Beyond memory impairment
Cognitive changes in Alzheimer’s disease

Edward Helmes,*, Truls Østbye

Department of Psychology, University of Western Ontario, London, ON, Canada
Department of Epidemiology and Biostatistics, University of Western Ontario, London, ON, Canada

Accepted 9 November 2000

Abstract

In addition to memory impairment, deficits in other cognitive processes are common in the advanced stages of Alzheimer’s disease (AD). The diagnosis of AD does not consider the relative prevalence of deficits in cognitive areas other than memory. We report on the prevalence of aphasia, apraxia, and other cognitive changes in individuals from a large representative sample of elderly Canadians. The proportion of these symptoms and the relevant neuropsychological test performance were compared in a group of 749 people over 65 years in age with AD and a control group of 563 people without cognitive impairment. Agnosia was less common in both groups than were deficits in complex visuomotor tasks, abstract thinking, aphasia, and constructional defects. The occurrence of all symptoms increased, and levels of performance on relevant neuropsychological tests decreased, with severity of Alzheimer disease. The tests did not, however, distinguish between possible and probable AD. Both these diagnostic groups showed similar levels of performance, which suggests that this distinction is not clinically meaningful. © 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

Keywords: Dementia; Alzheimer’s disease; Aphasia; Apraxia; Construction defect; Agnosia

* An earlier version of this paper was presented at the 103rd Annual Convention of the American Psychological Association, 14 August 1995, New York, NY.

* Corresponding author. School of Psychology, Edith Cowan University, 100 Joondalup Drive, Joondalup, WA 6027, Australia. Tel.: 61-8-9400-5543; fax: 61-8-9400-5834.

E-mail address: e.helmes@cowan.edu.au (E. Helmes).

Present address: Department of Community and Family Medicine, Duke University, Durham, NC 27710, USA.
1. Introduction

A basic feature of all forms of dementia is the gradual loss of both general and specific cognitive functions (Burns, Jacoby, & Levy, 1991; Zec, 1993). Disorders of language are common in the more advanced stages of many forms of dementia (Miller, 1989a; Thompson, 1987), as are apraxia (Rapcsak, Croswell, & Rubens, 1989), constructional defects (Edwards, Deuel, Baum, & Morris, 1991), and, less commonly, agnosia, the lack of recognition with adequate perception (Lincoln, 1995). In dementia of the Alzheimer type, impairments in these specific domains are part of common diagnostic criteria (American Psychiatric Association, 1987). The criteria do not, however, consider the relative frequency of occurrence of these disorders in dementing older people as the disorder progresses. If some disorders co-occur with memory problems, if they are as prevalent, but are present in different individuals, or occur very infrequently, then their inclusion in the diagnostic criteria does not add to the diagnostic power of the criteria. This report provides further insights into these matters, as well as the occurrence of the various disorders in a comparison group of cognitively intact older people.

Language disorders in people with Alzheimer’s disease (AD) are extensive, and include impaired naming, word fluency, comprehension, and word associations (Hart, Smith, & Swash, 1988; Miller, 1989a, 1989b; Murdoch, Chenery, Wilks, & Boyle, 1987; Nicholas, Barth, Obler, Au, & Albert, 1997; Price et al., 1993). The deficits in language arise from the widespread neuronal loss in AD, as the specific language centers become affected. Thus, the term aphasia (Greek: “no speaking”) may not be fully appropriate for the range of language difficulties in AD, although the term is often used more broadly to refer also to impairments in speaking (Bayles & Kaszniak, 1987).

Apraxia (Greek: “no action”) is a term applied to disorders of the organization of complex sequences of movements in space (Lezak, 1995). A distinction is often made between constructional apraxia, which involves drawing tasks and the assembly of two- and three-dimensional objects, and ideomotor apraxia, which is a failure to imitate or perform coordinated movements upon command in the absence of defects in motor control (Lucchelli, Lopez, Faglioni, & Boller, 1993). Both types of deficit have been noted to be impaired in AD (Edwards, Baum, & Deuel, 1991; Kirk & Kertesz, 1991; Rapcsak et al., 1989; Travniczek-Marterer, Danielczyk, Simanyi, & Fischer, 1993).

Agnosia (Greek: “no knowledge”), the absence of meaning in the presence of normal perception (Lezak, 1995), has not been the focus of much specific research in the context of AD, perhaps because as a complex concept, it has been viewed as part of the general deterioration in abstract thinking and judgement that are common in AD (Zec, 1993). Some specific agnostic disorders, such as anosognosia, the denial of illness, have recently been noted to be fairly common in AD (Starkstein et al., 1995).

Abstract thinking (concept formation) and judgement (reasoning) are also abilities that deteriorate with AD. Some authors regard these as distinctly different faculties (e.g., Lezak, 1995), while others treat them as closely related, in part because they are both highly correlated with age and education (Zec, 1993). Impaired reasoning ability is among the more common changes observed in AD, whether finer distinctions are made or not.
Few studies, if any, have compared the relative frequency of these cognitive disorders in older adults with or without dementia. Here, we describe the frequency of occurrence of these conditions in the Canadian Study of Health and Aging, a large, national representative sample of elderly Canadians with probable or possible AD.

2. Method

2.1. Participants

This study was based on a stratified random sample of 10,263 participants from the 10 Canadian provinces (Canadian Study of Health and Aging Working Group, 1994). The working group comprised multidisciplinary teams of clinical and methodological investigators from across Canada. Medicare systems file and electoral rolls were used to select the sample for the 18 study centers. The five major geographic regions were sampled equally, with the age groups over 75 being oversampled because of the higher prevalence of dementia in these groups. Individuals living in the community, as well as institutions, were included. Details of the study design and the sampling methodology are given by the Canadian Study of Health and Aging Working Group (1994).

All community participants received an initial screening interview that included the Modified Mini-Mental State Examination (3MS; Teng & Chui, 1987). Individuals in the community scoring below 78/100 and above 50/100 on the 3MS and a subsample of 494 who scored 78 or above, as well as all 1255 institutionalized participants, were invited to participate in a more detailed clinical assessment (N = 2914). This assessment included a clinical neurological examination, nurse’s evaluation, blood tests, and an extensive neuropsychological assessment. Diagnostic criteria from the American Psychiatric Association (1987) were used, and the final diagnosis was based upon a consensus conference of the neuropsychologist, nurse, and physician using all the clinical findings together with a computer algorithm to make a final diagnosis (Graham et al., 1996). This final diagnostic conference also used the NINCDS-ADRDA criteria for severity of dementia, and in the case of AD, whether this diagnosis was possible or probable. A total of 861 people undergoing the clinical assessment were determined to have some form of cognitive impairment but who did not meet the criteria of dementia, while a further 208 were diagnosed with vascular dementia and 175 with other forms of specific or unclassified dementia (Graham et al., 1996). The lack of imaging studies may have led to lower rates of diagnosis of other forms of dementia than if such tests had been performed routinely. Kappa (Cohen, 1960) was 0.81 (84.1% agreement) for the diagnosis of AD/no AD in a subsample of 210 participants who had an independent chart review (Graham et al., 1996). This reliability analysis did not include evaluations of the severity of the dementia or for the distinction between possible and probable AD. MacKnight, Graham, and Rockwood (1999) discuss the factors underlying some of the differences in diagnoses that were observed between physicians and neuropsychologists. The diagnoses used in the current analyses are based upon the final consensus of the case conferences.
Using the criteria of McKhann et al. (1984), a total of 749 people (193 men and 556 women) with a mean age of 84.3 years (S.D. = 6.83) were determined to have probable (n = 448) or possible AD (n = 301).

A comparison group, found to be cognitively normal after the clinical consensus conference, consisted of 563 people (181 males and 382 females), 398 from the community sample and 165 from the institutional sample. The mean age in this group was 78.9 years (S.D. = 6.62).

2.2. Procedure

After the clinical assessment, those who scored above 50 on a second 3MS administered during the nurse’s examination underwent a neuropsychological assessment (n = 1843, Tuokko, Kristjansson, & Miller, 1995). The battery of tests used reflected the constructs included in the diagnostic criteria. The final selection of tests was based upon a feasibility study and comprised short forms of three Wechsler Adult Intelligence Scale (WAIS)-R (Wechsler, 1981) subtests (Block Design, Comprehension, and Similarities; Satz & Mogel, 1962), the WAIS-R Digit Symbol subtest, and other measures of memory and higher cognitive functions (Tuokko et al., 1995) (see below).

In accordance with the diagnostic criteria used, the study distinguished ideational apraxia (performing integrated serial acts or tool use) or ideomotor apraxia (responding to verbal commands) from constructional apraxia (Zec, 1993) or constructional defect (Lezak, 1995). The terminology used here follows Lezak (1995): constructional defect refers to disorders of drawing and object assembly; apraxia refers to other forms of complex movement disorder. Both constructional defect and apraxia were diagnosed separately at the diagnosis meeting, along with the other diagnostic entities of aphasia and agnosia. Additional consensus judgements were made about defects in judgement and abstract thinking.

Aphasia was assessed by an 11-item Token test (Benton & Hamsher, 1989), Animal Naming (Rosen, 1980), naming 12 objects (Visual Identification) from the cued recall paradigm of Buschke (1984), and Verbal Fluency (Controlled Oral Word Association Test; Spreen & Benton, 1977). Verbal fluency is not normally considered a measure of aphasia, but is included here with other measures of language as a result of the decisions made in the formation of the battery (Tuokko et al., 1995). The naming tests were open-ended, with achieved maximum scores of 62 for Verbal Fluency and 27 for Animal Fluency. Visuomotor skills comprised assembly, coding, and drawing tests. The measures used were two tests from the WAIS (Block Design and Digit Symbol) and the sum of three scores (maximum of 10) derived from a figure of interlocking pentagons used in the 3MS screening test. These tests involve primarily construction and object manipulation tasks. The battery used in the study did not have any direct measures of ideational or ideomotor apraxia. Agnosia was measured by the color naming section of the Token test (maximum score of 5) and the number of objects correctly identified by category from the object naming section of the Buschke paradigm (maximum score of 12). Finally, the Similarities and Comprehension subtests of the WAIS-R assessed abstract thinking and judgement, respectively. This test battery has been shown to be able to discriminate among various forms of dementia (Steenhuis & Østbye, 1995).

All calculations for the present study were performed using SYSTAT (SPSS, 1998).
3. Results

3.1. Diagnosis

Not all of the 749 patients with probable or possible AD also had the individual cognitive disorders of aphasia, constructional defect, impaired abstract thinking, agnosia, or apraxia. In other words, the number of cases with a diagnosis for a given cognitive disorder was always less than the number with the diagnosis of the severity of AD. Excluding eight cases with missing data, 58.0% (430/741) of those with diagnosed AD were determined by the consensus conference to have aphasia, 51.7% (383/741) apraxia, 34.1% (253/741) agnosia, 87.2% (646/741) impaired abstract thinking, 77.3% (573/741) impaired judgment, and 65.0% (482/741) problems with construction skills. The last column of Table 1 indicates the number of cases with completed diagnoses of the specific disorders from the consensus conference and the total number who were evaluated at the consensus conference. The theoretical maximum for the latter figure is 1304. The denominator for the above figures for each disorder is therefore smaller than the denominators used in Table 1.

Individuals who did not complete the full clinical assessment did not receive diagnoses of the individual cognitive disorders. Their data are treated as missing in the analyses of diagnoses of cognitive dysfunction and are omitted from the analyses of the neuropsychological tests (n = 294). An examination of possible bias due to these missing data used a type I error rate of 0.01 to avoid spurious findings due to the large number of comparisons made in this section. Comparisons of gender ratios between those diagnosed with a disorder and those who were not were not statistically significant for any of the six cognitive disorders. While those missing a diagnosis did not differ in age from those with a diagnosis of agnosia, those lacking diagnoses for aphasia, apraxia, constructional and judgement defect, and impaired abstract thinking were younger than those with diagnoses of aphasia: 80.3 vs. 82.6 years, t(1310) = -5.07, P < .001; apraxia: 80.8 vs. 82.6 years, t(1310) = -4.20, P < .001; construction defect: 81.1 vs. 82.4 years, t(1310) = -3.15, P < .001; judgement: 79.7 vs. 82.9 years, t(1310) = 7.45; abstract thinking: 80.0 vs. 82.7 years, t(1310) = -5.93, P < .001).

<table>
<thead>
<tr>
<th>Diagnosed disorder</th>
<th>Level of impairment</th>
<th>None (N=563)</th>
<th>Mild (N=154)</th>
<th>Moderate (N=279)</th>
<th>Severe (N=308)</th>
<th>Number of AD cases diagnosed/total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agnosia</td>
<td>.000</td>
<td>.07</td>
<td>.30</td>
<td>.87</td>
<td>253/834</td>
<td></td>
</tr>
<tr>
<td>Apraxia</td>
<td>.003</td>
<td>.30</td>
<td>.52</td>
<td>.90</td>
<td>384/884</td>
<td></td>
</tr>
<tr>
<td>Aphasia</td>
<td>.011</td>
<td>.48</td>
<td>.51</td>
<td>.82</td>
<td>433/952</td>
<td></td>
</tr>
<tr>
<td>Judgement</td>
<td>.004</td>
<td>.52</td>
<td>.88</td>
<td>.996</td>
<td>574/942</td>
<td></td>
</tr>
<tr>
<td>Constructional defect</td>
<td>.007</td>
<td>.62</td>
<td>.77</td>
<td>.97</td>
<td>484/861</td>
<td></td>
</tr>
<tr>
<td>Abstract thinking</td>
<td>.011</td>
<td>.80</td>
<td>.97</td>
<td>1.00</td>
<td>649/968</td>
<td></td>
</tr>
</tbody>
</table>
3.2. Demographic factors

The initial rows of Tables 2 and 3 report age, years of education and 3MS scores for the control, severity level, and probable vs. possible AD groups. The groups with mild and moderate AD showed fewer years of education \( F(2,1218 \text{ df}) = 49.8, P < .001 \), and were older than the control group \( F(2,1271 \text{ df}) = 43.1, P < .001 \). Post-hoc Tukey HSD tests showed that all three groups differed on 3MS scores \( F(2,746 \text{ df}) = 926.3, P < .001 \). The same pattern was evident for the possible vs. probable AD grouping: the two groups with AD were older \( F(2,1309 \text{ df}) = 103.6, P < .001 \) and had fewer years of education \( F(2,1168 \text{ df}) = 102.6, P < .001 \).

Table 2

Scores on demographic measures and neuropsychological tests with increasing severity of AD

<table>
<thead>
<tr>
<th>Measure or test</th>
<th>None</th>
<th>Mild</th>
<th>Effect</th>
<th>Moderate</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
<td>S.D.</td>
<td>n</td>
<td>Mean</td>
</tr>
<tr>
<td>Age</td>
<td>563</td>
<td>78.9</td>
<td>6.62</td>
<td>267</td>
<td>82.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td>556</td>
<td>10.6</td>
<td>3.74</td>
<td>257</td>
<td>6.96</td>
</tr>
<tr>
<td>3MS</td>
<td>398</td>
<td>89.7</td>
<td>5.61</td>
<td>179</td>
<td>65.4</td>
</tr>
<tr>
<td>Abstract reasoning and judgement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarities</td>
<td>517</td>
<td>7.8</td>
<td>3.85</td>
<td>114</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>516</td>
<td>9.7</td>
<td>3.06</td>
<td>112</td>
<td>6.1</td>
</tr>
<tr>
<td>Aphasia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Token test</td>
<td>501</td>
<td>38.6</td>
<td>6.04</td>
<td>109</td>
<td>30.5</td>
</tr>
<tr>
<td>Verbal Fluency</td>
<td>512</td>
<td>28.4</td>
<td>11.57</td>
<td>106</td>
<td>16.9</td>
</tr>
<tr>
<td>Animal Fluency</td>
<td>508</td>
<td>14.6</td>
<td>3.96</td>
<td>109</td>
<td>8.9</td>
</tr>
<tr>
<td>Naming</td>
<td>506</td>
<td>11.8</td>
<td>0.47</td>
<td>112</td>
<td>11.2</td>
</tr>
<tr>
<td>Agnosia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objects</td>
<td>506</td>
<td>11.9</td>
<td>0.43</td>
<td>112</td>
<td>11.5</td>
</tr>
<tr>
<td>Colors</td>
<td>501</td>
<td>5.00</td>
<td>0.17</td>
<td>110</td>
<td>4.96</td>
</tr>
<tr>
<td>Visuomotor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block Design</td>
<td>499</td>
<td>10.2</td>
<td>4.57</td>
<td>107</td>
<td>4.3</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>476</td>
<td>28.7</td>
<td>11.49</td>
<td>93</td>
<td>13.8</td>
</tr>
<tr>
<td>Drawing</td>
<td>537</td>
<td>9.1</td>
<td>1.58</td>
<td>139</td>
<td>6.4</td>
</tr>
</tbody>
</table>

\( ^{a} \) Means sharing the same superscript are not significantly different by Tukey’s honestly significant difference procedure. The second row for each variable gives the standard scores based upon the mean and S.D. for the control group. The mean for the comparison group was 0.0, with an S.D. of 1.0 within rounding error.
All three groups differed on the 3MS by post-hoc Tukey HSD testing \( F(2,650) = 758.0, P < .001 \).

### 3.3. Severity level

The hypothesis that the prevalence of all types of cognitive dysfunction would increase with the severity of AD was tested by use of the consensus rating at each study center of the severity of the AD (mild, moderate, or severe) using DSM-III-R criteria (American Psychiatric Association, 1987). Eight cases did not receive severity ratings, leaving a total sample of 1304 for this analysis (370 males and 934 females).

At the level of clinical diagnosis, not all individuals showed the deficits, even at the most severe level of impairment. Table 1 reports the relative frequency of the diagnoses of aphasia, apraxia, agnosia, and constructional defects, excluding the 294 individuals from the comparison group who did not undergo the neuropsychological assessment and consensus diagnosis. The linear increase in frequency of diagnosed cognitive disorder was statistically significant for all diagnosed disorders: abstract thinking \( \chi^2(1, n=968)=120.4, P<.001 \); judgement \( \chi^2(1, n=942)=623.5, P<.001 \), agnosia \( \chi^2(1, n=834)=398.1, P<.001 \); aphasia \( \chi^2(1, n=952)=10.9, P<.001 \), apraxia \( \chi^2(1, n=884)=460.3, P<.001 \), and constructional defects \( \chi^2(1, n=861)=208.8, P<.001 \), using a statistic for trends in contingency tables due to Maxwell, described by Marascuilo and McSweeney (1977; pp. 198–202). In all six

\[
\text{df} = 48.9, P < .001.\]

All three groups differed on the 3MS by post-hoc Tukey HSD testing \( F(2,650 \text{ df}) = 758.0, P < .001 \).

### Table 3
Comparison of controls with patients having possible and probable AD

<table>
<thead>
<tr>
<th>Test or measure</th>
<th>Control</th>
<th>Possible Alzheimer’s</th>
<th>Probable Alzheimer’s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
<td>Mean S.D. ( n )</td>
<td>Mean S.D. ( n )</td>
</tr>
<tr>
<td>Age</td>
<td>563 78.9</td>
<td>6.62 448 84.5a</td>
<td>6.80 0.85</td>
</tr>
<tr>
<td>Years of education</td>
<td>556 10.6</td>
<td>3.74 360 8.4a</td>
<td>3.79 0.59</td>
</tr>
<tr>
<td>3MS</td>
<td>398 89.7</td>
<td>5.61 154 54.5</td>
<td>16.97 6.27</td>
</tr>
<tr>
<td>Abstract Reasoning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and Judgement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarities</td>
<td>517 7.8</td>
<td>3.85 104 3.0a</td>
<td>2.73 1.25</td>
</tr>
<tr>
<td>Comprehension</td>
<td>516 9.6</td>
<td>3.06 105 5.8a</td>
<td>3.08 1.24</td>
</tr>
<tr>
<td>Aphasia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Token test</td>
<td>501 38.6</td>
<td>6.04 97 31.1a</td>
<td>8.68 1.24</td>
</tr>
<tr>
<td>Verbal Fluency</td>
<td>512 28.4</td>
<td>11.57 96 15.0a</td>
<td>9.18 1.16</td>
</tr>
<tr>
<td>Animal Fluency</td>
<td>508 14.6</td>
<td>3.96 100 7.8a</td>
<td>3.35 1.72</td>
</tr>
<tr>
<td>Naming</td>
<td>506 11.8</td>
<td>0.47 101 11.0a</td>
<td>1.67 1.70</td>
</tr>
<tr>
<td>Agnosia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objects</td>
<td>506 11.9</td>
<td>0.43 101 11.5a</td>
<td>0.98 0.93</td>
</tr>
<tr>
<td>Colors</td>
<td>501 4.98a</td>
<td>0.17 99 4.9a</td>
<td>0.22 0.47</td>
</tr>
<tr>
<td>Visuomotor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block Design</td>
<td>499 10.2</td>
<td>4.57 93 4.3a</td>
<td>4.28 1.29</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>476 28.7</td>
<td>11.49 73 11.9a</td>
<td>8.24 1.46</td>
</tr>
<tr>
<td>Drawing</td>
<td>537 9.1</td>
<td>1.58 235 4.3a</td>
<td>3.72 3.04</td>
</tr>
</tbody>
</table>

\( a \) Means sharing the same superscript are not significantly different by Tukey’s honestly significant difference procedure.
cases, the values of the residual \( \chi^2 \) were sufficiently large to suggest that the actual relationship was curvilinear. This finding is consistent with the common observation of a relatively slow decline in the early stages of AD that is followed by a period of more rapid deterioration.

Table 2 gives the means for the age, education, and mean MMSE scores and for the neuropsychological tests at the different levels of severity of AD, excluding the group with the most severe dementia because of the comparatively small number of cases in that group \((n = 4)\). (A copy of the table containing the results pertaining to this group can be obtained by contacting the first author). It should be noted that higher scores indicate better performance on all the tests used, and the overall pattern of results followed the expectation of lower performance with increasing severity of AD, with the exception of the Token test, which showed a modest increase in scores in the moderate AD group over the mild AD group. In addition, not all test scores showed parallel rates of decline, and some were better preserved than others at the level of moderate impairment. In order to demonstrate this, the scores for the two AD groups were converted to z scores by using the mean and S.D. of the control group. These values are also reported in Table 2, along with values of \( \eta^2 \) and simple effect sizes to indicate the relative amount of variance accounted for by the independent variable for that particular test.

An initial multivariate analysis of variance was performed to determine if there were significant differences across levels of severity. This test was significant \([\text{Pillai’s trace} = 0.517, F(22,1222) = 19.37, P < .001]\). A type I error rate of 0.01 was selected for the univariate tests because of the number of comparisons being made. The Tukey honestly significant difference procedure was used to maintain the type I error rate within each post-hoc analysis.

The abstract reasoning and judgement measures both showed the same pattern of decline. The overall effect for Similarities was significant, \(F(2,742) = 156.40, P < .001\). The post-hoc tests showed that the scores of the two AD groups did not differ from one another. Comprehension showed a similar pattern, \(F(2,740) = 129.67, P < .001\), with Tukey’s test showing the same pattern as for Similarities.

Language test scores showed roughly comparable declines, except for the verbal fluency measure, which did not decline as quickly as the other measures in mild AD. Token test scores showed two outliers by the criterion of more than four Studentized residual units, and these individuals were deleted from the analysis, although the interpretation of the results remained the same. The corrected test showed significant differences, \(F(2,712) = 100.85, P < .001\), with the control group significantly greater than both groups with dementia, which did not differ from one another by Tukey’s test. The naming test showed all three groups to differ among one another (Table 2) by Tukey’s test \([\text{overall } F(2,722) = 53.48, P < .001]\). Scores on this measure were quite skewed. Consequently, a nonparametric Kruskal–Wallis analysis was performed. This showed the same results \([\chi^2(2, n = 725) = 116.1, P < .001]\). The fluency measures also had overall statistically significant differences, but these were more pronounced with the animal fluency measures than with the Verbal Fluency scores, \(F(2,721) = 238.07, P < .001\) and \(F(2,724) = 106.24, P < .001\), respectively. The post-hoc Tukey tests also showed more differences among groups for the Animal Fluency measure than for the F-A-S measure. All three groups differed significantly from one another on Animal Fluency, while only the controls differed from the two AD groups for Verbal Fluency.
Both language tests that showed lower scores in the moderate AD group, Naming and Verbal Fluency, also were statistically significantly different from the mild AD group.

The measures of agnosia showed little absolute differences across the three groups. The large samples led to statistically significant results for Object Identification, $F(2,723)=39.22$, $P<.001$, but not for Color Identification, $F(2,720)=4.21$, $P>.01$. Object Identification scores for all three groups differed significantly from one another by the post-hoc Tukey tests. Distributions of both agnosia measures were skewed. Kruskal–Wallis nonparametric analysis of variance showed the same results. Object Identification remained statistically significant $[\chi^2(2, n=720)=94.6, P<.001]$, but Color Identification did not differ significantly across the three groups $[\chi^2(2, n=720)=7.1, P>.01]$.

Visuomotor test scores showed more pronounced changes with increased severity of AD. Block Design scores dropped to less than half the level of the nondemented group, $F(2,705)=144.44$, $P<.001$, after the deletion of one outlier. The control group differed significantly from both AD groups, which did not differ from one another on the Tukey tests. One outlier was also deleted from the Digit Symbol analysis, which showed similar levels of decline, $F(2,644)=145.41$, $P<.001$. The post-hoc Tukey tests showed that the AD groups differed significantly from the control group, but not from each other. The Drawing test showed the greatest sensitivity to the degree of severity, $F(2,911)=220.66$, $P<.001$. The Tukey post-hoc test showed all groups differed significantly from one another.

The neuropsychological tests clearly did not show equivalent changes in performance from mild to moderate AD. The standard scores for the two AD groups (Table 2) show the effects most clearly. Some tests showed modest rates of decline of about one S.D. from no cognitive impairment to mild AD, such as Similarities, Comprehension, Verbal Fluency, and Object Identification. Others, such as Drawing and Animal Fluency, showed larger changes with mild AD. Some tests continued to decline from mild to moderate AD, such as Animal Fluency, Naming, Object Identification, and Drawing, while others showed a much less pronounced change (Comprehension, Similarities, Digit Span, Block Design, and the Token test). Animal Fluency showed the greatest association with the severity of AD, while Color and Object Identification showed the least association.

### 3.4. Probable vs. possible AD

Table 3 compares the means for the nondemented comparison group with the subgroups of possible and probable AD for the same basic demographic information and test scores as in Table 2. The same conventions were used for this set of analyses as for those reported in Table 2. In this case, the multivariate analysis of variance was also significant, Pillai’s trace $=0.492$, $F(22,1232)=18.29$, $P<.0001$. Values of $\eta^2$ are also given in Table 3 for the strength of association of the probability independent variable with the neuropsychological tests.

The measures of abstract reasoning and judgement showed no distinction across the two groups with AD, both of which differed from the control group. Both effects were roughly the same size, although the proportionate change was less for Comprehension, $F(2,750)=163.9$, $P<.001$, than it was for Similarities, $F(2,749)=132.50$, $P<.001$. Tukey tests showed that the control group differed from both AD groups, which did not differ from each other.
A similar pattern was shown with the aphasia measures. The control group had the highest scores on the Token test, while the two AD groups did not differ, $F(2,721) = 99.95, P < .001$, with one outlier removed. The retention of that person did not change the results of the analysis. Verbal Fluency and Animal Fluency showed the same pattern, $F(2,728) = 106.96, P < .001$ for Verbal Fluency, and $F(2,729) = 239.78, P < .001$ for Animal Fluency. This result was also the case for the object naming measure, $F(2,730) = 56.01, P < .001$. The distribution of scores for the naming test was highly skewed, and the Kruskal–Wallis statistic for this variable replicated the parametric result $[\chi^2(2, N = 733) = 113.7, P < .001]$. Scores on the measures of agnosia showed little differences across groups, although the overall tests were significant for Object Identification, but not for Color Identification, $F(2,731) = 37.44, P < .001$ and $F(2,724) = 2.67, P > .01$, respectively. Post-hoc Tukey tests showed that the control group differed from both AD groups on Object Identification. As was the case in the severity analyses, the agnosia measures both had negatively skewed distributions. Kruskal–Wallis analyses done in both cases led to the same conclusions as with the parametric analyses, $\chi^2(2, N = 734) = 85.9, P < .001$ for Object Identification and $\chi^2(2, N = 727) = 5.7, P > .01$ for Color Identification.

One outlier was removed from the final analysis for each of the Block Design and Digit Symbol visuomotor measures. In both cases, the conclusions remained the same as with all cases included. Both AD groups scored lower than the control group on Block Design, $F(2,712) = 148.34, P < .001$, and Digit Symbol, $F(2,649) = 148.43, P < .001$. The same pattern of results with no differences between the two AD groups was found on the Drawing measure, $F(2,1090) = 384.58, P < .001$ using the Tukey test.

The associations between the probability of AD and the neuropsychological measures were stronger than in the case of severity as reflected in the values of $\eta^2$. The Drawing measure and Animal Fluency both accounted for over two-thirds of the variance, while the agnosia measures again were at 10% or lower.

A $2 \times 2$ table was formed for the association between the diagnostic severity levels of mild and moderate AD and that of possible/probable AD in order to determine if the probability of AD was significantly associated with the severity of dementia. It was not; Fisher’s exact test, $P = .58$, Cohen’s kappa = −0.03.

4. Discussion

Overall, the results showed that the frequency of problems in the organization and regulation of visuomotor behavior and of abstract thinking and judgement were higher than that of problems with language functions in the group with AD. This was true both for the absolute level of performance on the neuropsychological tests and for the relative frequency of diagnosis.

The increase in impairment with increasing severity of dementia is well known, but has been quantified less often, particularly for visuomotor tests. Edwards, Deuel, et al. (1991) reported levels of apraxia at different levels of severity. However, the figures obtained here were somewhat lower at all levels of severity than those reported by Edwards, Deuel, et al., who also used a different method of staging the severity of the dementia, Berg’s (1988)
Clinical Dementia Rating. Yesavage, Brooks, Taylor, and Tinklenberg (1993) noted that the development of apraxia or aphasia was associated with more rapid deterioration than was the case in people with AD who did not have those disorders. At the same time, Forstl, Burns, Levy, and Cairns (1993) concluded that constructional defects were not specifically related to changes in the parietal lobes of people with advanced AD.

More rapid deterioration with severe aphasia was reported by Bracco et al. (1994) and by Kraemer, Tinklenberg, and Yesavage (1994). The present study did not specifically examine rates of deterioration using longitudinal data, but more severe changes in the semantic fluency measure (animals) than for the phonemic fluency (letters) measure were found. These measures are differentially sensitive to frontal and temporal lobe damage (Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998). The present results are thus consistent with the noted deterioration in both frontal and temporal lobes in AD, and highlight the disruption of frontal–temporal systems. However, this pattern of results is not unique to AD, but is seen both in normal adults and other forms of dementia (Hart et al., 1988; Monsch et al., 1994; Suhr & Jones, 1998). Our results are also consistent with Hodges et al. (1999) in finding greater impairment in semantic fluency over phonemic fluency (see Zec (1993) for a review on this issue).

Scores on the agnosia measures remained close to the maximum possible score in all groups. These measures were among the least sensitive to the progression of the disorder, although statistically significant differences in performance were found. There is little known about agnosia in AD, and basic references on the neuropsychology of AD, such as Salmon and Bondi (1997) and Zec (1993) do not refer to the assessment of agnosia. In this study, it was diagnosed least often of the various cognitive dysfunctions at the mild and moderate stages. It is also possible that the object and color identification tasks used in CSHA were not the most sensitive measures that could have been used. These two tasks can be classed as measures of anomia (Greek: “no name,” the inability to report the names of objects) and can be regarded more as measures of language function than of agnosia. In any event, it is clear that further study of agnosia and the other cognitive disorders in dementia are as worthy of study as those of memory functions. More research into agnosia in AD would certainly appear to be in order, as agnosia remains as one of the disorders that can contribute to the diagnosis of AD.

The mean scores of the neuropsychological tests suggested that all the test scores did not decline in parallel from the comparison group through the groups with mild and moderate AD. An analysis in which the scores were transformed to standard scores using the comparison group’s mean and S.D. confirmed the impression of differential sensitivity to the general change in overall cognitive functioning. The agnosia measures showed the least change while the Drawing test showed the largest change. The measures of abstract reasoning and judgement from the WAIS-R, Token test, and Verbal Fluency measures showed a pronounced decline in the group with mild AD, but then did not show a further sharp decline, whereas that was not the case for the semantic fluency and naming measures, which showed further declines in moderate AD. Such unequal rates of decline suggest that efforts to find stable subtypes of AD will be highly dependent upon the nature of the tests used in addition to the difficulty of obtaining samples at equivalent stages of cognitive decline.

The distinction between possible and probable AD is one that logically is made early in the diagnostic process, prior to there being sufficient evidence of impairment in several
areas of physical and cognitive functioning to confirm a diagnosis of definite AD. The
distinction is commonly repeated in educational material (e.g., Rosser, 1993), and so
becomes more widely used in daily practice. The distinction can refer to the likelihood of
AD being the cause of the observed cognitive deficits (Blacker et al., 1994). In practice,
however, it may be used instead in a less precise manner to describe the severity of the
deficits. The present study demonstrates that this latter usage is inaccurate and invalid, in
that there were no neuropsychological tests that showed any difference in performance
between groups described as possible and probable AD, and that there was no empirical
agreement between the two variables. Blacker et al. (1994) found that the distinction
between possible and probable AD was notably less reliable than other NINCDS-ADRDA
diagnoses. We also replicated Hom and Brewer’s (1991) report of similar results with a
different test battery in a small group (n = 42) of people with AD. The autopsy data that
would confirm the diagnosis of AD was not collected by the CSHA, and so the more
definitive analysis of the utility of the probable/possible AD distinction was not possible.
The distinction between mild and moderate AD was supported to a greater extent in that
generally test scores were lower in the group classified as moderately demented. This was
not uniformly the case, however. One likely cause for this finding is that different cognitive
functions are affected at different rates by the disease process. It is likely that clinicians
primarily base the distinction between mild and moderate AD upon the severity of the
memory impairment and of functional impairment. Cognitive functions that are affected
later in the disorder, such as some language and motor functions, would not readily
discriminate mild from moderate AD. This point could be addressed better in a longitudinai
study that assessed distinct cognitive functions.

The procedure used to select cases for neuropsychological assessment led to reduced
numbers of people with severe AD in the sample and the number of cases with missing
diagnoses of the specific cognitive disorders may both raise concerns over the generalizability
of the results. No differences in the frequency of missing diagnoses were noted for gender,
and the differences in age were never more than 2 years, a relatively minor difference in view
of the advanced age of the sample, with the missing cases invariably being younger for those
whose diagnosis was not made. While it is not possible to be entirely confident of the
generality of the results, the absence of any major differences between those with missing
diagnoses and those with complete data is of some reassurance.

The missing data for the neuropsychological tests were most frequent at the highest level
of severity, leading to the deletion of this group from some analyses. This group is obviously
also the most impaired and the presence of missing data is not surprising. In large part, the
missing data reflect the original decision that individuals scoring below 50 on the 3MS would
not undergo the neuropsychological test battery (Canadian Study of Health and Aging
Working Group, 1994; Graham et al., 1996). Such individuals would be very likely to have
moderate to severe cognitive impairment and to be missing from these analyses. Tuokko et al.
(1995) detail the causes for missing data from the clinical examination and the neuropsy-
chological assessment. The single largest cause (n = 281) was refusal to participate. The
missing data do have implications for the generalizability of the results, and should be
considered in their interpretation, although no more than roughly 10% of the sample was
missing from any given analysis.
Overall, this study shows the progression of impairment in specific cognitive functions both in terms of the frequency of diagnosis and in quantitative terms of impaired test performance with increasing severity of AD. The general pattern of results suggests that the likelihood of identifying stable subtypes of AD based upon cognitive dysfunction is not likely due to the apparent differential sensitivity of tests to the overall rate of cognitive decline. The implication of these results for clinical practice is that the choice of tests in the assessment of individuals suspected of having AD should be made with particular care. It is clear that some commonly used tests are comparatively insensitive to the changes in function in AD, particularly during the early stages. Tests that have been shown to be sensitive in the early stages, such as fluency and drawing tasks, are more useful for purposes of early diagnosis than are less sensitive ones. It is also clear that much more research on changes in cognitive functions with increasing severity of AD is in order, particularly with the increasing number of medications being marketed for use in the early treatment of people with AD.

Acknowledgments

We thank anonymous reviewers for suggesting the analysis comparing possible and probable AD, the addition of effect sizes, and the use of standard scores to compare the tests.

References


