Cognitive Assessment in Alzheimer’s Disease: Cross-Sectional and Longitudinal Perspectives

Alfred W. Kaszniak, Robert S. Wilson, Jacob H. Fox and Glenn T. Stebbins

ABSTRACT: This paper describes select results of a longitudinal study of 62 mild to moderate Alzheimer’s disease (AD) patients, in comparison to 60 age-matched healthy controls. Initial neurologic, radiologic, psychiatric, laboratory and cognitive examinations, required two full days, followed by one-day examinations at annual intervals. Of the total original sample, 31 AD patients and 39 controls could actually be followed for three annual examinations. Cognitive examination data confirmed cross-sectional (group discriminative) validity of memory and language measures, and showed the expected longitudinal deterioration in the AD sample, with controls maintaining consistent performance over the three years. However, those measures showing largest group differences at initial examination were not the best for tracking patient deterioration over time. Implications of these results for the selection of cognitive assessment measures are discussed.


Within recent years, considerable attention has focused upon the validation of cognitive measures appropriate for the assessment of dementia, particularly Alzheimer’s disease (AD).1-3 Given the prominence of memory impairment and decreased verbal fluency in the clinical presentation of AD, it is not surprising that memory and language assessment have frequently been the focus of test validation research. Studies comparing mildly demented AD patients to healthy controls have found psychometric measures of memory (e.g., the Wechsler Memory Scale) and of verbal fluency (e.g., generating as many words as possible that start with a given letter or belong to a given category) to be highly discriminating.6-9 Thus, available cross-sectional data provides evidence for the discriminative validity of particular memory and verbal fluency measures, in applications where differentiation of mild AD patients from healthy older persons is desired. However, there remains a relative lack of published data concerning the longitudinal validity of these same measures. Memory and verbal fluency assessment instruments have been shown to be predictive of differences, among AD patients, in overall rate of dementia progression over a one-year interval,10 but it is not yet clear whether these measures will validly track the expected longitudinal deterioration of cognition in AD. It cannot be assumed that measures able to differentiate mild AD patients from healthy older persons will also show longitudinal validity, since instruments sufficiently sensitive to mild dementia may show a truncation of variance ("baseline effects"), as dementia severity increases over time.

Questions concerning longitudinal validity are of considerable importance for neuropsychological research in dementia. An instrument with good initial discriminative validity, but variance truncation as dementia severity progresses, might not be the best choice for studies concerned with neurobiological correlates of cognitive deficit in AD. This is because such truncation in measure variance will limit the magnitude of any relationship which can be observed between a given neurobio-
logic (e.g., anatomic, physiologic, chemical) variable and the
cognitive measure in question. The purpose of the present
study is to examine both the discriminative (cross-sectional)
and longitudinal validity of select memory, verbal fluency, and
other language measures, in samples of AD patients and age-
matched healthy controls.

METHODS AND MATERIALS

Subjects

The initial sample consisted of 62 AD patients and 60 healthy
controls. The AD patients were referred to the study by cooper­
ating neurologists and internists, while controls were volun­
teers from the community who were paid for their project
participation. All subjects received comprehensive evaluations,
including physical and neurologic examination, a computerized
tomographic (CT) scan, electroencephalogram (EEG), psychi­
atric interview, interview with a close family member, and
extensive cognitive assessment. The AD patients met DSM­
III criteria for Primary Degenerative Dementia. A series of
laboratory procedures (including blood counts, red blood cell
indices, erythrocyte sedimentation rate, blood urea nitrogen,
vitamin B12 levels, folic acid levels, urinalysis, electrolytes,
sugar, calcium, phosphorus, liver function, T4, serum protein
electrophoresis, electrocardiogram, and chest and skull radiogra­
gms) were performed in order to rule out various infectious,
toxic or metabolic disorders. The Hachinski ischemic scale was
used to assist in eliminating from the sample any patients
suspect of multi-infarct dementia. The present sample of patients
meets recently established criteria for clinical diagnosis of
probable Alzheimer's disease. No patients were institutional­
zation at the time of entry into the sample, and all patients had
either mild or early moderate dementia severity, by current
criteria.

The initial examinations required two full days, followed by
one day examinations at annual intervals. The total duration of
the study was five years, with new subjects entering during
during years one through four. Eight patients and 14 controls entered in
eyear four, and therefore had only two annual examinations.
This reduced the potential sample which could be followed for
three years to 54 patients and 46 controls. Of these, 31 patients
and 39 controls were actually followed for at least three annual
examinations. Five controls (11 percent) were lost to follow-up,
and 14 controls refused. An additional two controls, although followed, were
not included in analyses, as neurological examination raised
the possibility of minor focal neurologic abnormality at re-
examination. Twenty-three patients (43 percent) were lost to
follow-up, with five deaths, four medical complications, five
moves out of state, and nine refusals.

Demographic data for subjects followed for three years, and
those lost to follow-up is shown in Table I. The 70 subjects who
were followed for three years did not significantly differ from
the original sample of 122 in age, sex, education, physical and
neurologic examination results, medical history, or initial
dementia severity. However, AD patients lost to follow-up had
a slightly longer dementia history, as estimated by a close
family member, than patients who remained in the sample for
three annual examinations. AD versus control subjects fol­
lowed for three years were not significantly different in age, sex
or education.

Cognitive Measures

All subjects were administered a large battery of memory,
language, perceptual and emotional assessment instruments,
the published cross-sectional results of which have been re­
viewed elsewhere.2 For purposes of the present report, a subset
of measures of memory, verbal fluency, and naming were
selected to be consistent with those reported to have discrimina­
tive (mild AD versus healthy control) validity in previously
published research.67 These include the Logical Memory and
Visual Reproduction subtests of the Wechsler Memory Scale
(WMS),2 and the Animal Naming, Visual Confrontation Nam­
ing, and Spoken Commands subtests of the Boston Diagnostic
Aphasia Examination (BDAE).17 Both the Wechsler Memory
Scale and the Boston Diagnostic Aphasia Examination subtests
have normative data available for adults through 80 and 85
years of age, respectively, facilitating their clinical utility.
The Logical Memory subtest of the WMS requires the subject
to immediately recall each of two brief paragraphs spoken by
the examiner, while the WMS Visual Reproduction Subtest
involves the presentation, for ten seconds each, of three differ­
cent non-representational geometric forms, which are then to be
immediately drawn from memory. The BDAE Animal Naming
subtest asks the subject to verbally generate as many names of
animals as possible within a 60 second time period. The BDAE
Visual Confrontation Naming subtest involves the examiner
pointing to each of several line drawings which must be named
by the subject. Finally, the BDAE Spoken Commands subtest
requires the subject to execute simple commands spoken by the
examiner. All subtests were individually administered to patients
and controls, in a quiet and private examination room, by
graduate research assistants experienced in maintaining the
cooperation of dementia patients.

Table 1. Demographic data on total study sample

<table>
<thead>
<tr>
<th>N</th>
<th>Followed Controls</th>
<th>AD Patients</th>
<th>Lost to Follow-up Controls</th>
<th>AD Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>31</td>
<td>7</td>
<td>39</td>
<td>7</td>
</tr>
<tr>
<td>Age at entry (years)</td>
<td>69.28</td>
<td>68.81</td>
<td>68.86</td>
<td>66.70</td>
</tr>
<tr>
<td>Education (years)</td>
<td>10.82</td>
<td>11.81</td>
<td>13.10</td>
<td>10.80</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>21</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>18</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Length of Illness (months)</td>
<td>33.14</td>
<td>44.18</td>
<td>33.14</td>
<td>44.18</td>
</tr>
</tbody>
</table>
RESULTS

In order to determine initial group differences on the memory, verbal fluency, and other language measures, as well as any possible differences between subjects in each group who were followed for three years versus those who were not, a series of twoway (group by stay-in/drop-out status) analyses of variance (ANOVA) were computed. Significant group main effects (p<.0001) were obtained on all dependent variables, confirming the discriminant (diagnostic) validity of the measures selected. No significant main effects of followed versus lost to follow-up, nor significant group by follow-up interaction effects were obtained. Given the absence of follow-up status (followed versus not followed) main effects and group by follow-up interactions, it can be concluded that the sample available for each of the three annual reexaminations is representative, in terms of initial memory and language functioning, of the entire sample originally enrolled into the study.

While the group main effects were highly significant for all dependent variables, these variables did differ in the degree to which they differentiated the groups (discriminative efficiency). Employing the statistic η2 (the square of which indicates the proportion of dependent measure variance accounted for by the group difference), subtests demonstrated the following discriminative efficiency rank ordering (from greatest to least): (1) WMS Logical Memory (F = 101.74, η2 = .79), (2) WMS Visual Reproduction (F = 100.02, η2 = .69), (3) BDAE Animal Naming (F = 82.39, η2 = .65), (4) BDAE Visual Confrontation Naming (F = 29.64, η2 = .46), (5) BDAE Spoken Commands (F = 21.04, η2 = .41).

Longitudinal data analyses, while also confirming the longitudinal validity of each of the subtests, indicate a quite different ordering, in terms of their ability to track progression of dementia severity. Analyses of longitudinal data are based upon a smaller group of 22 AD patients and 39 controls, since nine of the AD patients were too demented at either the second or third examination for meaningful collection of all subtest data. Group means and standard deviations for the AD and control groups, across the three annual evaluations, are shown in Table 2.

Multivariate analyses of variance (MANOVA) were performed for each of the subtests, using a group (AD versus control) by repeated measures (each of the three annual examinations) design. As would be expected from the cross-sectional analyses described above, group main effects were highly significant (p<.0001) for all subtests. As can be seen in Table two, control subjects showed consistent subtest performance across the three annual examinations, while AD patients showed progressive deterioration (confirmed by significant (p< .05) interaction term multivariate F ratios).

The magnitude of the interaction effect (group by annual evaluation) multivariate F ratio can be employed as an index of a subtest’s ability to track longitudinal deterioration in the AD patients (the greater the change, over repeated examinations, in the AD group, relative to controls, the larger the interaction effect). Using this index, the subtests demonstrated the following rank ordering in ability to track AD longitudinal deterioration: (1) BDAE Spoken Commands (interaction multivariate F = 18.75), (2) BDAE Visual Confrontation Naming (interaction multivariate F = 13.91), (3) WMS Visual Reproduction (interaction multivariate F = 10.11), (4) BDAE Animal Naming (interaction multivariate F = 7.36), (5) BDAE Logical Memory (interaction multivariate F = 4.31). This ordering, with the minor exception of the interchanged positions of WMS Visual Repro-

| Table 2: Means (and Standard Deviations) for Cognitive Tests Across Annual Examinations |
|---------------------------------|-----------------|-----------------|-----------------|
|                                 | Control         | AD              | Control         |
| WMS Logical Memory              |                 |                 |                 |
| 1                               | 8.33 (2.51)     | 2.82 (2.67)     | 7.97 (3.36)     |
| 2                               | 9.10 (2.70)     | 2.54 (1.99)     | 9.82 (3.56)     |
| 3                               | 8.66 (2.78)     | 1.23 (1.57)     | 8.10 (3.48)     |
| WMS Visual Reproduction         |                 |                 |                 |
| AD                              | 2.04 (1.84)     | 1.50 (1.84)     | 0.55 (1.14)     |
| BDAE Animal Naming              |                 |                 |                 |
| AD                              | 10.61 (4.31)    | 8.26 (4.32)     | 5.52 (3.65)     |
| BDAE Visual Naming              |                 |                 |                 |
| AD                              | 10.47 (0.08)    | 10.41 (0.18)    | 10.42 (0.16)    |
| BDAE Spoken Commands            |                 |                 |                 |
| AD                              | 9.76 (0.78)     | 9.21 (0.98)     | 8.03 (2.01)     |
|                                 | 19.67 (0.53)    | 19.59 (0.72)    | 19.46 (0.75)    |
|                                 | 18.33 (1.94)    | 17.44 (2.41)    | 15.67 (3.98)    |

WMS = Wechsler Memory Scale; BDAE = Boston Diagnostic Aphasia Examination; AD = Alzheimer's disease
duction and BDAE Animal Naming, is directly opposite that obtained for group discriminative validity. As revealed by inspection of subtest standard deviations, across repeat examinations, in Table 2, this appears due to a systematic reduction in variability over time, for those subtests most sensitive to initial group difference.

**DISCUSSION**

As expected, based upon results of previous research, the presently employed measures of verbal (WMS Logical Memory) and nonverbal (WMS Visual Reproduction) recent memory, as well as measures of verbal fluency (BDAE Animal Naming), object naming (BDAE Visual Confrontation Naming), and auditory comprehension (BDAE Spoken Commands) all show validity in documenting differences between mild to moderate AD and age-matched healthy controls. Further, consistent with clinical impressions and previous research, memory and verbal fluency measures provide more efficient group discrimination than object naming or auditory comprehension measures, and therefore appear to be good choices for application in initial diagnostic assessment.

All of the presently employed subtests also demonstrate validity in documenting the progressive deterioration of AD (as confirmed by significant group by repeated examination multivariate interaction effects). However, their relative efficiency in tracking AD patient deterioration is generally opposite their ordering of efficiency in initial group discrimination. This reversal of subtest efficiency, when comparing discriminative and longitudinal validity, appears attributable to the effects of variance truncation. That is, those measures most sensitive to initial differences between AD patients and controls (recent memory and verbal fluency measures) are also those which show “basement effects” (reduction in range of variability) over repeated yearly examinations, for the AD patients. It should be pointed out that these basement effects are even more marked than the present data might suggest, since those AD patients too demented at follow-up examinations to comprehend task instructions were eliminated from the longitudinal sample. As we have previously demonstrated, AD patients who cannot be validly tested with a particular task, at follow-up evaluation, show more rapid dementia progression (as indicated by a simple mental status examination), than those patients who continue to be testable.

Thus, while immediate verbal and nonverbal memory and verbal fluency measures may be good choices for initial diagnostic assessment, other measures, such as those of object naming and auditory comprehension, may provide better approaches to tracking the cognitive deterioration of AD, across a range of dementia severity (despite the fact that the latter measures are less efficient at initial group discrimination). Overall, these results argue for an approach to cognitive assessment in dementia which explicitly recognizes the deteriorative nature of dementing illness, and provides different instruments for the assessment of mild, early impairment, as well as for those more pervasive and severe deficits seen with dementia progression.

**ACKNOWLEDGEMENTS**

The research reported in this manuscript was supported by NIH grant AG00905 to Rush Medical College. Preparation of the manuscript was supported by NIMH grant MH40827 to the University of Arizona.

**REFERENCES**