Cognitive Impairment in Preclinical Alzheimer’s Disease: A Meta-Analysis

Lars Bäckman
Karolinska Institute, Stockholm Gerontology Research Center, and Max Planck Institute for Human Development

Sari Jones, Anna-Karin Berger, and Erika Jonsson Laukka
Karolinska Institute and Stockholm Gerontology Research Center

Brent J. Small
University of South Florida

To determine the size of the impairment across different cognitive domains in preclinical Alzheimer’s disease (AD), a meta-analysis based on 47 studies involving 9,097 controls and 1,207 preclinical AD cases was conducted. There were marked preclinical deficits in global cognitive ability, episodic memory, perceptual speed, and executive functioning; somewhat smaller deficits in verbal ability, visuospatial skill, and attention; and no preclinical impairment in primary memory. Younger age (< 75 years) and shorter follow-up intervals (< 3 years) were associated with larger effect sizes for both global cognitive ability and episodic memory. For global cognitive ability, studies that used population-based sampling yielded larger effect sizes; for episodic memory, larger differences were seen in studies that preidentified groups in terms of baseline cognitive impairment. Within episodic memory, delayed testing and recall-based assessment resulted in the largest effect sizes. The authors conclude that deficits in multiple cognitive domains are characteristic of AD several years before clinical diagnosis. The generalized nature of the deficit is consistent with recent observations that multiple brain structures and functions are affected long before the AD diagnosis.

Keywords: preclinical, Alzheimer’s disease, cognition, dementia, meta-analysis, moderating factors

During the past decade, considerable evidence has accumulated indicating that individuals who will develop Alzheimer’s disease (AD) exhibit cognitive deficits several years before a clinical diagnosis of dementia can be rendered (e.g., Chen et al., 2001; Elias et al., 2000; B. J. Small, Fratiglioni, Viitanen, Winblad, & Bäckman, 2000; for an overview, see Bäckman, Small, & Fratiglioni, 2004). This is evidenced by a performance decrement at a baseline assessment point among persons who are currently without dementia but who will receive an AD diagnosis after a follow-up interval. Such preclinical deficits have been observed across multiple cognitive domains, including episodic memory (e.g., Bäckman, Small, & Fratiglioni, 2001), executive functioning (e.g., Albert, Moss, Tanzi, & Jones, 2001), verbal ability (e.g., Jacobs et al., 1995), visuospatial skill (e.g., B. J. Small, Herlitz, Fratiglioni, Almkvist, & Bäckman, 1997), attention (e.g., Linn, Wolf, Bachman, & Knoefel, 1995), and perceptual speed (e.g., Fabrigoule et al., 1998). Given this apparent breadth of impairment, it is only logical that similar deficits have been observed for global indicators of cognitive functioning, such as the Mini-Mental State Examination (MMSE; e.g., B. J. Small, Viitanen, & Bäckman, 1997) and composite measures of cognitive ability (e.g., Fabrigoule et al., 1998).

Knowledge pertaining to the preclinical phase of AD is important for theoretical and clinical reasons alike. Theoretically, knowledge regarding the transition from normal aging to dementia is vital in furthering an understanding of how the disease evolves. From a clinical perspective, identifying individuals at risk for developing AD as early as possible is imperative for maximizing treatment efficacy (Flicker, 1999; Post, 1999).

Although multiple cognitive domains are affected in preclinical AD, several studies have suggested that tasks assessing episodic memory (e.g., word recall, face recognition) may be particularly effective at identifying at-risk individuals (Elias et al., 2000; B. J. Small, Herlitz, et al., 1997; Tierney, Szalai, Snow, Fisher, Nores, et al., 1996). These observations are consistent with findings that brain structures known to be critical to episodic remembering, such as the hippocampus and neighboring regions (Nyberg, McIntosh, Houle, Nilsson, & Tulving, 1996; Vargha-Khadem et al., 1997), show pathological alterations long before clinical diagnosis (Braak & Braak, 1995; Fox et al., 1996; S. A. Small, Perera, DeLaPaz, Mayeux, & Stern, 1999). A recurrent debate in the literature on episodic memory in preclinical AD has centered on the issue of whether some episodic memory tasks (e.g., free recall) are more effective than others (e.g., recognition) in signaling an impending...
literature review in order to minimize the possibility of overlooking studies.

Search Strategy and Selection Criteria

The diagnostic criteria for AD (McKhann et al., 1984). The search was performed because it corresponds to the introduction of more systematic and reliable diagnostic criteria. The year 1985 was chosen as a starting point for the literature search of online databases (MEDLINE and PsycINFO). The outcome measure was baseline cognitive performance for incident AD cases and controls. An inspection of the studies included revealed that the outcome measure was baseline cognitive performance for incident AD cases and controls.

Our goal in the present study was threefold. First, we sought to determine the magnitude and pattern of impairment across multiple cognitive domains in preclinical AD. Thus, baseline cognitive performance of individuals who would and would not develop AD across a follow-up period constituted the outcome measure in the current analysis. The following domains were identified: global cognitive ability, episodic memory, executive functioning, verbal ability, visuospatial skill, attention, perceptual speed, and primary memory. Of chief interest was whether there would be systematic differences among different cognitive domains in the size of the preclinical impairment. Second, we examined the influence of study characteristics on the pattern of outcome in tests of global cognitive ability and episodic memory. These analyses involved comparisons across (a) participant age, (b) the time between preclinical assessment and diagnosis, and (c) the nature of the study sample. With regard to the latter, we examined differences between studies that used population-based versus convenience samples, as well as between studies in which at-risk individuals were identified a priori on the basis of some classificatory system (e.g., mild cognitive impairment, or MCI) versus those in which no such procedures were adopted. A third goal was to determine whether there were differences in effect size as a function of task characteristics (i.e., retention interval, retrieval demands, nature of the learning materials) within the key domain of episodic memory.

Method

Search Strategy and Selection Criteria

Articles published between January 1985 and February 2003 were identified through an extensive literature search of online databases (MEDLINE and PsycINFO). The year 1985 was chosen as a starting point because it corresponds to the introduction of more systematic and reliable diagnostic criteria for AD (McKhann et al., 1984). The search was performed in March 2003 and was limited to published English-language articles with human participants. The keywords used were Alzheimer’s disease, Alzheimer’s disease, Alzheimer’s, or AD in combination with preclinical, subclinical, prodromal, prediagnostic, pre-diagnostic, presymptomatic, pre-symptomatic, early stages, early symptoms, early diagnosis, early detection, mild cognitive impairment, MCI, cognitive impairment not dementia, CIND, conversion, prediction, transition, incidence AND cognition, incidence AND longitudinal, or incidence AND follow-up. In addition, we examined the tables of contents of relevant journals published within the 12 months prior to the completion of the literature review in order to minimize the possibility of overlooking studies that may not yet have been included in computerized databases.

The journals that were searched were American Journal of Epidemiology; American Journal of Psychiatry; Archives of General Psychiatry; Archives of Neurology; Brain; Brain and Cognition; British Journal of Psychiatry; International Journal of Geriatric Psychiatry; Journal of the American Geriatrics Society; Journal of the American Medical Association; Journal of Clinical and Experimental Neuropsychology; Journal of the International Neuropsychological Society; Journals of Gerontology, Series B: Psychological Sciences and Social Sciences; Journals of Gerontology, Series A: Biological Sciences and Medical Sciences; Journal of Neurology, Neurosurgery, and Psychiatry; Neurology; Neuropsychology; Psychology and Aging; and Psychological Medicine.

Studies included in the meta-analysis had to satisfy several criteria. These criteria were implemented to ensure that the studies were sufficiently similar to justify their contribution to the current analysis. First, participants had to be free of clinical dementia at a baseline assessment point, with a portion of the sample receiving, after a follow-up period, a diagnosis of AD according to standardized clinical criteria (e.g., National Institute of Neurological and Communication Disorders and Stroke—Alzheimer’s Disease and Related Disorders Association [NINCDS–ADRDA; McKhann et al., 1984]), the revised third edition of the Diagnostic and Statistical Manual of Mental Disorders [DSM–III–R; American Psychiatric Association, 1987], the fourth edition of the DSM [DSM–IV; American Psychiatric Association, 1994], and the 10th edition of the International Statistical Classification of Diseases and Related Health Problems [ICD–10; World Health Organization, 1992]). Note that we only selected studies in which AD was the diagnostic outcome. Studies that used dementia as the outcome were excluded because of the potential heterogeneity of this category. Second, all participants must have received an assessment of cognitive functioning at baseline. Finally, the studies had to include sufficient statistical information to allow for effect sizes to be calculated, which could include means and standard deviations, exact p values, t values, or F values, and sample sizes. In cases in which this information was not originally reported, we contacted the study authors directly to obtain the relevant statistics.

In total, we obtained results from 50 studies that met the criteria for inclusion in the present meta-analysis. However, in some cases, the retrieved articles came from the same research group and were not derived from completely independent samples. Specifically, studies from the Kungsholmen Project (Bäckman & Small, 1998; Bäckman, Small, & Fratiglioni, 2001; Berger, Fratiglioni, Forsell, Winblad, & Bäckman, 1999; B. J. Small et al., 2000; B. J. Small, Hertzl, et al., 1997; B. J. Small, Viitanen, & Bäckman, 1997), studies from the University of California, San Diego research group (Bondi, Salmon, Galasko, Thomas, & Thal, 1999; Lange et al., 2002; Olichney et al., 2002), and work by Tierney and colleagues (Tierney, Szalai, Dunn, Geslani, & McDowell, 2000; Tierney, Szalai, Snow, & Fisher, 1996; Tierney, Szalai, Snow, Fisher, Nores, et al., 1996; Tierney, Szalai, Snow, Fisher, Tsuda, et al., 1996) exhibited enough similarities, with the exception of differences in the numbers of participants, to raise concerns about the independence of the study samples. In these cases, the inclusion of studies was determined by whether each study contributed unique information. As a result, three studies (Berger et al., 1999; Tierney, Szalai, Snow, & Fisher, 1996; Tierney, Szalai, Snow, Fisher, Tsuda, et al., 1996) were excluded. All remaining studies contributed unique information.

Thus, a total of 47 studies contributed data to the meta-analysis, comprising a total of 1,207 preclinical AD cases and 9,097 controls. The basic characteristics of each of the studies included are shown in Table 1. A full reference list of all studies included in the meta-analysis is provided on the Web at http://dx.doi.org/10.1037/0022-006X.54.4.520.supp.

Cognitive Outcome Measures

The outcome measure was baseline cognitive performance for incident AD cases and controls. An inspection of the studies included revealed that
Table 1
Characteristics of the 47 Studies Included in the Meta-Analysis

<table>
<thead>
<tr>
<th>Author</th>
<th>Year published</th>
<th>Cognitive ability domain(s) examined</th>
<th>Preclinical AD</th>
<th>Diagnostic criteria</th>
<th>Follow-Up interval (years)</th>
<th>Participant characteristics</th>
<th>Sampling method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert et al.</td>
<td>2001</td>
<td>AT, EM, EX, PM, PS, VA, VS</td>
<td>23</td>
<td>NIH</td>
<td>71.8</td>
<td>3</td>
<td>intact</td>
</tr>
<tr>
<td>Andreassen et al.</td>
<td>1999</td>
<td>GC</td>
<td>16</td>
<td>DSM/NIH</td>
<td>71.5</td>
<td>1.1</td>
<td>MCI</td>
</tr>
<tr>
<td>Arnáz et al.</td>
<td>2001</td>
<td>GC</td>
<td>9</td>
<td>DSM/NIH</td>
<td>62.3</td>
<td>3.2</td>
<td>MCI</td>
</tr>
<tr>
<td>Bäckman &amp; Small</td>
<td>1998</td>
<td>PM</td>
<td>24</td>
<td>DSM/NIH</td>
<td>82.8</td>
<td>2.9</td>
<td>intact</td>
</tr>
<tr>
<td>Bäckman et al.</td>
<td>2001</td>
<td>EM, PM</td>
<td>15</td>
<td>DSM/NIH</td>
<td>81.2</td>
<td>6.7</td>
<td>intact</td>
</tr>
<tr>
<td>Bacon et al.</td>
<td>1998</td>
<td>GC</td>
<td>8</td>
<td>DSM/NIH</td>
<td>74.2</td>
<td>—</td>
<td>intact</td>
</tr>
<tr>
<td>Bizdan &amp; Bizdan</td>
<td>2002</td>
<td>GC</td>
<td>19</td>
<td>DSM/NIH</td>
<td>75.5</td>
<td>5</td>
<td>cimprd</td>
</tr>
<tr>
<td>Bondi et al.</td>
<td>1999</td>
<td>PM</td>
<td>7</td>
<td>DSM/NIH</td>
<td>70.8</td>
<td>3.7</td>
<td>intact</td>
</tr>
<tr>
<td>Chen et al.</td>
<td>2000</td>
<td>AT, EM, EX, GC, VA, VS</td>
<td>120</td>
<td>DSM/NIH</td>
<td>75.6</td>
<td>1.5</td>
<td>intact</td>
</tr>
<tr>
<td>Convit et al.</td>
<td>2000</td>
<td>EM, GC, VA</td>
<td>14</td>
<td>DSM/NIH</td>
<td>72.4</td>
<td>3.2</td>
<td>cimprd</td>
</tr>
<tr>
<td>Daly et al.</td>
<td>2000</td>
<td>EM, EX</td>
<td>23</td>
<td>NIH</td>
<td>71.8</td>
<td>3</td>
<td>cimprd</td>
</tr>
<tr>
<td>de Leon et al.</td>
<td>1993</td>
<td>GC</td>
<td>25</td>
<td>DSM/NIH</td>
<td>71.1</td>
<td>3.9</td>
<td>cimprd</td>
</tr>
<tr>
<td>Devanand et al.</td>
<td>2000</td>
<td>GC</td>
<td>19</td>
<td>DSM/NIH</td>
<td>65.7</td>
<td>1.6</td>
<td>cimprd</td>
</tr>
<tr>
<td>Fabrigoule et al.</td>
<td>1998</td>
<td>EM, GC, PS, VA</td>
<td>16</td>
<td>DSM/NIH</td>
<td>72.9</td>
<td>2</td>
<td>intact</td>
</tr>
<tr>
<td>Fowler et al.</td>
<td>2002</td>
<td>EM, GA, VA</td>
<td>9</td>
<td>NIH</td>
<td>59.0</td>
<td>2</td>
<td>intact/cimprd</td>
</tr>
<tr>
<td>Fox et al.</td>
<td>1998</td>
<td>EM, PS, GA</td>
<td>10</td>
<td>NIH</td>
<td>44.7</td>
<td>2.6</td>
<td>intact</td>
</tr>
<tr>
<td>Grober &amp; Kawas</td>
<td>1997</td>
<td>EM, GC, VA</td>
<td>20</td>
<td>NIH</td>
<td>79.1</td>
<td>4</td>
<td>intact</td>
</tr>
<tr>
<td>Hall et al.</td>
<td>2000</td>
<td>EM</td>
<td>35</td>
<td>DSM/NIH</td>
<td>80.3</td>
<td>2.2</td>
<td>intact</td>
</tr>
<tr>
<td>Howieson et al.</td>
<td>1997</td>
<td>EM, GC, PM, VA, VS</td>
<td>16</td>
<td>NIH</td>
<td>86.8</td>
<td>2.8</td>
<td>intact</td>
</tr>
<tr>
<td>Huang et al.</td>
<td>2000</td>
<td>GC</td>
<td>14</td>
<td>NIH</td>
<td>62.2</td>
<td>2.1</td>
<td>MCI</td>
</tr>
<tr>
<td>Jacobs et al.</td>
<td>1995</td>
<td>EM, VA</td>
<td>41</td>
<td>DSM/NIH</td>
<td>73.3</td>
<td>2.1</td>
<td>intact</td>
</tr>
<tr>
<td>Jacobson et al.</td>
<td>2002</td>
<td>EM, VA</td>
<td>20</td>
<td>NIH</td>
<td>75.2</td>
<td>1.3</td>
<td>intact</td>
</tr>
<tr>
<td>Jelic et al.</td>
<td>2000</td>
<td>GC</td>
<td>14</td>
<td>NIH</td>
<td>59.1</td>
<td>1.8</td>
<td>MCI</td>
</tr>
<tr>
<td>Kogure et al.</td>
<td>2000</td>
<td>EM, GC, PM, VS</td>
<td>32</td>
<td>DSM/NIH</td>
<td>69.8</td>
<td>2</td>
<td>cimprd</td>
</tr>
<tr>
<td>Lange et al.</td>
<td>2002</td>
<td>GC</td>
<td>20</td>
<td>DSM/NIH</td>
<td>74.6</td>
<td>2</td>
<td>intact</td>
</tr>
<tr>
<td>Lindeboom et al.</td>
<td>2002</td>
<td>GC</td>
<td>24</td>
<td>DSM/NIH</td>
<td>73.2</td>
<td>2</td>
<td>intact</td>
</tr>
<tr>
<td>Lindsay et al.</td>
<td>2002</td>
<td>GC</td>
<td>30</td>
<td>DSM/NIH</td>
<td>73.3</td>
<td>5</td>
<td>intact</td>
</tr>
<tr>
<td>Linn et al.</td>
<td>1995</td>
<td>EM, PM, VA</td>
<td>55</td>
<td>NIH</td>
<td>72.5</td>
<td>6</td>
<td>intact</td>
</tr>
<tr>
<td>Meyer et al.</td>
<td>2002</td>
<td>GC</td>
<td>35</td>
<td>DSM/NIH</td>
<td>73.7</td>
<td>3.9</td>
<td>MCI</td>
</tr>
<tr>
<td>Nielson et al.</td>
<td>1999</td>
<td>AT, EM, GC, VS</td>
<td>102</td>
<td>DSM/NIH</td>
<td>72.8</td>
<td>3.9</td>
<td>intact</td>
</tr>
<tr>
<td>Okamura et al.</td>
<td>2002</td>
<td>GC</td>
<td>8</td>
<td>NIH</td>
<td>71.4</td>
<td>3.1</td>
<td>MCI</td>
</tr>
<tr>
<td>Olichney et al.</td>
<td>2002</td>
<td>GC</td>
<td>7</td>
<td>NIH</td>
<td>74.3</td>
<td>2</td>
<td>MCI</td>
</tr>
<tr>
<td>Riemenscheider et al.</td>
<td>2002</td>
<td>EM, GC</td>
<td>10</td>
<td>NIH</td>
<td>69.5</td>
<td>1.5</td>
<td>MCI</td>
</tr>
<tr>
<td>Rubin et al.</td>
<td>1998</td>
<td>AT, EM, PM, VA, VS</td>
<td>16</td>
<td>NIH</td>
<td>71.7</td>
<td>1.5</td>
<td>intact</td>
</tr>
<tr>
<td>Schaub et al.</td>
<td>2002</td>
<td>GC</td>
<td>17</td>
<td>DSM/NIH</td>
<td>—</td>
<td>3</td>
<td>intact</td>
</tr>
<tr>
<td>Schmand et al.</td>
<td>2000</td>
<td>EM, GC</td>
<td>25</td>
<td>DSM/NIH</td>
<td>73.5</td>
<td>3</td>
<td>intact</td>
</tr>
<tr>
<td>Small, Herlitz et al.</td>
<td>1997</td>
<td>GC</td>
<td>32</td>
<td>DSM/NIH</td>
<td>83.5</td>
<td>3.1</td>
<td>intact</td>
</tr>
<tr>
<td>Small, Viitanen et al.</td>
<td>1997</td>
<td>EM, PM, VA</td>
<td>26</td>
<td>DSM/NIH</td>
<td>83.5</td>
<td>3.1</td>
<td>intact</td>
</tr>
<tr>
<td>Small et al.</td>
<td>2000</td>
<td>GC</td>
<td>71</td>
<td>DSM/NIH</td>
<td>80.6</td>
<td>6.7</td>
<td>intact</td>
</tr>
<tr>
<td>Smith et al.</td>
<td>1998</td>
<td>GC</td>
<td>35</td>
<td>DSM/NIH</td>
<td>79.8</td>
<td>3.5</td>
<td>MCI</td>
</tr>
<tr>
<td>Tabert et al.</td>
<td>2002</td>
<td>GC</td>
<td>23</td>
<td>NIH</td>
<td>67.2</td>
<td>2.0</td>
<td>MCI</td>
</tr>
<tr>
<td>Tierney et al.</td>
<td>1996</td>
<td>AT</td>
<td>29</td>
<td>DSM/NIH</td>
<td>72.7</td>
<td>2</td>
<td>cimprd</td>
</tr>
<tr>
<td>Tierney et al.</td>
<td>2000</td>
<td>EM, GC, VA</td>
<td>29</td>
<td>DSM/NIH</td>
<td>71.7</td>
<td>2</td>
<td>cimprd</td>
</tr>
<tr>
<td>Tröster et al.</td>
<td>1994</td>
<td>GC</td>
<td>4</td>
<td>DSM/NIH</td>
<td>67.2</td>
<td>5</td>
<td>intact</td>
</tr>
<tr>
<td>Tuokko et al.</td>
<td>1991</td>
<td>EM</td>
<td>18</td>
<td>DSM/NIH</td>
<td>68.7</td>
<td>1.3</td>
<td>intact</td>
</tr>
<tr>
<td>Visser et al.</td>
<td>2001</td>
<td>EM, GC, VA</td>
<td>23</td>
<td>DSM/NIH</td>
<td>68.9</td>
<td>5</td>
<td>MCI</td>
</tr>
<tr>
<td>Wolf et al.</td>
<td>2000</td>
<td>GC</td>
<td>8</td>
<td>DSM/ACD</td>
<td>72.0</td>
<td>2.4</td>
<td>MCI</td>
</tr>
</tbody>
</table>


There was sufficient information on enough individual cognitive tests that they could be organized into several broad domains of functioning. The tasks were organized into eight categories of cognitive ability on the basis of the typical association between tests and ability domains seen in the neurologic and neuropsychological literature (e.g., Lezak, 1995; Spreen & Strauss, 1998). The specific domains identified (and associated tests) were as follows: global cognitive ability (e.g., the MMSE [Folstein, Folstein, & McHugh, 1975], the Cognitive test section of the Cambridge Examination for Mental Disorders of the Elderly [Roth, Huppert, Mountjoy, & Tyn, 1988]; episodic memory (e.g., the California Verbal Learning Test [Delis,
allows one to determine whether the two groups are clearly separable. This statistic (O/L%) represents the degree of overlap in the distributions of the variables used were either at the level of study characteristics or in the nature of the episodic memory tests employed. Among study characteristics, we were interested in the effect of participant age (<75 years, ≥75 years), length of follow-up period (<3 years, ≥3 years), sampling method (population based, convenience), and initial cognitive status of participants (intact, impaired). For age and length of follow-up, the division was based on a median split across the studies included. With respect to sampling method, population-based studies were those that used epidemiologic-based methods to recruit persons from a specified population (e.g., Fabricolle et al., 1998; B. J. Small, Herlitz, et al., 1997). By contrast, convenience samples included persons who were selected from clinical settings (e.g., Andreaan et al., 1999; Devanand et al., 2000) or recruited through appeals in the media or other public outlets (e.g., Grober & Kawas, 1997; Howieson et al., 1997). In terms of initial cognitive status, participants were considered impaired if they had been classified with cognitive dysfunction (i.e., MCI [e.g., Petersen et al., 1999] or memory impairment [e.g., Kogure et al., 2000]) at baseline.

In addition to study characteristics, we examined the extent to which group differences in episodic memory functioning varied as a function of task characteristics. Specifically, we were interested in the effect of retention interval (immediate, delayed), study materials (verbal, nonverbal), and retrieval support (recall, recognition) on the magnitude of observed effects.

**Data Extraction and Statistical Analysis**

The system for the meta-analysis was based on a random-effects procedure, as described by Hedges and Olkin (1985). From the data reported in each study, the effect-size estimate, g, was calculated, indicating the difference between the control and incident AD groups divided by the pooled standard deviation. Thus, g represents the standardized difference between the two groups within each study. In order to pool the results across studies, the effect size, d, was computed, which represents the standardized difference between the preclinical AD and nondementia groups weighted by the sample sizes of the individual studies. The weighting procedure affects the variance estimate from each study because, as Hedges and Olkin (1985) noted, “the variance estimates for studies with larger sample sizes are more precise than those for studies with smaller effect sizes” (p. 107). For both g and d, the direction of the effect size was positive if the performance of the incident AD group was worse than that of the control participants. In studies in which more than one dependent measure was present for a cognitive domain (e.g., multiple tests of episodic memory), an averaged effect size was calculated for the overall analysis to avoid having one study dominate the results (Heinrichs & Zakzanis, 1998). For example, if a study had tests of immediate and delayed episodic memory, these effect sizes were averaged to generate the overall effect size for episodic memory. However, when examining effect sizes as a function of retention interval, we treated the two tests separately.

We also calculated the chi-square statistic Q, indicating homogeneity of results across effect sizes. If a significant Q value is observed, this indicates heterogeneity of results and may result in a search for potential moderating variables. In the moderator analysis, the QM statistic indicates the degree of homogeneity within a class of studies or a cognitive domain. The QM statistic refers to a test of differences between categories and, if statistically significant, suggests the influence of a moderating variable. Because of the large sample sizes in this study, an alpha level of .01 was adopted as the statistical criterion.

The degree of overlap between the control and incident AD groups was determined using the U statistic (Heinrichs & Zakzanis, 1998). The overlap statistic (O/L%) represents the degree of overlap in the distributions of the cognitive scores for the incident AD and control groups. This statistic allows one to determine whether the two groups are clearly separable.

**Moderator Variables**

In addition to overall effect sizes, we examined the influence of several potential moderating variables, using categorical models. The moderator variables used were either at the level of study characteristics or in the nature of the episodic memory tests employed. Among study characteristics, the degree of overlap between the control and incident AD groups was calculated using the U statistic (Heinrichs & Zakzanis, 1998). The overlap statistic (O/L%) represents the degree of overlap in the distributions of the cognitive scores for the incident AD and control groups. This statistic allows one to determine whether the two groups are clearly separable.

**Results**

**Overall Effect Sizes**

A concern in meta-analytic studies is the potential existence of publication bias and other biases related to the process of locating, selecting, and combining studies (Easterbrook, Berlin, Gopalan, & Matthews, 1991). One method for identifying outliers, and the possibility of biases, is the use of funnel plots, which assess the relationship between sample size and effect size. Using funnel plots (Egger, Davey Smith, Schneider, & Minder, 1997; Light & Pillemer, 1984), we addressed this issue for global cognitive ability and episodic memory, the domains that had the largest numbers of observations. As shown in Figure 1, for global cognitive ability, the majority of observations yielded g values between 0.5 and 2.0. A similar pattern was found for episodic memory, with g values ranging from 0.5 to 3.0. Note that the effect-size measures shown in Figure 1 are indexed by g, because data from individual studies are being plotted. The distributions approximated a funnel shape, with small-sample studies scattered at the bottom of the graphs and the peak converging on the average effect size for both cognitive domains. Thus, the shape of the plots provides no evidence for systematic bias within either cognitive domain.

![Figure 1](image-url)
Effect sizes for the eight cognitive domains are shown in Table 2. The incident AD group exhibited statistically significant deficits across all domains except primary memory. Global cognitive ability, episodic memory, perceptual speed, and executive functioning yielded the largest effect sizes, all above 1.0. Significant group differences were also observed for verbal ability, visuospatial skill, and attention, with $d$ values ranging from 0.62 to 0.79.

The overlap statistic is also shown in Table 2. Of special note here is that even for global cognitive ability, episodic memory, perceptual speed, and executive functioning, almost half of the distributions between the control and incident AD groups overlapped.

The $Q_w$ statistic revealed significant heterogeneity of effect sizes for all domains except executive functioning. In terms of potential moderators of this variability, we examined only global cognitive ability and episodic memory. We did so for two principal reasons. First, the number of studies that included measures of these domains was sufficient to permit a more fine-grained analysis. Second, the a priori categorization of episodic memory tasks, in terms of retention interval, study materials, and retrieval support, lent itself to examining these factors as potential moderators.

**Influence of Study Characteristics on Effect Sizes**

Table 3 shows the influence of the moderator variables on effect sizes for global cognitive ability and episodic memory. For both global cognitive ability and episodic memory, age and follow-up interval were associated with the magnitude of group differences. Specifically, younger age and shorter follow-up intervals resulted in larger differences between the incident AD and control groups. Studies in which population-based sampling was used revealed larger effect sizes for global cognitive ability, whereas there was a trend in the same direction for episodic memory ($p < .05$). Finally, for episodic memory, studies in which participants were preselected on the basis of cognitive impairment at baseline yielded larger effect sizes.

**Effect Sizes Within Episodic Memory**

Table 4 displays the effect sizes for episodic memory as a function of retention interval, retrieval support, and study materials. As evidenced by the $Q_w$ statistic, the results indicated significant differences in effect size as a function of retention interval and retrieval condition. Specifically, delayed testing and recall-based retrieval resulted in larger group differences. A trend was seen for study materials, with verbal materials resulting in somewhat larger group differences ($p < .05$).

**Discussion**

The results of this meta-analysis indicate a global impairment of cognitive functioning in the preclinical phase of AD. Specifically, for three of the specific cognitive domains targeted, the overall $d$ values were 1.03 (episodic memory), 1.07 (executive functioning), and 1.11 (perceptual speed), reflecting large effect sizes according to the nomenclature of Cohen (1988). For verbal ability ($d = 0.79$), visuospatial skill ($d = 0.64$), and attention ($d = 0.62$), the group difference was of moderate size, whereas it was nonexistent for primary memory ($d = 0.00$).

These findings are consistent with claims made by many investigators in this field (e.g., Bäckman, Small, & Fratiglioni, 2001; Elias et al., 2000; Grober & Kawas, 1997; Linn et al., 1995; B. J. Small, Herlitz, et al., 1997) that episodic memory impairment is a cardinal feature of an impending dementia disease, likely reflecting the fact that memory-relevant regions in the medial-temporal lobe are affected prior to the time at which dementia can be clinically diagnosed (e.g., Fox et al., 1996; S. A. Small et al., 1999) At the same time, these results demonstrate that episodic memory does not have a unique status among categories of cognitive markers for identifying forthcoming AD. The confidence intervals around the effect sizes overlapped for episodic memory, executive functioning, and perceptual speed. For verbal ability, visuospatial skill, and attention, the effect sizes were somewhat smaller but still quite sizable. The only domain that deviated markedly from the pattern of clear preclinical impairment was that of primary memory. However, given that basic primary memory operations are well preserved in early clinical AD (e.g., Martin, Brouwers, Cox, & Fedio, 1985; Simon, Leach, Winocur, & Moscovitch, 1995), there is little reason to believe that they would be affected preclinically.

The observation that the deficit in preclinical AD generalized across several major domains of cognition was corroborated by the fact that the category of global cognitive ability showed an effect size of 1.19. However, the large effect size obtained for this category does not imply that each cognitive domain represented in global assessment instruments such as the MMSE is equally impaired. For example, B. J. Small, Viitanen, and Bäckman (1997)

---

Table 2

Effect Sizes for the Eight Domains of Cognitive Functioning

<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of studies ($k$)</th>
<th>Controls</th>
<th>Cases</th>
<th>$d$</th>
<th>99% CI</th>
<th>O/L%</th>
<th>$Q_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Cognitive Ability</td>
<td>35</td>
<td>6,509</td>
<td>868</td>
<td>1.19*</td>
<td>1.15–1.24</td>
<td>38.8</td>
<td>538.53*</td>
</tr>
<tr>
<td>Episodic Memory</td>
<td>24</td>
<td>6,694</td>
<td>704</td>
<td>1.03*</td>
<td>0.98–1.08</td>
<td>42.1</td>
<td>768.85*</td>
</tr>
<tr>
<td>Verbal Ability</td>
<td>15</td>
<td>3,625</td>
<td>438</td>
<td>0.79*</td>
<td>0.73–0.85</td>
<td>57.0</td>
<td>391.35*</td>
</tr>
<tr>
<td>Visuospatial Skill</td>
<td>9</td>
<td>3,159</td>
<td>363</td>
<td>0.64*</td>
<td>0.58–0.71</td>
<td>61.8</td>
<td>24.44*</td>
</tr>
<tr>
<td>Primary Memory</td>
<td>8</td>
<td>1,567</td>
<td>205</td>
<td>0.00</td>
<td>−0.09–0.09</td>
<td>100.0</td>
<td>80.10*</td>
</tr>
<tr>
<td>Attention</td>
<td>5</td>
<td>2,969</td>
<td>289</td>
<td>0.62*</td>
<td>0.55–0.68</td>
<td>61.8</td>
<td>20.50*</td>
</tr>
<tr>
<td>Perceptual Speed</td>
<td>4</td>
<td>1,274</td>
<td>65</td>
<td>1.11*</td>
<td>1.00–1.22</td>
<td>42.1</td>
<td>163.47*</td>
</tr>
<tr>
<td>Executive Functioning</td>
<td>3</td>
<td>547</td>
<td>166</td>
<td>1.07*</td>
<td>0.92–1.21</td>
<td>42.1</td>
<td>2.45</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval; O/L% = overlap statistic.

* $p < .01$. 

---
demonstrated that three MMSE subscales, Orientation to Time, Orientation to Place, and Delayed Recall, were most effective in differentiating incident AD cases from controls 3 years before diagnosis.

The finding of a multiple cognitive system breakdown is consistent with observations that conversion rates to AD over a 2.5-year interval were considerably greater for persons who exhibited deficits in episodic memory and some other cognitive domain (e.g., verbal ability, visuospatial skill) at baseline than for persons who had isolated memory impairment (Bozoki, Giordani, Heidebrink, Berent, & Foster, 2001). Further, the breadth of cognitive impairment in preclinical AD is interesting to consider in view of the literature on normal cognitive aging. It is well established that episodic memory, executive functioning, and speed not only deteriorate markedly across the adult life span but also show strong associations in samples of individuals without dementia (e.g.,

| Table 3 | Effect Sizes for Global Cognitive Ability and Episodic Memory Across Study Characteristics |
|------------------------|-------------------------------|---------------------|------------------|------------------|-------------------|-----------------|-------------------|
| Moderator variable     | Number of studies (k) | Controls | Cases | d      | 99% CI          | $Q_N$ | $Q_B$ |
| Global cognitive ability | Age of participants  | < 75 years | 27     | 4,872 | 538 | 1.32 | 1.26–1.37 | 237.22** | 110.35** |
|                        | ≥ 75 years         | 8        | 1,808 | 347  | 0.91 | 0.83–0.99 | 195.68** |         |
| Follow-up interval     | < 3 years          | 20       | 2,511 | 444  | 1.29 | 1.22–1.37 | 212.97** | 22.85** |
|                        | ≥ 3 years          | 14       | 3,936 | 416  | 1.12 | 1.06–1.18 | 302.42** |         |
| Sampling method        | Population based   | 9        | 5,311 | 440  | 1.22 | 1.17–1.27 | 363.77** | 11.37** |
|                        | Convenience        | 26       | 1,198 | 428  | 1.08 | 0.98–1.17 | 163.38** |         |
| Participant characteristics | Intact            | 14       | 5,449 | 490  | 1.19 | 1.14–1.27 | 427.65** | 0.21   |
|                        | Impaired           | 21       | 1,060 | 378  | 1.17 | 1.07–1.27 | 110.67** |         |

Episodic memory

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of studies (k)</th>
<th>Controls</th>
<th>Cases</th>
<th>d</th>
<th>99% CI</th>
<th>$Q_N$</th>
<th>$Q_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention interval</td>
<td>Immediate</td>
<td>15</td>
<td>3,356</td>
<td>411</td>
<td>0.96</td>
<td>0.90–1.02</td>
<td>380.17**</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>17</td>
<td>3,388</td>
<td>453</td>
<td>1.23</td>
<td>1.16–1.30</td>
<td>516.89**</td>
</tr>
<tr>
<td>Retrieval support</td>
<td>Recall</td>
<td>22</td>
<td>6,493</td>
<td>656</td>
<td>1.02</td>
<td>0.97–1.06</td>
<td>561.19**</td>
</tr>
<tr>
<td></td>
<td>Recognition</td>
<td>6</td>
<td>937</td>
<td>194</td>
<td>0.79</td>
<td>0.68–0.90</td>
<td>74.92**</td>
</tr>
<tr>
<td>Study material</td>
<td>Verbal</td>
<td>24</td>
<td>6,694</td>
<td>704</td>
<td>1.05</td>
<td>1.00–1.09</td>
<td>756.17**</td>
</tr>
<tr>
<td></td>
<td>Nonverbal</td>
<td>8</td>
<td>2,476</td>
<td>187</td>
<td>0.96</td>
<td>0.89–1.02</td>
<td>431.70**</td>
</tr>
</tbody>
</table>

Note. Participant characteristics: Intact = not classified with cognitive impairment; Impaired = mild cognitive impairment or cognitively impaired but no dementia. CI = confidence interval.

*p < .05. **p < .01.

Table 4 | Effect Sizes for Episodic Memory Across Conditions |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>Number of studies (k)</td>
<td>Controls</td>
<td>Cases</td>
<td>d</td>
<td>99% CI</td>
<td>$Q_N$</td>
<td>$Q_B$</td>
</tr>
<tr>
<td>Retention interval</td>
<td>Immediate</td>
<td>15</td>
<td>3,356</td>
<td>411</td>
<td>0.96</td>
<td>0.90–1.02</td>
<td>380.17**</td>
</tr>
<tr>
<td></td>
<td>Delayed</td>
<td>17</td>
<td>3,388</td>
<td>453</td>
<td>1.23</td>
<td>1.16–1.30</td>
<td>516.89**</td>
</tr>
<tr>
<td>Retrieval support</td>
<td>Recall</td>
<td>22</td>
<td>6,493</td>
<td>656</td>
<td>1.02</td>
<td>0.97–1.06</td>
<td>561.19**</td>
</tr>
<tr>
<td></td>
<td>Recognition</td>
<td>6</td>
<td>937</td>
<td>194</td>
<td>0.79</td>
<td>0.68–0.90</td>
<td>74.92**</td>
</tr>
<tr>
<td>Study material</td>
<td>Verbal</td>
<td>24</td>
<td>6,694</td>
<td>704</td>
<td>1.05</td>
<td>1.00–1.09</td>
<td>756.17**</td>
</tr>
<tr>
<td></td>
<td>Nonverbal</td>
<td>8</td>
<td>2,476</td>
<td>187</td>
<td>0.96</td>
<td>0.89–1.02</td>
<td>431.70**</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
Bäckman, Small, & Wahlin, 2001; Park et al., 2002; Salthouse, 1996). Thus, the fact that the largest effect sizes were observed for these domains indicates continuity in patterns of cognitive impairment from normal aging to preclinical AD.

In addition, it is of interest to consider recent brain imaging and histopathological findings comparing controls and incident AD cases. Emerging evidence indicates that multiple brain structures and functions in addition to those related to the medial-temporal lobe may be affected before the AD diagnosis. These findings include volume reductions of anterior cingulate and temporal sulcus (Killiany et al., 2000), posterior cingulate and neocortical temporoparietal regions (Fox et al., 2001), and frontal regions (van der Flier, van den Heuvel, Weverling-Rijnsburger, Bollen, et al., 2002); decreased blood flow in posterior cingulate and precuneus (Kogure et al., 2000); reduced glucose metabolism in temporoparietal regions (Arnó et al., 2001); and deposits of amyloid plaques in temporal (Morris & Price, 2001) and frontal (Yamaguchi, Sugiura, Ogawa, Oshima, & Ihara, 2001) cortex. In addition, more general alterations of brain functions in preclinical AD have been observed, including an increase of white-matter hyperintensities (Wolf, Ecke, Bettin, Dietrich, & Gertz, 2000) as well as a reduction of whole-brain glucose metabolism (Silverman et al., 2001). Thus, given that these studies indicate a rather widespread affection of brain structures and functions in preclinical AD, it should come as no surprise that a similar extent of impairment is observed at the behavioral level. This point is further strengthened by recent evidence indicating strong relationships of volumetric differences (Wolf, Ecke, Bettin, Dietrich, & Gertz, 2000) as well as a reduction of whole-brain glucose metabolism (Silverman et al., 2001). However, note that the degree of impairment was quite sizable also for studies that used longer retest intervals, with $d$ values of 1.12 and 0.76 for global cognitive ability and episodic memory, respectively. This pattern of results is interesting in light of longitudinal evidence on the trajectory of cognitive decline in preclinical AD. In two studies, stability of cognitive impairment among incident AD patients was observed from 6 to 3 years prior to eventual diagnosis (Bäckman, Small, & Fratiglioni, 2001; B. J. Small et al., 2000). In another study, disproportionate decline was seen among incident AD cases from 3.5 to 1.5 years before diagnosis (Chen et al., 2001). Thus, the preclinical period in AD appears to be characterized by an early onset followed by relative stability until a few years before diagnosis, when precipitous cognitive decline occurs (e.g., Fox et al., 1996; Rubin et al., 1998; B. J. Small, Viitanen, & Bäckman, 1997).

For global cognitive ability, studies in which participants were recruited through population-based sampling revealed larger effect sizes than did studies that used convenience sampling (e.g., clinical settings, advertisements). There was a trend in the same direction for episodic memory. This may reflect the fact that memory and other cognitive problems are common among people who actively seek participation in studies on aging and cognition irrespective of whether they are in a preclinical phase of dementia. As a consequence, memory performance would also be expected to be lower for the controls in these types of studies, resulting in smaller group differences.

A key difference between studies in this area concerns whether participants have already been classified as cognitively impaired at the baseline assessment using categories such as MCI. On the one hand, it may appear reasonable that studies that follow a group of preclassified persons prospectively should yield larger effect sizes than studies that work retrospectively from diagnosis to baseline. On the other hand, recent evidence indicates that categories such as MCI are rather heterogeneous. For example, it has been shown that although a large portion of cognitively impaired older adults go on to develop AD within a few years, a sizable portion remain stable or even improve across the same time period (e.g., Helkala et al., 1997; Palmer, Wang, Bäckman, Winblad, & Fratiglioni, 2002). Obviously, the latter fact speaks against prospective studies resulting in larger effect sizes. For global cognitive ability, the confidence intervals for the two types of studies were overlapping, whereas for episodic memory, larger effect sizes were observed for MCI-type studies. The latter finding was expected in view of the fact that episodic memory impairment constitutes a cardinal criterion for inclusion in the MCI category (Petersen et al., 1999).

Within the domain of episodic memory, there were two significant effects pertaining to specific task characteristics. First, the effect size was larger for delayed than for immediate testing of memory. Provided that delayed testing taxes consolidation processes to a greater extent than does immediate testing (e.g., Haist, Gore, & Mao, 2001; Squire, 1986), this result is consistent with the view that failure in transferring information from temporary storage to a more permanent memory representation is a characteristic feature of the episodic memory impairment in preclinical AD (Bäckman & Small, 1998; Martin et al., 1985). Second, effect sizes were larger for recall than for recognition. Given that the retrieval demands are typically greater in recall than in recognition (e.g., Craik, 1983; Craik & McDowd, 1987; Jacoby, Toth, & Yonelinas, 1993), these data suggest that retrieval problems, in addition to difficulties in encoding and consolidation, may be characteristic of...
preclinical AD. There was also a trend in the direction of larger effect sizes for verbal than for nonverbal materials. This may reflect the fact that verbal materials (e.g., words, paired-associates) are typically poorer in terms of the features available at encoding than are nonverbal materials (e.g., faces, pictures). As a result, the requirement of self-initiated elaborative encoding operations is generally greater for verbal materials, which may be particularly handicapping for preclinical AD cases. In general, then, the moderating effects observed for episodic memory suggest that performance deficits in preclinical AD are exacerbated with increasing cognitive demands.

However, despite the observed influences of study and task characteristics, it is important to note that the effect sizes for global cognitive ability and episodic memory were uniformly large. Thus, the differences in effect sizes observed in this meta-analysis should not conceal the fact that the preclinical cognitive impairment in AD was found to generalize across (a) several key domains of cognitive functioning, (b) major characteristics of research studies and sample composition, and (c) multiple aspects of episodic memory.

Although the present meta-analysis revealed generally large effect sizes, the overlap in cognitive performance between cases and controls was quite substantial, with overlap scores ranging from 39% to 62% for those task domains yielding reliable group differences. Thus, a sizable proportion of the cases had scores that fell within normal ranges and vice versa. To be sure, it may not be realistic to expect nonoverlapping distributions of cognitive performance scores many years before diagnosis given the large interindividual variability in preclinical AD (e.g., Bäckman, Jones, Small, Agüero-Torres, & Fratiglioni, 2003; Fox et al., 1996; Palmer et al., 2002) as well among those who will remain without dementia (e.g., Anstey & Christensen, 2000; Bäckman, Small, Wahlin, & Larsson, 1999; Hultsch, Hertzog, Dixon, & Small, 1998). There are numerous reasons for exhibiting cognitive impairment in old age in addition to being in a preclinical phase of AD. These include psychiatric (e.g., Bäckman & Forsell, 1994), metabolic (e.g., Calvaresi & Bryan, 2001), immunological (e.g., C. J. Wilson, Finch, & Cohen, 2002), hormonal (e.g., Wahlin, Robins Wahlin, Small, & Bäckman, 1998), and circulatory (e.g., Fahlander et al., 2000) disturbances, all of which could potentially result in false negatives. Indeed, there is evidence that some persons classified as cognitively impaired at baseline not only remained free of AD many years later but actually exhibited improved cognitive performance at follow-up (e.g., Palmer et al., 2002). It is important to note that these improvers had no increased risk of being diagnosed with AD at a subsequent follow-up examination.

Further, persons who will be diagnosed with AD vary both in terms of onset of precipitous decline and rate of decline during the preclinical period (Bäckman et al., 2003; Rubin et al., 1998). In particular, incident AD cases who exhibit relatively normal cognitive performance until shortly before follow-up, when decline occurs very rapidly, will be difficult to catch. As with the data on patterns of cognitive impairment in normal aging and preclinical AD discussed above, these observations are consistent with a continuity view of AD development. Specifically, labels such as “normal aging,” “preclinical AD,” “MCI” and “early AD” may best be viewed as instances on a dimension of brain and cognitive functioning rather than as discrete categories. These facts obvi-ously hamper the utility of cognitive tests in identifying persons eligible for pharmacological or other interventions designed to slow down the disease process.

Having said that, we should note that the meta-analysis reported in this article provides a rather conservative test of differences between cases and controls in the sense that effect sizes were assessed in a univariate fashion (i.e., d values were estimated for single-task variables). There is evidence that the ability to identify persons at risk for developing AD increases substantially when tasks assessing different cognitive domains (e.g., episodic memory, executive functioning, verbal ability) are combined into the same prediction model (e.g., Albert et al., 2001; Chen et al., 2001; B. J. Small, Viitanen, & Bäckman, 1997). Obviously, this was not feasible in the current meta-analysis.

In addition, indicators of other domains of functioning may be useful in identifying at-risk individuals. As noted, these include multiple measures of brain structure and function such as volumetric measures (e.g., Fox et al., 1996, 2001; Killiany et al., 2000; van der Flier, van den Heuvel, Weaverling-Rijnsburger, Bollen, et al., 2002), glucose metabolism (e.g., Arnáiz et al., 2001; Silverman et al., 2001), blood flow (e.g., Kogure et al., 2000; S. A. Small et al., 1999), amyloid deposits (Morris & Price, 2001; Yamaguchi et al., 2001), and white-matter hyperintensities (Wolf et al., 2000). They also include markers for genetic predisposition such as the presence of the apolipoprotein e4 allele (for an overview, see Farrer et al., 1997), subjective memory complaints (Geerlings, Jonker, Bouter, Adèr, & Schmand, 1999; Palmer, Bäckman, Winblad, & Fratiglioni, 2003), family reports of cognitive impairment (Daly et al., 2000), and depressive symptoms (Berger et al., 1999; R. S. Wilson, Schneider, et al., 2003). The identification of specific risk factors acting as precipitating factors may provide further help in differentiating between preclinical AD and nonprogressive cognitive impairment. Possible precipitating factors include medical events such as unrecognized hypertension (Larner et al., 2000), head trauma (Plassman et al., 2000), and stroke (Snowdon et al., 1997) as well as factors such as engagement in intellectual activities (e.g., Scarmeas, Levy, Tang, Manly, & Stern, 2001; Wang, Karp, Winblad, & Fratiglioni, 2002), proneness to distress (R. S. Wilson, Evans, et al., 2003), and social isolation (Fratiglioni, Wang, Ericsson, Maytan, & Winblad, 2000).

From a clinical perspective, then, the point is that although a variety of different factors may show relatively good predictive accuracy in identifying persons in a preclinical phase of AD, these factors are rarely used in the same prediction model. An important avenue for future research is to combine predictors from different behavioral and biological domains. To the extent that the factors included account for unique variance in group classification, such a multivariate approach will increase overall prediction accuracy. Further, it enables the examination of possible interactive effects among various preclinical markers.

Although the present analysis of cognitive markers in preclinical AD provides a comprehensive account of the current state of knowledge in this domain of research, there are limitations to note. The first concern has to do with the precision with which the diagnosis of AD could be made among the studies included in the current analysis. We included only studies in which standardized diagnostic instruments were used (e.g., DSM–III–R, DSM–IV, ICD–10, NINCDS–ADRSA criteria). Recent evidence suggests that in research settings comparable to those represented in the
meta-analysis, accuracy of the clinical diagnosis is high. Specifically, a diagnostic accuracy around .90 has been reported in both clinical (e.g., Salmon et al., 2002) and community-based (e.g., Massoud et al., 1999) samples when verified against autopsy data. Nonetheless, occasional failures to identify mildly impaired persons as early AD cases at baseline assessment may have resulted in an overestimation of the cognitive differences between preclinical AD cases and controls.

Second, among the studies included in the analysis few had information on all, or even most, of the ability domains targeted. The fact that the effect-size estimates for some cognitive domains (i.e., global cognitive ability and episodic memory) were based on more than 7,000 participants, whereas those for others (i.e., executive functioning) were based on fewer than 1,000 participants has to be taken into account when comparing effect sizes across domains. As a result of these differences, the analyses of moderator variables were restricted to global cognitive ability and episodic memory, and it was not possible to examine systematically whether effect sizes varied across cognitive domains depending on the time between baseline assessment and diagnosis. For example, proponents of the view that episodic memory is the earliest cognitive marker of incipient AD could still argue their case; the present analysis is not informative regarding the onset of decline for different cognitive domains. However, it may be noted that the effect size was larger for global cognitive ability (d = 1.12) than for episodic memory (d = 0.76) for follow-up intervals over 3 years. Note also that those studies that used assessment periods spanning several decades found clear preclinical impairment in domains other than episodic memory, such as linguistic skill (Snowdon et al., 1996) and general intelligence (Whalley et al., 2000).

Another consequence of the limited database is that we were unable to assess the influence of the moderator variables simultaneously. In some cases, moderator variables may be confounded. For example, studies in which participants are preselected on the basis of their cognitive performance may primarily use convenience sampling rather than population-based sampling. Disentangling the effects of these and other variables would require considerably more studies than are available in the current literature. Further, the existing database does not allow for partitioning the length of the follow-up period into shorter time intervals. Future longitudinal studies should address this issue in order to provide more definite information about the specific time at which precipitous cognitive decline normally occurs in preclinical AD. A final concern has to do with the way in which different outcome measures were classified into cognitive domains. Obviously, for certain tasks (e.g., MMSE, word recall) classification is unproblematic, whereas for others (e.g., verbal fluency, Digit Symbol) it is more ambiguous.

Given the present findings, it is interesting to note that there is also emerging evidence of a preclinical period with cognitive deficits in other dementing disorders, such as vascular dementia (e.g., Ingles, Wentzel, Fisk, & Rockwood, 2002; Jonsson Laukka, Jones, Small, Fratiglioni, & Bäckman, 2004), frontotemporal dementia (Geschwind et al., 2001), and Huntington’s disease (Lawrence et al., 1998; Siemers et al., 1996). Thus, preclinical cognitive impairment is not isolated to AD. Measures of cognitive performance may prove useful in the early identification of at-risk individuals over a wide spectrum of dementia diseases, although the specific patterns of impairment may vary across etiologies.

In summary, the current meta-analysis has provided compelling evidence that impairment in multiple cognitive domains several years before clinical diagnosis is characteristic of AD. The generality of the cognitive impairment observed is highly consistent with recent observations that numerous brain structures and functions are affected prior to the AD diagnosis. Finally, we would like to highlight the finding that the magnitude of the preclinical cognitive impairment was affected relatively little by various study and task characteristics.

References


COGNITIVE IMPAIRMENT IN PRECLINICAL AD


Received May 3, 2004
Revision received October 6, 2004
Accepted October 15, 2004