., & D'Agostino, R.B. (2000). The preclinical phase of Alzheimer disease: A 22-year prospective study of the Framingham cohort. *Archives of Neurology*, 57, 808–813.

Faber-Langendoen, K., Morris, J.C., Knesevich, J.W., LaBarge, E., Miller, J.P., & Berg, L. (1988). Aphasia in senile dementia of the Alzheimer type. *Annals of Neurology*, 23, 365–370.

Fisher, N.J., Rourke, B.P., & Bieliauskas, L.A. (1999). Neuropsychological subgroups of patients with Alzheimer's disease: An examination of the first 10 years of CERAD data. *Journal of Clinical and Experimental Neuropsychology*, 21, 488–518.

Fletcher, R.H., Fletcher, S.W., & Wagner, E.H. (1996). Diagnosis. In *Clinical epidemiology: The essentials* (pp. 42–75). Baltimore: Williams & Wilkins.

Hy, L.X. & Keller, D.M. (2000). Prevalence of AD among whites: A summary by levels of severity. *Neurology*, 55, 198–204.

Ivnik, R.J., Malec, J.F., Smith, G.E., Tangalos, E.G., & Petersen, R.C. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, STROOP, TMT, and JLO. *Clinical Neuropsychologist*, 10, 262–278.

Ivnik, R.J., Malec, J.F., Smith, G.E., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1992). Mayo's Older Americans Normative Studies: WMS-R norms for ages 56 to 94. *Clinical Neuropsychologist*, 6(Suppl.), 48–81.

Ivnik, R.J., Smith, G.E., Cerhan, J.H., Boeve, B.F., Tangalos, E.G., & Petersen, R.C. (2001). Understanding the diagnostic capabilities of cognitive tests. *Clinical Neuropsychologist*, 15, 114–124.

Ivnik, R.J., Smith, G.E., Petersen, R.C., Boeve, B.F., Kokmen, E., & Tangalos, E.G. (2000). Diagnostic accuracy of four approaches to interpreting neuropsychological test data. *Neuropsychology*, 14, 163–177.

Jacobs, D.M., Sano, M., Dooneief, G., Marder, K., Bell, K.L., & Stern, Y. (1995). Neuropsychological detection and character-

- ization of preclinical Alzheimer's disease. *Neurology*, *45*, 957–962.
- Jacobson, M.W., Delis, D.C., Bondi, M.W., & Salmon, D.P. (2002). Do neuropsychological tests detect preclinical Alzheimer's disease: Individual-test versus cognitive-discrepancy score analyses. *Neuropsychology*, *16*, 132–139.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia: Lea & Febiger.
- Kaszniak, A.W., Wilson, R.S., Fox, J.H., & Stebbins, G.T. (1986). Cognitive assessment in Alzheimer's disease: Cross-sectional and longitudinal perspectives. *Canadian Journal of Neurological Science*, *13*, 420–423.
- Katzman, R. & Fox, P.J. (1999). The worldwide impact of dementia: projections of prevalence and costs. In R. Mayeux & Y. Christen (Eds.), *Epidemiology of Alzheimer's disease: From gene to prevention* (pp. 1–17). Heidelberg, Germany: Springer-Verlag.
- Kirshner, H.S., Webb, W.G., & Kelly, M.P. (1984). The naming disorder of dementia. *Neuropsychologia*, *22*, 23–30.
- Locascio, J.J., Growdon, J.H., & Corkin, S. (1995). Cognitive test performance in detecting, staging, and tracking Alzheimer's disease. *Archives of Neurology*, *52*, 1087–1099.
- Lucas, J.A., Ivnik, R.J., Smith, G.E., Bohac, D.L., Tangalos, E.G., Graff-Radford, N.R., & Petersen, R.C. (1998). Mayo's older Americans normative studies: Category fluency norms. *Journal of Clinical and Experimental Neuropsychology*, *20*, 194–200.
- Malec, J.F., Ivnik, R.J., & Smith, G.E. (1993). Neuropsychology and normal aging. In R.W. Parks, R.F. Zec, & R.S. Wilson (Eds.), *Neuropsychology of Alzheimer's disease and other dementias* (pp. 81–111). New York: Oxford University Press.
- Martin, A., Brouwers, P., Lalonde, F., Cox, C., Teleska, P., Fedio, P., Foster, N.L., & Chase, T.N. (1986). Towards a behavioral typology of Alzheimer's patients. *Journal of Clinical and Experimental Neuropsychology*, *5*, 594–610.
- Masur, D.M., Fuld, P.A., Blau, A.D., Crystal, H., & Aronson, M.K. (1990). Predicting development of dementia in the elderly with the Selective Reminding Test. *Journal of Clinical and Experimental Neuropsychology*, *12*, 529–538.
- Masur, D.M., Sliwinski, M., Lipton, R.B., Blau, A.D., & Crystal, H.A. (1994). Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. *Neurology*, *44*, 1427–1432.
- Mattis, S. (1973). *Mattis Dementia Rating Scale (MDRS)*. Odessa, FL: Psychological Assessment Resources.
- 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.
- Monsch, A.U., Bondi, M.W., Butters, N., Salmon, D.P., Katzman, R., & Thal, L.J. (1992). *Archives of Neurology*, *49*, 1253–1258.
- Morris, J.C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, *43*, 2412–2414.
- Nielsen, H., Lolk, A., Andersen, K., Andersen, J., & Kragh-Sørensen, P. (1999). Characteristics of elderly who develop Alzheimer's disease during the next two years—A neuropsychological study using CAMCOG. The Odense Study. *International Journal of Geriatric Psychiatry*, *14*, 957–963.
- Petersen, R., Kokmen, E., Tangalos, E.G., Ivnik, R.J., & Kurland, L.T. (1990). Mayo Clinic's Alzheimer's Disease Patient Registry. *Aging*, *2*, 408–415.
- Petersen, R., Smith, G., Ivnik, R., Kokmen, E., & Tangalos, E. (1994). Memory function in very early Alzheimer's disease. *Neurology*, *44*, 867–872.
- Petersen, R., Smith, G., Waring, S., Ivnik, R., Tangalos, E., & Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology*, *56*, 303–308.
- Rey, A. (1964). *L'examen clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.
- Rubin, E., Storandt, M., Miller, P., Kinscherf, D., Grant, E., Morris, J., & Berg, L. (1998). A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of Neurology*, *55*, 395–401.
- Sackett, D.L., Haynes, R.B., Guyatt, G.H., & Tugwell, P. (1991). The interpretation of diagnostic data. *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston: Little, Brown & Company.
- Salmon, D.P., Thomas, R.G., Pay, M.M., Booth, A., Hofstetter, C.R., Thal, L.J., & Katzman, R. (2002). *Neurology*, *7*, 1022–1028.
- Schenk, D. (2002). Amyloid-beta immunotherapy for Alzheimer's disease: The end of the beginning. *National Review of Neuroscience*, *3*, 824–828.
- Small, B.J., Fratiglioni, L., Viitanen, M., Winblad, B., & Bäckman, L. (2000). The course of cognitive impairment in preclinical Alzheimer disease. *Archives of Neurology*, *57*, 839–844.
- Smith, G.E., Cerhan, J.H., & Ivnik, R.J. (2003). Diagnostic validity. In D. Tulskey & D. Saaklofske (Eds.), *Clinical interpretation of the WAIS-III and WMS-III*. New York: Academic Press.
- Soininen, H.S. & Scheltens, P. (1998). Early diagnostic indices for the prevention of Alzheimer's disease. *Annals of Medicine*, *30*, 553–559.
- Tierney, M., Szalai, J., Snow, W., Fisher, R., Nores, A., Nadon, G., Dunn, E., & St. George-Hyslop, P. (1996). Prediction of probable Alzheimer's disease in memory-impaired patients: A prospective longitudinal study. *Neurology*, *46*, 661–665.
- Wechsler, D. (1987). *Wechsler Memory Scale-Revised*. New York: The Psychological Corporation.
- Williams, B.W., Mack, W., & Henderson, V.W. (1989). Boston Naming Test in Alzheimer's disease. *Neuropsychologia*, *27*, 1073–1079.