Preclinical prediction of AD using neuropsychological tests

MARILYN S. ALBERT,1,2 MARK B. MOSS,3 RUDOLPH TANZI,2 AND KENNETH JONES4
Departments of Psychiatry 1 and Neurology 2, Massachusetts General Hospital, Harvard Medical School;
Department of Anatomy and Neurobiology, Boston University 3; and the 4Heller School of Social Policy, Brandeis University
(Received January 27, 2000; Revised January 16, 2001; Accepted January 16, 2001)

Abstract
Normals (N = 42) and patients with mild memory difficulty (N = 123) were given a neuropsychological test battery,
and then followed annually for 3 years to determine which individuals developed sufficient functional change that
they met clinical criteria for AD. Twenty-three of the 123 participants with mild memory difficulty converted to a
diagnosis of probable Alzheimer’s disease (AD) within 3 years of follow-up. Four of the 20 neuropsychological
measures obtained at baseline, were useful in discriminating the groups on the basis of their status 3 years after
the tests were given. The 4 discriminating tests pertained to assessments of memory and executive function. When
the controls were compared to the individuals with memory impairments who ultimately developed AD (the converters),
the accuracy of discrimination was 89%, based on the neuropsychological measures at baseline. The
discrimination of the controls from the individuals with mild memory problems who did not progress to the point
where they met clinical criteria for probable AD over the 3 years of follow-up (the Questionables) was 74% and
the discrimination of the questionables from the converters was 80%. The specific tests that contributed to these
discriminations, in conjunction with recent neuropathological and neuroimaging data from preclinical cases, have
implications for which brain regions may be affected during the prodromal phase of AD. (JINS, 2001, 7, 631–639.)

Keywords: Preclinical Alzheimer’s disease, Memory, Executive function, MCI

INTRODUCTION
The discrimination of those destined to develop Alzheimer’s disease (AD) from the larger pool of individuals with
mild memory loss is of increasing importance, as strategies for the prevention or delay of dementia are developed. Since
progressive difficulty with memory is the earliest sign of AD in most patients (Cummings & Benson, 1983; Moss
et al., 1986; Welsh et al., 1992), neuropsychological testing has been administered in an attempt to identify patients
before they progress to the point where they meet clinical criteria for AD. This approach to the early detection of AD
has, however, been complicated by the fact that mild memory impairments are quite common in the elderly (e.g., Light,
1991), and not everyone with subtle memory difficulties progresses to develop AD (Daly et al., 2000; Rubin et al.,
1989).

A number of studies have demonstrated that tests of memory can, nevertheless, prove useful in differentiating
individuals in the prodromal phase of AD from those who have memory problems that will not progress to AD within
a few years time (e.g., Bondi et al., 1994; Howieson et al., 1997; Jacobs et al., 1995; Petersen et al., 1994; Rubin et al.,
1998; Small et al., 1995; Tuokko et al., 1991). Tests that assess other cognitive domains (e.g., language function,
praxis) have also been reported to contribute to this discrimination, but the overall accuracy has been modest (De-
vanand et al., 1997; Masur et al., 1994; Small et al., 1997) or it was unclear if the results were specific to the diagnosis
of AD (Flicker et al., 1991). Only one study has reported that selected neuropsychological tests can predict with a
high degree of accuracy which individuals with memory problems would develop AD over time (Tierney et al., 1996).
In this study, which based prediction on tests given 2 years before AD was diagnosed, the best predictors were delayed
recall of a word list (the Rey Auditory Verbal Learning Test) and the mental manipulation of well learned se-
quences (the Mental Control subtest from the Wechsler Mem-
ory Scale). Conversely, at least one study has reported that neuropsychological testing does not help in the prediction of who will develop AD over time (Bowen et al., 1997). Taken together, these studies suggested that memory tests, particularly those that emphasized retention of information over delays, would be useful in predicting who, in the present study, would develop AD over time.

We also hypothesized that tests of executive function would contribute significantly to the discrimination of individuals in the prodromal phase of AD. This hypothesis was based on two earlier studies indicating that tests of executive function were impaired in patients with mild AD, when other cognitive domains (apart from memory) did not show significant deficits (Grady et al., 1988; Lafleche & Albert, 1995). In the latter study, a broad range of executive function tests were given and it was found that those requiring set shifting, sequencing and self-monitoring were particularly impaired, whereas those that required abstraction and concept formation were only marginally affected in very mild AD patients (Lafleche & Albert, 1995). A similar finding was therefore anticipated in the present study.

To test these hypotheses, we administered a neuropsychological battery to 165 participants, 123 of whom met criteria for questionable AD when the study began. The battery evaluated all major domains of cognitive ability, including memory, executive function, language, spatial ability, attention, and general intelligence. After 3 years of follow-up, 23 of the 123 questionable individuals had progressed to the point where they met clinical research criteria for probable AD (McKhann et al., 1984). It was therefore possible to determine which neuropsychological measures at baseline could be used to predict the status of the participants at the end of the following 3 years.

RESEARCH PARTICIPANTS AND METHODS

Participant Recruitment and Selection Procedures

A total of 165 individuals were participants in the present study. They had been recruited through the print media (rather than from a clinic or other medical referral source). The advertisements for participants indicated that a research study was seeking individuals both with and without memory difficulty. Volunteers who responded to the ads then underwent a multistage screening procedure.

To be included in the study, participants needed to be 65 and over, to be free of significant underlying medical, neurologic or psychiatric illness, to have a collateral source who was willing to participate on an annual basis, to have a Clinical Dementia Rating (CDR; Hughes et al., 1982) of either normal (CDR = 0) or questionable AD (CDR = 0.5), and to be willing to participate in the study procedures. At baseline, the study procedures included a medical examination (consisting of a physical exam and medical history, EKG and standard laboratory tests), a semistructured interview, neuropsychological testing, a magnetic resonance imaging (MRI) scan, a single photon emission computed tomography scan (SPECT), and blood for genetic analysis. Only the semistructured interview was repeated annually; the remaining study procedures were repeated in a subset of the participants. All provided informed consent prior to the initiation of the study.

The multistage screening procedures were designed to yield a sample that consisted of 165 individuals meeting specific health requirements, approximately \( \frac{1}{2} \) of whom met CDR criteria for normal and \( \frac{1}{2} \) of whom were questionable. All consecutive individuals who met the study criteria were enrolled until the target sample size was obtained. A total of 1,095 individuals were screened in order to meet the recruitment goals. Participants were excluded for a variety of reasons, including the presence of specific illnesses (e.g., insulin dependent diabetes) or medications (e.g., Percocet, Ambien; 38%), the absence of a collateral source (13%), or severity of cognitive impairment (1%).

The primary goal of the study as a whole was the examination of cognitive, brain structure/function and genetic factors involved in the prodromal phase of AD. The study population was therefore intentionally enriched with a large number of individuals who had an increased likelihood of developing AD within a few years (i.e., those with a CDR rating of 0.5). For purposes of comparison, a smaller number of controls were also evaluated. The absolute number of individuals in the cohort was based on estimates of the number of questionable individuals who would meet criteria for AD within a few years (Rubin et al., 1989), and the power needed to address the primary hypotheses of interest.

The CDR ratings, central to the categorization of the participants, were derived from the annual semistructured interview, which was administered to each participant and his or her collateral source. This interview was based on the Initial Subject Protocol which was used to the initial development of the CDR scale (Hughes et al., 1982). It includes a semistructured set of questions regarding functional status administered to both the respondent and collateral, and a neurologic, psychiatric and mental status evaluation of the respondent. In the present study, each interview was administered by a skilled clinician (e.g., psychiatrist, neuropsychologist, physician’s assistant) and took approximately 1 \( \frac{1}{2} \) to 2 h to complete. The mean interrater reliability of the CDR ratings was high \( (r = 0.99, p < .0001) \), as was the interrater reliability of the 6 CDR subcategories \( (r = .90) \) that were used to generate the overall CDR rating (Daly et al., 2000).

Participant Baseline Characteristics

At baseline, the subjects were divided into two groups, based on their functional status, as indicated above. One group consisted of 42 participants with normal cognition (CDR = 0.0) and the other group consisted of 123 individuals with questionable AD (CDR = 0.5). They had a mean age of
71.4 and 72.2, respectively. The educational level of the two groups was equivalent (14.4 years and 14.9 years, respectively), as was the mean Mini-Mental State Exam (MMSE; Folstein et al., 1975) score (29.4 ± 0.7 and 29.1 ± 1.2, respectively). The gender distribution within both groups was also similar; approximately 60% female and 40% male.

**Participant Follow-Up Characteristics**

After 3 years of follow-up, 9 participants were deceased. One of these individuals died prior to the first follow-up assessment and is not included in this report. For those who remained alive, the annual follow-up rate was 99%.

Based on their 3-year trajectory of functional change, the participants could be categorized into six groups:

1. **Group 1: Normals**: Participants who had normal cognition at baseline (CDR = 0) and continued to be categorized as normal at follow-up (n = 32). This represented 76% of the normal participants. (Ten of 42 participants with a CDR rating of zero at baseline were categorized as questionable after 3 years of follow-up), but none converted to AD.

2. **Group 2: Questionables**: Participants who met criteria for questionable AD at baseline (CDR = 0.5) and were still categorized as CDR = 0.5 after 3 years of follow-up (n = 91). This represented 73% of the questionable participants.

3. **Group 3: Converters**: Participants who met CDR criteria for questionable AD at baseline, but progressed to the point where they were coded CDR = 1 within 3 years of follow-up and met NINCDS/ADRDA criteria for probable AD (McKhann et al., 1984; n = 23). (The annual medical, neurologic, psychiatric and laboratory evaluation was augmented, as needed, to assure the participants met these criteria. The participants who converted to AD at follow-up were slightly older than those who did not (73.0 and 71.9, respectively). The proportion of males was also slightly greater among those who were diagnosed with AD on follow-up than in those who remained questionable (52% vs. 48%). Neither of these group differences was statistically significant. Two converters are now deceased and an autopsy was obtained on one of them, which confirmed a diagnosis of definite AD (NIA-Reagan Work Group, 1997). The converters represented 19% of the questionable participants, thus, the annual conversion rate among the questionables was about 6% per year. It should be emphasized that none of the controls in the present study converted to AD during the follow-up period.

4. **Group 4: Non-AD**: Participants who were questionable at baseline and were coded CDR = 1 on follow-up but did not meet clinical research criteria for probable AD (n = 3). These 3 individuals had strokes.

5. **Group 5: Declining normals**: Participants who were categorized as normal at baseline but met criteria for questionable AD within 3 years of follow-up (n = 10). None of these individuals converted to AD within the 3-year follow-up.

6. **Group 6: Improvers**: Participants who were categorized as questionable AD at baseline but subsequently met criteria for normal cognition (n = 6). Other studies that have followed cohorts such as those in the present study and have, likewise, found a small number of individuals who appear to fluctuate in their functional status (Rubin et al., 1989; Flicker et al., 1991).

The present report focuses on the participants in Groups 1 to 3. Only a small number of secondary analyses were applied to the other groups. An examination of these groups will be the subject of separate publications.

**Neuropsychological Procedures**

The neuropsychological battery that was administered to the participants at baseline consisted of 20 test scores based on 17 tests: (1) five memory tests (California Verbal Learning Test, Cued Selective Reminding Test, Rey-Osterrieth Complex Figure Test, Delayed Word Recall Test, and Visual Reproduction Subtest of the Wechsler Memory Scale; Delis et al., 1987; Grober & Buschke, 1987; Knopman & Ryberg, 1989; Rey, 1941; Wechsler, 1988); (2) six tests of executive function (Trail Making Test, Part B; Stroop Interference Test; Self-Ordering Test; Porteus Mazes; Alpha Span Test; and Digit Span Backward; Craik, 1986; Petrides & Milner, 1982; Porteus 1959; Reitan, 1958; Stroop, 1935; Wechsler, 1988); (3) three language tests (Controlled Word Association Test for letters and for categories; 15 items from the Boston Naming Test; Benton & Hamsher, 1976; Kaplan et al., 1982); (4) two tests of spatial ability (copying the Rey Complex Figure; copying the figures from the Wechsler Memory Scale; Rey, 1941; Wechsler, 1945); (5) three tests of sustained attention (Digit Span Forward; Trail Making Test, Part A; Cued Reaction Time; Baker et al., 1985; Reitan, 1958; Wechsler, 1988); and (6) an assessment of general intelligence (estimated IQ based on a reduced version of the Wechsler Adult Intelligence Scale–Revised; Satz & Mogel, 1962). These tests were not administered by the same individuals who conducted the interview that was used to generate the CDR ratings, and the test scores were not used in the assignment of the CDR ratings.

**Genetic Assessment**

The apolipoprotein (APOE) gene was also examined in the participants (n = 160) because the E4 allele of this gene is overrepresented in AD patients compared to the general population (Saunders et al., 1993), and is now widely recognized as a risk factor for AD. We therefore sought to determine whether APOE status, either alone or in combi-
nation with the neuropsychological measures, was useful as a predictor of which individuals would convert to AD over time.

Data Analysis

The neuropsychological data were analyzed by analysis of variance (ANOVA) and discriminant function analysis (Press, 1972). Sensitivities and specificities (Bland & Altman, 1986) were also calculated, where appropriate. It should be emphasized that the neuropsychological data presented here were obtained at Baseline, but the groups used in the analyses were based on the status of the participants following three annual follow-up visits.

RESULTS

Comparison of Neuropsychological Scores Among Groups Based on Status After 3 Years of Follow-Up

The 20 neuropsychological measures from baseline were first examined individually with three-way ANOVAs that compared performance among Groups 1 to 3. Statistically significant ANOVAs were followed by planned pairwise comparisons between the groups (i.e., the controls vs. converters, controls vs. questionables, and questionables vs. converters), using a Scheffé adjustment for multiple comparisons.

Almost all of the three-way ANOVAs were statistically significant ($p < .01$ or greater), but the number of pairwise comparisons that were significant were relatively small. The mean test scores for each group and the results of the planned comparisons are presented in Table 1. These analyses were designed to highlight the major differences among the groups and are intended as descriptive statistics only.

Only one measure per test was used for each of the cognitive domains in order to eliminate the colinearity that would occur if multiple measures from the same test were employed (Pedhazur, 1997). When more than one measure per test was available (e.g., the Selective Reminding Test, the California Verbal Learning Test), the specific measure selected was based on theoretical considerations.

As shown in Table 1, tests of memory, executive function, and language demonstrated significant differences among the groups on the basis of the pairwise comparisons. There were no significant differences between the groups on tests of spatial skill or attention.

We then examined the baseline test measures to determine if the questionables and the converters differed by more than 1.5 standard deviations from the controls. The purpose of this analysis was to compare our study population with those categorized as having mild cognitive impairment (MCI). MCI is a concept that has recently been introduced to refer to individuals who have memory complaints, normal activities of daily living, have normal general cognitive function but are memory impaired for their age, and are not demented (Petersen et al., 1999). It is typical for individuals with MCI to score 1.5 standard deviations below the mean of their peer group on tests of memory, and, in fact, some research studies use this range of performance as one of the selection criteria for MCI (Petersen, 2000).

Of the 20 test scores in the neuropsychological battery, only three differed between the controls and the converters at baseline by 1.5 standard deviations or more (see Table 1). The Total Learning Score on the California Verbal Learning Test and the total score on the Self-Ordering Test differed by 1.5 standard deviations between the controls and the converters. In addition, time to completion on Part B of the Trail Making Test differed by more than 2 standard deviations between the controls and the converters. However, no test differed by 1.5 standard deviations between the controls and the questionables at baseline, although the Total Learning Score on the California Verbal Learning Test, differed by 1 standard deviation between these groups.

Neuropsychological Performance From Baseline as Predictor of Group Membership After 3 Years of Follow-Up

We then examined the manner in which the 20 neuropsychological test scores, taken in combination, best differentiated the groups. To accomplish this, several discriminant function analyses were performed. This method of analysis was used to avoid selecting variables that appeared significantly different either on the basis of the post-hoc comparisons or the 1.5 standard deviation criterion, and thus capitalizing on chance.

The first discriminant analysis was conducted to determine whether the 20 test scores in the battery, when taken together, significantly differentiated the three groups from one another (controls vs. questionables vs. converters). This discriminant function also included a measure of age, gender and years of education for each participant, in order to adjust for any possible differences between the groups based on these variables. This overall analysis was highly statistically significant ($\chi^2 = 86.03, df = 40, p < .00001$). It should be noted that the statistical power of such an analysis is based on the overall sample size, rather than the size of any of the individual groups (Press, 1972). Moreover, the homogeneity of the group variances was examined and found to be satisfactory (assuring that the modest size of some of the groups did not compromise the analysis).

We then performed a stepwise discriminant function analysis, in order to select the neuropsychological variables that best differentiated the controls, questionables, and converters from one another. The term best refers here to the process by which each variable is selected by the algorithm, and the order by which the variable improves the significance of the overall function. In the stepwise analysis, age, gender and educational attainment were entered at the first step, and then the algorithm of the discriminant function
selected the variables that, when combined, best differentiated the groups from one another ($\chi^2 = 72.91$, $df = 14$, $p < .00001$). Four of the 20 test scores were selected as the best discriminators between the controls, questionable, and converters. They were (1) the total learning score on the California Verbal Learning Test (CVLT), (2) the immediate recall of the figures from the Wechsler Memory Scale (F–WMS), (3) the time to completion on Part B of the Trail Making Test (TMT), and (4) the total score on the Self-Ordering Test (SOT). The covariates entered at the first step (age, gender, and years of education) were not significant. It should be noted that one of these measures (F–WMS) was not significantly different among the groups in the pairwise comparisons, but when all of the measures were combined with one another it improved the accuracy of discrimination significantly. Likewise, several measures that were significantly different between the groups in the pairwise comparisons were not selected because they did not significantly contribute to the discrimination of the groups (e.g., letter fluency).

Three separate post-hoc discriminant functions were then performed in which these four tests were entered and the three groups were compared, pairwise. Age, gender, and years of education were entered at the first step in each of these analyses, even though the overall discriminant function indicated that they did not significantly contribute to the discrimination. This was done to assure that any subtle, though non-significant, effect of these variables was reduced as much as possible. The results of the stepwise discriminant functions can be briefly summarized as follows:

1. **Controls versus converters**: The discriminant function comparing the controls and the converters yielded one function that was highly significant ($\chi^2 = 39.8$, $df = 7$, $p < .00001$). Three of the four neuropsychological variables significantly differentiated the groups: the total score on the CVLT ($p < .005$), the immediate recall of the figures on the WMS ($p < .007$), and the total time to completion on Part B of the TMT ($p < .016$). The over-

### Table 1. Mean (SD) scores and group differences on neuropsychological tests among controls, questionable, and converters

<table>
<thead>
<tr>
<th>Domain/Test</th>
<th>Controls</th>
<th>Questionables</th>
<th>Converters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$(SD)$</td>
<td>$M$</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVLT Total $\sim$</td>
<td>58.3 (8.2)$#^*$</td>
<td>49.7 (9.7)$^*^\wedge$</td>
<td>44.8 (11.2)$^#^\wedge$</td>
</tr>
<tr>
<td>Selective Reminding</td>
<td>48.6 (5.7)$#$</td>
<td>45.3 (6.3)$\wedge$</td>
<td>40.0 (10.8)$^#^\wedge$</td>
</tr>
<tr>
<td>Delayed Word Recall</td>
<td>6.0 (2.1)</td>
<td>5.8 (1.8)</td>
<td>5.0 (1.9)</td>
</tr>
<tr>
<td>WMS Figure Recall</td>
<td>8.5 (2.7)</td>
<td>8.3 (3.8)</td>
<td>7.7 (3.5)</td>
</tr>
<tr>
<td>Rey Delayed Recall</td>
<td>35.8 (10.6)</td>
<td>34.9 (13.9)</td>
<td>31.9 (17.3)</td>
</tr>
<tr>
<td><strong>Executive function</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail Making Part B $\sim$</td>
<td>82.9 (33.6)$#$</td>
<td>102.3 (50.3)$^\wedge$</td>
<td>164.5 (95.9)$^#^\wedge$</td>
</tr>
<tr>
<td>Self Ordering Total $\sim$</td>
<td>9.6 (3.5)$#$</td>
<td>11.7 (4.0)$^\wedge$</td>
<td>15.0 (5.9)$^#^\wedge$</td>
</tr>
<tr>
<td>Stroop Interference</td>
<td>5.8 (7.3)</td>
<td>4.1 (7.8)</td>
<td>3.5 (5.0)</td>
</tr>
<tr>
<td>Porteus Mazes</td>
<td>14.8 (2.3)</td>
<td>14.9 (1.8)</td>
<td>14.6 (2.5)</td>
</tr>
<tr>
<td>Alpha Span</td>
<td>4.8 (0.8)$#$</td>
<td>4.7 (0.8)$^\wedge$</td>
<td>4.0 (0.9)$^#^\wedge$</td>
</tr>
<tr>
<td>Digit Span Backward</td>
<td>5.8 (1.6)$#$</td>
<td>5.2 (1.6)</td>
<td>5.0 (1.0)</td>
</tr>
<tr>
<td><strong>Language ability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naming</td>
<td>14.9 (0.4)</td>
<td>14.6 (0.8)</td>
<td>14.5 (0.7)</td>
</tr>
<tr>
<td>Letter Fluency</td>
<td>88.0 (18.9)$^*^\wedge$</td>
<td>81.8 (22.0)</td>
<td>66.7 (32.7)$^#^\wedge$</td>
</tr>
<tr>
<td>Category Fluency</td>
<td>20.8 (5.4)$#$</td>
<td>19.1 (5.6)</td>
<td>17.5 (5.9)$#$</td>
</tr>
<tr>
<td><strong>Spatial ability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMS Figure Copy</td>
<td>12.4 (1.1)</td>
<td>12.4 (1.1)</td>
<td>12.4 (1.4)</td>
</tr>
<tr>
<td>Rey Figure Copy</td>
<td>66.5 (4.8)</td>
<td>67.3 (3.5)</td>
<td>65.5 (6.6)</td>
</tr>
<tr>
<td><strong>Sustained attention</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span Forward</td>
<td>7.3 (1.4)</td>
<td>6.8 (1.3)</td>
<td>6.5 (1.3)</td>
</tr>
<tr>
<td>Trail Making Part A</td>
<td>56.8 (16.4)</td>
<td>62.9 (18.4)</td>
<td>70.1 (34.1)</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>370.9 (42.4)</td>
<td>389.2 (42.3)</td>
<td>387.3 (57.6)</td>
</tr>
<tr>
<td><strong>Intelligence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced WAIS–R</td>
<td>129.6 (15.7)$#$</td>
<td>124.6 (15.9)</td>
<td>118.4 (17.6)$#$</td>
</tr>
</tbody>
</table>

$\sim =$ tests that differed between controls and converters by 1.5 $SD$.

$\#$ = significant difference between controls and converters.

$^*$ = significant difference between controls and questionable.

$^\wedge$ = significant difference between questionable and converters.
2. **Controls versus questionables**: The discriminant function comparing the controls and the questionables yielded one function which was highly significant ($\chi^2 = 35.14, df = 7, p < .00001$). Two of the four neuropsychological variables significantly differentiated the groups: the total score on the CVLT ($p < .00001$) and the total score on the SOT ($p < .026$). The overall accuracy of this discrimination was 74% (sensitivity = .71, specificity = .81).

3. **Questionables versus converters**: The discriminant function comparing the questionables and the converters yielded one function which was also highly significant ($\chi^2 = 37.07, df = 7, p < .00001$). Three of the four neuropsychological variables significantly differentiated the groups: time to completion on Part B of the TMT ($p < .00001$), immediate recall of the figures from the WMS ($p < .001$), and the total score on the SOT ($p < .009$). The accuracy of this discrimination was 80% (sensitivity = .74, specificity = .83).

Two other discriminant functions were conducted as secondary analyses. They examined the ability of this set of four variables to differentiate the participants in Group 5 and Group 6 from those in Group 1. The first discriminant function demonstrated that there was a significant difference between the participants in Group 1 (the controls who remained controls) and Group 5 (the controls who became questionable; $\chi^2 = 22.3, df = 7, p < .002$; accuracy = 77%; sensitivity = .60, specificity = .84). There was, however, no significant difference between the participants in Group 1 and Group 6 (the questionable subjects who improved and were coded as controls).

**Contribution of APOE Status to Prediction of Group Membership After 3 Years of Follow-Up**

Discriminant function analyses were then used to examine the contribution of APOE genotype, alone or in combination with the neuropsychological variables, to the discrimination of the controls, questionables and converters. APOE status was coded as E4 present or absent. The percentage of individuals with an E4 allele was lower in the controls than in the questionables or the converters (22%, 30%, and 39%, respectively). However, when APOE status alone was used to differentiate these three groups, the discriminant function was not statistically significant. APOE status was then added to each of the discriminant function analyses in which the neuropsychological data had been examined. Addition of APOE status did not significantly improve the discrimination of the three groups above that of the neuropsychological data alone.

In addition, the distribution of APOE genotypes within the study sample was examined to determine whether it differed from the population at large. Published data from a large community ($N = 1209$) in the Boston area were used for purposes of comparison. Like the participants in the present study, this community sample (Framingham, Massachusetts) is predominantly White and of European ancestry (Wilson et al., 1994). The APOE genotypes for these two groups are presented in Table 2. A chi-square comparing the two distributions was not statistically significant, indicating that the relative frequency of APOE genotypes in the two populations are similar.

**DISCUSSION**

These data suggest that neuropsychological measures of memory and executive function, obtained at baseline, are useful in discriminating individuals destined to convert to AD from controls, and from individuals with mild memory impairments at baseline who do not progress sufficiently in 3 years to warrant a clinical diagnosis of AD. Four specific measures contributed to this discrimination: (1) two tests of memory—the total learning score on the California Verbal Learning Test, immediate recall of the figures from the Wechsler Memory Scale; and (2) two tests of executive function—time to completion on Part B of the Trail Making Test and the total score on the Self-Ordering Test.

The greatest accuracy was found when these measures were used to differentiate the controls from the converters (i.e., 89%). The comparison of the controls and questionables, and the comparison of the questionables and converters is lower, though substantial (74% and 80%, respectively). The lower accuracy of these latter comparisons is perhaps not surprising since, based on previous experience, it is anticipated that over subsequent years of follow-up more questionable individuals will progress to the point where they meet criteria for AD. Thus, some of the participants who are currently categorized as questionable are in the prodromal phase of AD, making differentiation from the individuals who have already converted to AD particularly challenging. It will be important to follow the individuals who remained questionable after 3 years of evaluation to determine the earliest point at which the neuropsychological measures can be used to discriminate individuals who are destined to later develop AD. These findings are therefore of potential clinical importance.

### Table 2. APOE genotype frequencies in the study sample and the participants in the Framingham Study

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Study sample</th>
<th>Framingham Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>e2–e2</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>e2–e3</td>
<td>21</td>
<td>13.1</td>
</tr>
<tr>
<td>e2–e4</td>
<td>3</td>
<td>1.9</td>
</tr>
<tr>
<td>e3–e3</td>
<td>93</td>
<td>58.1</td>
</tr>
<tr>
<td>e3–e4</td>
<td>40</td>
<td>25.0</td>
</tr>
<tr>
<td>e4–e4</td>
<td>3</td>
<td>1.9</td>
</tr>
</tbody>
</table>
These data reaffirm the well established finding that memory impairments, specifically those related to the ability to learn and retain new information, characterize the very earliest stage of AD (e.g., Moss et al., 1986; Newmann et al., 1994; Welsh et al., 1992). It is, however, surprising that the most discriminating measure did not pertain to delayed recall, as tests of delayed recall have consistently been the best discriminator of controls from patients with mild AD. Analyses conducted to examine this issue further (not shown) indicated that tests of delayed recall (such as free recall on the Selective Reminding Test or delayed recall on the CVLT) were second in discriminating power to measures that reflect consistency of learning (such as the Total score on the CVLT). Interestingly, the retention score on the CVLT also did not differentiate the groups. Thus, problems with delayed recall are clearly present at this early stage of disease, but learning consistency (as reflected by the accumulated number of words learned over the five trials of the CVLT) appears to be a better discriminator of those who will meet criteria for probable AD within the following few years.

These findings also confirm the importance of executive function abnormalities in the early stage of AD (e.g., Baddeley et al., 1986; Grady et al., 1988; Lafleche & Albert, 1995; Sahakian et al., 1990). Moreover, they demonstrate that tests requiring set shifting and sequencing are particularly impaired during the prodromal phase of disease. Though the Mental Control subtest of the WMS was not included here, it seems likely that its sensitivity in the model developed by Tierney et al. (1996) is related to the fact that one aspect of this subtest involves the reversal of a well-known sequence.

These findings are also informative on a theoretical level. The fact that tests of memory function are more prominent in the discrimination of the controls from the other two groups, whereas the tests of executive function were more prominent in the discrimination of the questionables from the other two groups, suggests that executive function deficits follow those of memory. This implies that the development of executive function difficulty in an individual who already has memory problems is likely to be a harbinger of AD.

These neuropsychological results parallel recent neuropathological and neuroimaging studies suggesting that brain regions important for memory and executive function are altered in prodromal AD. Neuropathological studies indicate that the striking memory deficits of individuals in the early stage of AD derive from the accumulation of pathology in memory-related neural systems in the brain. It has been known for some time that medial temporal lobe areas show the greatest damage in end-stage AD (Ball, 1977; Hyman et al, 1984) and medial temporal lobe structures are essential for normal memory (Beeson-Held et al., 1999; Zola et al., 2000). More recently, it has become apparent that the initial neuronal lesions of AD (e.g., the neurofibrillary tangles and neuritic plaques) develop in the entorhinal cortex, a portion of the anterior parahippocampal gyrus which receives projections from widespread limbic and association areas and gives rise to the perforant pathway, the major cortical excitatory input to the hippocampus itself (Gomez-Isla et al., 1996). In fact, some layers of the entorhinal cortex undergo 40 to 60% neuronal depopulation even in the earliest phase of AD (Gomez-Isla et al., 1996).

Recent neuroimaging data have provided increasing evidence for the involvement of the anterior cingulate gyrus in the early stage of AD. Studies using either SPECT or MRI have demonstrated alterations in the anterior cingulate among individuals in the prodromal phase of AD (Johnson et al., 1998; Killiany et al., 2000). The anterior cingulate gyrus may be vulnerable in early AD because it is reciprocally and strongly connected with memory-related structures, including the entorhinal cortex (Van Hoesen, 1993). In addition, however, it is strongly and reciprocally connected with the prefrontal cortex. It has been hypothesized that the anterior cingulate plays a major role in executive function abilities, primarily through these reciprocal connections with the prefrontal cortex (Arikuni et al., 1994).

This brain region is known to develop severe neuronal loss in AD (Vogt et al., 1991), but the stage at which this occurs is not yet known. The present findings, together with the neuroimaging data, mentioned above, suggest that neuronal loss in the anterior cingulate may begin early in the disease and may, in part, be responsible for the executive function deficits seen in the early stage of AD.

Taken together, these findings suggest that tests of memory and executive function, or measurements of brain regions related to these abilities, are implicated in early AD, and may be useful in discriminating individuals with AD several years before they meet clinical criteria for dementia.

The negative findings reported here regarding APOE status as a predictor of conversion to AD are consistent with several recent studies, including a large multicenter study that examined the use of APOE genotype as a diagnostic test for AD (Mayeux et al., 1998). In this multicenter study, the sensitivity and specificity of APOE alone in discriminating autopsy proven cases of AD from those of cases with other forms of dementia (e.g., Pick’s disease, cerebrovascular disease, etc) was 65% and 68%, respectively. In the present study, the accuracy of discriminating controls from converters was not significant.

The relationship of these findings to ongoing studies of MCI is also of interest. Individuals with MCI are generally selected in such a way that, in addition to evidence of progressive memory difficulty in daily life, their performance on selected tests of memory is 1.5 standard deviations lower than the norm for their age group (Petersen, 2000). Although the categorization of the participants in the present study was not based on findings from the neuropsychological battery, the test scores indicate that the converters differed by 1.5 standard deviations from the controls at baseline on three tests, one of which was in the area of memory. It therefore appears likely that the converters in the present study are comparable to participants who meet criteria for MCI. Since the other two tests that most differed between converters and controls pertained to executive function, the
present findings suggest that it would be particularly interesting to assess executive function ability in individuals with MCI.

Our data also suggest, however, that the questionable in the present study (i.e., the individuals who were questionable at both baseline and follow-up) have lower levels of cognitive impairment than is typical of individuals with MCI, as they did not differ by 1.5 standard deviations from the controls on any of the 20 neuropsychological test scores in the battery. This is consistent with the finding that 55% of the questionable participants in the present study demonstrated progressive functional difficulty on follow-up, but that 29% did not (Daly et al., 2000). It will be important to follow these questionable individuals to determine how many remain unchanged over time and how many ultimately meet criteria for probable AD.

Although tests of memory and executive function appear to be most impaired during the prodromal phase of AD, the present findings also emphasize the challenge in attempting to identify the boundary between the earliest stage of AD and normal aging. It is possible that an examination of the strategies participants use in performing these tasks will be even more useful at identifying individuals in the prodromal phase of AD than the overall test scores. Alternatively, combining neuropsychological measures with assessments from other domains, such as neuroimaging, may be more effective at predicting the trajectories of change that will be experienced by older individuals with memory problems. Such evaluations are under way.

It should also be emphasized that the composition of the study sample was constructed to have a specific distribution of persons with memory problems and controls. Thus, the sample is not representative of the elderly population at large. To determine the broad applicability of these findings, it will be necessary to evaluate and follow a sample of individuals selected to represent the general population at large.

ACKNOWLEDGMENTS

This work was supported by Grant #P01-AG04953 from the National Institute on Aging to the Massachusetts General Hospital. The authors would like to thank Dr. Mary Hyde for assistance with computer programming and data analysis.

REFERENCES


Johnson, K., Jones, K., Holman, B.L., Becker, J.A., Spiers, P.,


