Original Research Article

Dementia and Geriatric Cognitive Disorders

Dement Geriatr Cogn Disord 2009;27:439–450
DOI: 10.1159/000215390

Accepted: January 30, 2009
Published online: April 28, 2009

Minimizing Misdiagnosis: Psychometric Criteria for Possible or Probable Memory Impairment

Brian L. Brooks[a, b] Grant L. Iverson[b, c] Howard H. Feldman[c] James A. Holdnack[d]

[a]Alberta Health Services and University of Calgary, Calgary, Alta., [b]British Columbia Mental Health & Addiction Services, Coquitlam, B.C., [c]University of British Columbia, Vancouver, B.C., Canada; [d]Pearson Assessment, San Antonio, Tex., USA

Key Words

Memory · Assessment · False positive · Guidelines · Alzheimer's disease · Mild cognitive impairment · Wechsler Memory Scale · Misdiagnosis

Abstract

Background/Aims: Memory impairment can be easily misdiagnosed in older adults because obtaining some low scores is common. The objective of the present study is to present new psychometric criteria for determining 'possible' and 'probable' memory impairment. ***Methods:*** We propose criteria based on an analysis of performance from 450 healthy older adults (55–87 years old) on 3 measures from the WMS-III: Logical Memory, Word List, and Visual Reproduction. These measures yield 8 age-adjusted scores for learning, recall, and recognition. The proposed criteria for memory impairment are based on the prevalence of low scores when simultaneously examining all 8 scores and are stratified by current intelligence, estimated premorbid intelligence, and education. The criteria are subsequently validated on 100 healthy older adults and 34 patients with 'possible' or 'probable' Alzheimer's Disease (AD). ***Results:*** Tables with cutoffs and false-positive rates are presented for clinical use. In the validation cohort there were no misclassifications in AD patients. ***Conclusion:*** This study presents steps in the development of proposed psychometric criteria that, in conjunction with clinical judgment, could minimize the misdiagnosis of memory impairment. It is important to reduce misdiagnosis in order to (a) optimize patient care, (b) provide an accurate foundation for identifying biological and neurological markers, and (c) successfully develop disease-modifying treatments. Further validation in a sample of older adults with lesser degrees of cognitive impairment is needed.

Introduction

Accurate assessment of cognition underlies the potential for effective treatment of the dementias. When assessing an older adult for the presence of cognitive impairment, clinicians and researchers must be aware of and guard against its misdiagnosis because this would negatively impact the therapeutic progress. Newly proposed criteria for Alzheimer's disease (AD) expand the definition of the disease state to include milder degrees of cognitive impairment that are supported by biological evidence of the disease. These criteria have a core requirement for memory impairment, which in turn must be correctly identified [1]. Unfortunately, identifying a decline in memory abilities, regardless of the underlying etiology, is particularly prone to false positives when faulty psychometric principles are used.

Iverson and Brooks [2] presented a discussion of 5 psychometric principles to consider when interpreting mul-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1420–8008/09/0275–0439$26.00/0

Accessible online at:
www.karger.com/dem

Brian L. Brooks, PhD
Alberta Children's Hospital, Neurosciences Program, 2888 Shaganappi Trail NW
Calgary, Alta. T3B 6A8 (Canada)
Tel. +1 403 955 2597, Fax +1 403 955 7501
E-Mail brian.brooks@albertahealthservices.ca

tiple test scores. Although any cutoff results in a known low false-positive rate on a single measure (i.e., 16% if the 1-SD cutoff is used or 7% if the 1.5-SD cutoff is used), clinicians and researchers administer and interpret memory tests in combination, not in isolation. These psychometric principles for interpreting multiple test scores are designed to counter the faulty psychometric assumptions utilized in single-score interpretation, which might be employed in daily practice and can potentially lead to a false-positive identification of memory impairment. It is important to consider that, when multiple tests are administered, healthy people obtain some low scores [3–8]. In fact, the more tests that are administered, the more likely it is to obtain a low score [2, 9]. As an example, consider the Neuropsychological Assessment Battery [10]. Iverson et al. [5] reported that 2 or more scores <1 SD are found in 78.4% of healthy adults when considering all 36 scores. However, 2 or more scores <1 SD are found in 51.3% of healthy adults when considering an abbreviated version with 16 scores [2].

Once it is appreciated that (a) low scores are common in healthy people, (b) as the number of tests administered increases, the more likely it is to obtain a low score, and (c) the number of low scores decreases with more stringent cutoff scores (i.e., 2nd percentile), then it is important to consider that there are variables related to the people being tested that can also impact multivariate test interpretation. Perhaps one of the most important psychometric principles is that people with fewer years of education and/or lesser intelligence are expected to have lower performance on cognitive measures [11–15], and thus have more low scores [5]. For example, when considering performance across the 20 scores from the Wechsler Adult Intelligence Scale – Third Edition (WAIS III) Wechsler Memory Scale – Third Edition (WMS-III) battery [16, 17], 73.5% of healthy people with estimated low average intelligence obtain 5 or more scores ≤5th percentile compared to only 5.6% of those with estimated high average intelligence [2].

The psychometric principles for interpreting multiple test scores are applicable to batteries that cover multiple domains of cognitive abilities, as well as batteries of tests that assess specific abilities (e.g., a battery of memory measures). Because healthy older adults may obtain some low memory scores [18] there is substantial concern regarding false-positive diagnosis of memory impairment [19–23]. If memory impairment is misdiagnosed [see ref. 23 for a review], then this drastically limits the accuracy of all subsequent investigations and treatments.

Despite a longstanding interest in identifying memory impairment consistent with very early dementia, there have yet to be psychometrically derived and functionally stratified criteria for memory impairment. One term (along with accompanying criteria) that has gained considerable, but not universal, popularity with clinicians and researchers is mild cognitive impairment (MCI). Petersen et al. [24, 25] defined MCI as being characterized by (a) a subjective memory complaint, (b) an unusually low score on an objective memory measure (based on age only or age and education adjusted normative data), (c) normal general cognitive functioning, (d) normal activities of daily living, and (e) not meeting criteria for dementia. The criterion for an unusually low score has generally been set at 1.5 SDs below the normative mean for healthy older adults. Of course, this cutoff remains somewhat arbitrary and often clinicians and researchers might select a criterion for impairment that ranges anywhere from 1 to 2 SDs below the mean. In addition, the same criterion might be applied to people with varying levels of functioning.

In the recently proposed Working Group criteria [1], a framework for earlier AD diagnosis is specified to include the core criterion of memory impairment with an emphasis on the importance of measuring delayed recall utilizing a memory test paradigm that includes encoding specificity [26–28]. Despite this, the working group authors [1] 'have not defined a magnitude of deficit or the comparative norms that should be utilized' [p. 742]. As a result, clinicians have been left to use clinical judgment pertaining to what constitutes impairment, what tests are most appropriate to identify memory impairment, what types of normative data are best utilized to quantify the impairment based on demographic characteristics, and they do not have clear guidelines for identifying cognitive impairment that are solidly rooted in psychometrics.

The clinical implications for assessing memory functioning in older adults, without considering the psychometric realities of interpreting multiple test scores or considering different levels of functioning, are striking [e.g. 19, 20, 21, 23]. Brooks et al. [23] used standardization data from the WMS-III battery of memory tests to illustrate how easy it is to misdiagnose a healthy older adult as having memory impairment, particularly if he or she has lower functioning. For example, Brooks et al. [23] reported that one or more delayed memory scores ≤5th percentile on the WMS-III battery of memory tests is found in 26.8% of those with 8 or less years of education, in 33.3% of those with 9–11 years of education, in 46.9% of those with borderline intelligence, and in 28.4% of those with low average intelligence. In other words, low scores are common across a battery of memory tests and the chanc-

es of misdiagnosing a healthy older adult as having memory impairment increases with lower functioning. The risk of misdiagnosing memory impairment in older adults is not trivial and is likely reflected in many longitudinal studies with older adults [for a review, see table 1 in ref. 23].

The purpose of this study is to propose psychometrically derived criteria for identifying 'possible' and 'probable' memory impairment. This paper describes the initial steps in the development of these memory criteria and provides a validation study. The foundation for these criteria is premised on the knowledge that healthy older adults do get some low memory scores [e.g. 19, 20, 21, 23]; therefore, criteria must be derived from these psychometric analyses pertaining to what constitutes a clearly abnormal number of low memory scores. The criteria are developed by examining performance across a small battery of memory measures simultaneously in a large sample of healthy older adults, and then validated in a separate group of healthy older adults and a clinical sample of patients with known memory impairment (i.e., 'possible' or 'probable' AD). We hypothesized that: (1) healthy older adults will obtain some low memory scores; (2) the number of low memory scores will increase in those with lesser intelligence or fewer years of education; (3) the false-positive rates, which will be established in a large sample of healthy older adults, will be validated in a separate healthy sample, and (4) all patients with 'possible' or 'probable' AD would meet criteria for memory impairment.

Table 1. Demographic information for 3 samples of older adults

	Development sample	Healthy control validation sample	Clinical validation sample (possible or probable AD)
n	450	100	34
Age	72.8 ± 9.1	72.9 ± 8.6[a]	72.7 ± 7.6[b]
Range	55 – 89	55 – 88	56 – 88
Education	11.7 ± 3.0	11.7 ± 3.0[c]	14.6 ± 2.9[d, *]
Range	7 – 18	7 – 18	7 – 18
Sex			
Male, %	41.1	46.0[e]	55.9[f]
Female, %	58.9	54.0	44.1
Ethnicity			
Caucasian, %	84.7	88.0[g]	88.2[h]
African American, %	9.6	6.0	8.8
Hispanic, %	4.4	4.0	0.0
Other, %	1.3	2.0	2.9
WAIS-III FSIQ	100.1 ± 14.8	100.6 ± 15.7[i]	86.7 ± 13.3)[j, *]
Range	61 – 153	69 – 135	67 – 120
WTAR scores	99.8 ± 15.1	100.5 ± 15.1[k]	103.0 ± 14.0[l]
Range	57 – 126	63 – 126	66 – 122
WTAR-Demo FSIQ	99.5 ± 11.8	100.3 ± 11.5[m]	105.7 ± 12.2[n, *]
Range	69 – 125	74 – 122	76 – 123
MMSE	–	–	22.1 ± 2.5
Range	–	–	19–29
DRS	–	–	119.7 ± 11.6
Range	–	–	97 – 138

Unless otherwise indicated, numbers are mean values ± SD. Pairwise Student's t tests were used for statistical comparisons between the development group and either the healthy control validation sample or the clinical validation sample: [a] $t(548) = 0.10$, $p = 0.92$; [b] $t(482) = 0.06$, $p = 0.95$; [c] $t(548) = 0.00$, $p = 1.0$; [d] $t(482) = 5.45$, $p < 0.001$; [e] $\chi^2(1) = 0.80$, $p = 0.37$; [f] $\chi^2(1) = 2.83$, $p = 0.09$; [g] $\chi^2(1) = 0.72$, $p = 0.40$; [h] $\chi^2(1) = 0.31$, $p = 0.58$; [i] $t(548) = 0.30$, $p = 0.76$; [j] $t(482) = 5.12$, $p < 0.001$; [k] $t(548) = 0.42$, $p = 0.68$; [l] $t(482) = 1.20$, $p = 0.23$; [m] $t(548) = 0.62$, $p = 0.54$; [n] $t(482) = 2.95$, $p = 0.003$. χ^2 analyses for ethnicity involved collapsing all non-Caucasian groups because of small sample sizes. * Indicates that the AD group is significantly different from the 'development' group.

Method

Participants

Three groups of older adults were selected for this study. The first group (the 'development sample'), which was used to derive a proposed set of psychometric criteria, consisted of 450 healthy community-dwelling older adults. The second sample, which was used as the first step in validating the psychometric criteria, consisted of 100 healthy community-dwelling older adults. The third sample, which was used to validate the psychometric criteria, used a clinical sample that consisted of 34 patients clinically diagnosed to have Alzheimer's disease. A summary of the demographics for these three groups is presented in table 1.

Our development sample (n = 450) was derived from the WMS-III [16] standardization sample. This sample was recruited from 28 cities across the United States. The treatment of participants and the collection of data were done in compliance with the Helsinki Declaration. Participants were included if they were medically and psychiatrically healthy, based on a self-report questionnaire. The exclusion criteria included color-blindness, uncorrected hearing loss and/or visual impairment, upper extremity motor problems that might interfere with testing, current treatment for alcohol or drug dependence, current or recent consumption of three or more alcoholic beverages on two or more nights per week, seeking attention from a professional for memory of thinking problems, a history of traumatic brain injury involving loss of consciousness for 5 or more minutes and/or requiring hospitalization for more than 24 h, any medical or psychiatric condition that could potentially impact cognitive functioning (e.g., stroke, epilepsy, brain surgery, encephalitis, meningitis, multiple sclerosis, Parkinson's disease, Huntington's chorea, AD, schizophrenia, or bipolar disorder), or currently receiving any treatment for a medical or psychiatric condition [29].

Our second group was a random sample of 100 healthy older adults selected from the normative group, prior to creating the

psychometric guidelines, to serve as the 'healthy control validation sample'. There were no significant differences between the development group and the healthy control validation group in their age [t(548) = 0.10, p = 0.92], education [t(548) = 0.00, p = 1.0], sex [$\chi^2(1) = 0.80$, p = 0.37], ethnicity [Caucasian vs. non-Caucasian; $\chi^2(1) = 0.72$, p = 0.40], WAIS-III Full Scale IQ [t(548) = 0.30, p = 0.76], WTAR Reading score [t(548) = 0.42, p = 0.68], or WTAR-demographics predicted Full Scale IQ [t(548) = 0.62, p = 0.54].

Our third group, the 'clinical validation sample', included 34 patients with 'possible' or 'probable' AD from the WMS-III validation studies. This AD sample was diagnosed according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimers' Disease and Related Disorders Association (NINCDS-ADRDA) criteria [30], had Mini Mental State Examination (MMSE) scores >18 (mean MMSE = 22.1, SD = 2.5, range = 19–29), Dementia Rating Scale (DRS) scores >95 (mean DRS = 119.7, SD = 11.6, range = 97–138), and were not judged to be depressed (a score of 18 or less on the Beck Depression Inventory or a score of 14 or less on the Geriatric Depression Scale). The details of the AD sample have been previously published in the WMS-III technical manual [29]. Notably, WMS-III data were not utilized within the diagnostic process of the patients with AD. The AD patients were not significantly different from the development sample on age [t(482) = 0.06, p = 0.95], sex [$\chi^2(1) = 2.83$, p = 0.09], ethnicity [$\chi^2(1) = 0.31$, p = 0.58], and WTAR Reading score [t(482) = 1.20, p = 0.23]. However, they did have more years of education [t(482) = 5.45, $p < 0.001$], higher estimated premorbid intellectual abilities [t(482) = 2.95, p = 0.003], and lower 'current' intelligence [t(482) = 5.12, $p < 0.001$]. These findings would suggest that the AD sample had somewhat higher premorbid functioning and had sustained a considerable decline in intellectual abilities.

Measures

The WMS-III [16] is a battery of 6 measures designed to evaluate working memory, learning, immediate recall, delayed recall, and recognition of information. The WMS-III is appropriate for adults aged 16–89 years and was normalized using a stratified, US representative sample of 1,250 healthy adults.

For the present study, only a few WMS-III measures were examined, rather than the entire battery. These selected measures likely represent a typical battery of episodic-memory tests that would be administered to older adults with suspected memory problems. We included Logical Memory (verbal learning and memory of stories), Visual Reproduction (visual learning and memory of designs), and Word List (verbal learning and memory of a word list). These 3 measures produce 8 age-adjusted standard scores (mean = 10, SD = 3) pertaining to immediate recall and learning (i.e., Logical Memory I, Visual Reproduction I, Word List I Total Recall), delayed recall (i.e., Logical Memory II, Visual Reproduction II, Word List II Total Recall), and recognition (i.e., Visual Reproduction Recognition and Word List Recognition).

Current intelligence was measured using the WAIS III [17]. The WAIS-III Full Scale IQ (FSIQ) score is derived from performance on 13 subtests and provides an overall measure of intelligence. The WAIS-III FSIQ has very strong psychometric properties in older adults (i.e., internal consistency reliability, r = 0.97–0.98; test-retest reliability, r = 0.96) [29] and it is the most widely used measure of intelligence in North America [31].

Premorbid intellectual abilities were estimated using the Wechsler Test of Adult Reading (WTAR) [32]. The WTAR is a measure of single-word reading that requires the reading and pronunciation of words with irregular grapheme-to-phoneme translation. The WTAR does not rely on comprehension or knowledge of word meaning, but rather relies on previous learning. An advantage of the reading-recognition paradigm is that it is relatively unaffected by mild damage to the structure or function of the brain [33–35], although it might underestimate premorbid intellectual abilities with more serious forms of neurological injury or disease [36–38]. The WTAR was conormalized with the WAIS-III [17]. Therefore, the WTAR reading score can be combined with demographic variables to estimate intellectual abilities (i.e., WTAR-Demographically predicted Full Scale IQ, with mean = 100 and SD = 15). The WTAR-demographically predicted FSIQ (WTAR-demo FSIQ) was chosen as the measure of predicted premorbid intellectual abilities because (a) it is appropriate for use up to the age of 89 years, (b) the WTAR performance remains relatively stable in the presence of mild cognitive declines, (c) it is a brief test compared to administering the entire WAIS-III, and (d) the WTAR has very strong reliability across the older adult age groups (i.e., internal consistency reliability, r = 0.92–0.95; test-retest reliability, r = 0.94) [29].

Analyses

To derive the proposed psychometric criteria for memory impairment, we examined the prevalence of memory scores, with all 8 normative scores being examined *simultaneously*, in the 'development sample' of healthy older adults (n = 450). Two cutoff scores were included for the analyses: (1) at or below the 16th percentile (i.e., ≤1 SD, scaled score ≤7; mean = 10, SD = 3) and (2) at or below the 5th percentile (i.e., scaled score ≤5; mean = 10, SD = 3). The 16th percentile cutoff was selected so that the new criteria could be applicable to higher-functioning people (i.e., in order to be able to identify a more subtle decline in higher functioning, older adults might require a higher cutoff score). The 5th percentile was also selected because it is approximately equal to the 1.5-SD criterion that has been suggested by Petersen et al. [24, 25].

Our goal was to create criteria for memory impairment that could take into account a range of functioning levels. To achieve this goal, the analyses were stratified based on: (1) current intelligence (i.e., unusually low or borderline, WAIS-III FSIQ <80; low average, WAIS-III FSIQ = 80–89; average, WAIS-III FSIQ = 90–109; high average, WAIS-III FSIQ = 110–119, and superior/very superior, WAIS-III FSIQ = 120+); (2) estimated premorbid intelligence (i.e., unusually low or borderline, WTAR-Demo FSIQ <80; low average, WTAR-Demo FSIQ = 80–89; average, WTAR-Demo FSIQ = 90–109; high average, WTAR-Demo FSIQ = 110–119, and superior/very superior, WTAR-Demo FSIQ = 120+), and (3) highest level of formal education (i.e., 8 years or fewer, 9–11 years, 12 years, 13–15 years, and 16+ years).

Our goal was to establish specificity as close to 0.80 for *possible memory impairment* and as close to 0.90 for *probable memory impairment* as possible. If the number of low scores below a specific cutoff was found in approximately 20% or fewer of healthy older adults, then we determined this was *possible memory impairment*. If the number of low scores below the cutoff was found in approximately 10% or fewer of healthy older adults, then we determined this was *probable memory impairment*. The probable memory impairment designation included those individuals having possible memory impairment. Of course, there is some slight

Table 2. Guidelines for determining memory impairment, based on level of functioning, when considering a cutoff of ≤16th percentile

	n	Memory scores below cutoff		
		broadly normal	possible memory impairment	probable memory impairment
Level of intelligence				
Unusually low (FSIQ ≤79)	40	0–6	7 (12.5%)	8 (5.0%)
Low average (FSIQ = 80–89)	68	0–4	5 (22.1%)	6+ (10.3%)
Average (FSIQ = 90–109)	213	0–2	3 (21.1%)	4+ (9.4%)
High average (FSIQ = 110–119)	83	0–2	3 (13.3%)	4+ (6.0%)
Superior/very superior (FSIQ ≥120)	46	0–1	2 (17.4%)	3+ (6.5%)
Level of estimated premorbid intelligence				
Unusually low (WTAR-FSIQ ≤79)	28	0–6	7 (14.3%)	8 (7.1%)
Low average (WTAR-FSIQ = 80–89)	62	0–5	6 (16.1%)	7+ (1.6%)
Average (WTAR-FSIQ = 90–109)	255	0–3	4 (13.7%)	5+ (7.8%)
High average (WTAR-FSIQ = 110–119)	80	0–2	3 (12.5%)	4+ (7.5%)
Superior/very superior (WTAR-FSIQ ≥120)	16	0	1 (25.0%)	2+ (0.0%)
Years of education				
8 years or less	101	0–4	5–6 (19.8%)	7+ (2.0%)
9–11 years	70	0–4	5 (15.7%)	6+ (7.1%)
12 years	151	0–2	3 (22.5%)	4+ (9.9%)
13–15 years	71	0–2	3 (14.1%)	4+ (7.0%)
16+ years	57	0–1	2–3 (26.3%)	4+ (12.3%)

≤16th percentile (i.e., ≤1 SD) is a scaled score of 7 (mean = 10, SD = 3). The false-positive rates in healthy older adults, which are presented in parentheses, are presumed because the healthy community-dwelling adult sample was not followed longitudinally to determine if some of them were experiencing prodromal AD. Intelligence is based on FSIQ scores from the WAIS-III [22]. Intellectual abilities are estimated using the WTAR-demographics prediction method [32].

variation in the actual specificity rates and these exact values could not always be achieved.

Preliminary validation of the new psychometric criteria established using 450 healthy older adults was carried out on the random sample of 100 healthy adults (i.e., 'healthy validation sample') and the archival sample of patients with AD (i.e., 'clinical validation sample'). The method for evaluating the criteria for memory impairment involved a case-by-case examination of each person's test performance on the 8 age-adjusted scores from the 3 WMS-III tests. This method is reflective of the clinical decision-making model that involves an individual patient's psychometric data. Moreover, unlike a group comparison, this case-by-case method allows us to make use of the 'stratification by functioning' criteria.

Results

The number of low scores required for determining *possible* or *probable memory impairment* was calculated for the 8 WMS-III scores. Based on the analyses with the total sample (n = 450), *possible memory impairment* would be based on having 4–5 scores ≤16th percentile (i.e., found in 18.2% of healthy older adults) or having 2 scores ≤5th percentile (i.e., found in 13.3% of healthy adults). *Probable memory impairment* would be based on having 6 or more scores ≤16th percentile (i.e., found in 6.0% of healthy older adults) or having 3 or more memory scores ≤5th percentile (i.e., found in 5.1% of healthy older adults). Considering the 16th percentile cutoff, having 7 or more low scores is found in 1.6% and having 8 low scores is found in 0.7%. With the 5th percentile cutoff, having 4 or more low scores is found in 2.0% and 6 or more low scores is found in 0.9%. However, we recommend examining the number of low scores stratified by current intelligence, estimated intellectual abilities, and/or by years of education (tables 2, 3), rather than for the entire sample.

Proposed criteria for identifying memory impairment using the 16th percentile as a cutoff are presented in table 2. When stratifying the criteria across levels of intelligence (i.e., WAIS-III FSIQ), the number of low memory scores required to identify probable impairment is 8 in those with unusually low intelligence, 6 or more in those

Table 3. Guidelines for determining memory impairment, based on level of functioning, when considering a cutoff of ≤5th percentile

	n	Memory scores below cutoff		
		broadly normal	possible memory impairment	probable memory impairment
Level of intelligence				
Unusually low (FSIQ ≤79)	40	0–3	4–5 (12.5%)	6+ (7.5%)
Low average (FSIQ = 80–89)	68	0–2	3 (11.8%)	4+ (4.4%)
Average (FSIQ = 90–109)	213	0	1 (18.8%)	2+ (5.6%)
High average (FSIQ = 110–119)	83	0	1 (14.5%)	2+ (7.2%)
Superior/very superior (FSIQ ≥120)	46	0	-	1+ (8.7%)
Level of estimated premorbid intelligence				
Unusually low (WTAR-FSIQ ≤79)	28	0–2	3 (25.0%)	4+ (7.1%)
Low average (WTAR-FSIQ = 80–89)	62	0–2	3 (11.3%)	4+ (4.8%)
Average (WTAR-FSIQ = 90–109)	255	0	1 (27.1%)	2+ (10.6%)
High average (WTAR-FSIQ = 110–119)	80	0	1 (11.3%)	2+ (5.0%)
Superior/very superior (WTAR-FSIQ ≥120)	16	0	–	1+ (6.3%)
Years of education				
8 years or less	101	0–1	2 (19.8%)	3+ (5.0%)
9–11 years	70	0–1	2 (22.9%)	3+ (12.9%)
12 years	151	0	1 (22.5%)	2+ (7.9%)
13–15 years	71	0	1 (22.5%)	2+ (11.3%)
16+ years	57	0	1 (14.0%)	2+ (7.0%)

≤5th percentile is a scaled score of 5 (mean = 10, SD = 3). The false-positive rates in healthy older adults, which are presented in parentheses, are presumed because the healthy community-dwelling adult sample was not followed longitudinally to determine if some of them were experiencing prodromal AD. Intelligence is based on FSIQ scores from the WAIS-III [22]. Intellectual abilities are estimated using the WTAR-demographics prediction method [32].

with low average intelligence, 4 or more in those with average or high average intelligence, and 3 or more in those with superior/very superior intelligence. Across levels of estimated intellectual abilities (i.e., WTAR-Demo FSIQ), the number of low memory scores needed to identify probable memory impairment is 8 or more in those with unusually low estimated intelligence, 7 or more in those with low average estimated intelligence, 5 or more in those with average estimated intellectual abilities, 4 or more in those with high average estimated intelligence, and 2 or more in those with superior/very superior estimated intelligence. When stratified by level of education, the number of memory scores at or below the 16th percentile needed for probable memory impairment is 7 or more in those with 8 years or less, 6 or more in those with 9–11 years, and 4 or more in those with at least high school (i.e., 12, 13–15 and 16+ years).

Proposed criteria for identifying memory impairment using the 5th percentile as a cutoff are presented in table 3. The number of memory scores at or below the 5th percentile, based on various levels of intelligence (i.e., WAIS-III FSIQ), required to identify *probable memory impairment* is 6 or more in those with unusually low intelligence, 4 or more in those with low average intelligence, 2 or more in those with average or high average intelligence, and 1 or more in those with superior/very superior intelligence. Across levels of estimated intellectual abilities (i.e., WTAR-Demo FSIQ), the number of low memory scores needed to identify *probable memory impairment* is 4 or more in those with unusually low or low average estimated intelligence, 2 or more in those with average or high average estimated intellectual abilities, and 1 or more in those with superior/very superior estimated intelligence. When stratified by level of education, the number of low memory scores (i.e., at or below the 5th percentile) needed for *probable memory impairment* is 3 or more in those with fewer than 12 years of education (i.e., 8 years or less, and 9–11 years). In those older adults with at least high school education (i.e., 12, 13–15 and 16+ years), having 2 or more low memory scores is suggestive of *probable memory impairment*.

The new psychometric criteria for determining memory impairment, based on the development sample of 450 healthy older adults, were examined in a random sample of 100 healthy older adults (i.e., 'healthy validation sample'; data not shown). The new criteria were applied on a case-by-case basis (i.e., each person's performance on the WMS-III battery was examined based on the criteria presented in tables 2, 3). Using the 16th percentile cutoff and stratifying by intelligence (i.e., WAIS-III FSIQ), 80% of the sample was classified as being *broadly normal*. *Possible memory impairment* was identified in 20%, with 13% of those people being classified as having *probable memory impairment*. The exact same results were obtained using the criteria stratified by estimated 'premorbid' intelligence (i.e., WTAR-Demo FSIQ). Using the criteria stratified by education and the 16th percentile cutoff, 74% of the sample was classified as being *broadly normal* and 26% was classified as having at least *possible memory impairment*. Of those with memory impairment, 12% were classified as having *probable memory impairment*.

When using the 5th percentile cutoff and considering the criteria stratified by current intelligence (i.e., WAIS-III FSIQ), 16% of the 'healthy validation sample' was classified as having at least *possible memory impairment*, of whom 6% were classified as having *probable memory impairment*. For the criteria stratified by estimated 'premorbid' intelligence (i.e., WTAR-Demo FSIQ), 22% were classified as having at least *possible memory impairment*, of whom 11% were classified as having *probable memory impairment*. Similar results were obtained for the education-stratified criteria, with 23% having at least *possible memory impairment* and 11% of that sample having *probable memory impairment*.

The archival sample of patients with AD was also administered the WMS-III (i.e., 'clinical validation sample'). Mean performance on the 8 selected subtests included: (a) Logical Memory Immediate, mean = 4.3, SD = 2.9, range = 1–13; (b) Logical Memory Delayed, mean = 3.0, SD = 2.5, range = 1–10; (c) Word List Immediate, mean = 4.8, SD = 2.3, range = 1–9; (c) Word List Delayed, mean = 6.4, SD = 1.1, range = 6–12; (e) Word List Recognition, mean = 3.1, SD = 1.4, range = 1–7; (f) Visual Reproduction Immediate, mean = 5.1, SD = 2.2, range = 1–10; (g) Visual Reproduction Delayed, mean = 5.7, SD = 1.8, range = 4–11, and (h) Visual Reproduction Recognition, mean = 6.5, SD = 2.3, range = 2–12. The new psychometric criteria for determining memory impairment (tables 2, 3) were applied on a case-by-case basis to the 34 patients with possible or probable AD (table 4). Similar to the 'healthy validation sample', examination of the criteria in the 'clinical validation sample' was done using the two cutoff scores (i.e., at or below the 16th percentile and at or below the 5th percentile) and based on levels of functioning (i.e., current intelligence, estimated premorbid intelligence, and years of education).

Using the 16th percentile cutoff in this clinical validation sample, 91.2% of patients were identified as having at least *possible memory impairment* [$\chi^2(1) = 54.53$, $p < 0.001$; odds ratio (OR) = 41.3 (95% CI = 12.1–139.4); positive predictive value (PPV) = 0.80 (95% CI = 0.76–0.82); negative predictive value (NPV) = 0.91 (95% CI = 0.79–0.97)] and 82.4% were identified as having *probable memory impairment* [$\chi^2(1) = 57.47$, $p < 0.001$; OR = 31.23 (95% CI = 11.1–87.9); PPV = 0.87 (95% CI = 0.83–0.90); NPV = 0.82 (95% CI = 0.70–0.91)] using the criteria stratified by WAIS-III FSIQ. When considering the criteria stratified by estimated premorbid intelligence (i.e., WTAR-Demo FSIQ), 94.1% had at least *possible memory impairment* [$\chi^2(1) = 58.70$, $p < 0.001$; OR = 64.0 (95% CI = 15.5–259.5); PPV = 0.80 (95% CI = 0.76–0.81); NPV = 0.94 (95% CI = 0.83–0.98)] and 94.1% were identified as having *probable memory impairment* [$\chi^2(1) = 74.85$, $p < 0.001$; OR = 107.1 (95% CI = 24.9–447.7); PPV = 0.87 (95% CI = 0.83–0.88); NPV = 0.94 (95% CI = 0.83–0.98)]. When considering the criteria stratified by years of education, 97.1% had at least *possible memory impairment* [$\chi^2(1) = 52.0$, $p < 0.001$; OR = 93.9 (95% CI = 15.3–563.6); PPV = 0.74 (95% CI = 0.70–0.75); NPV = 0.97 (95% CI = 0.87–0.99)] and 94.1% were identified as having *probable memory impairment* [$\chi^2(1) = 77.58$, $p < 0.001$; OR = 117.3 (95% CI = 27.0–494.3); PPV = 0.88 (95% CI = 0..84–0.89); NPV = 0.94 (95% CI = 0.83–0.98)].

When considering performance at or below the 5th percentile in this 'clinical validation sample', 91.2% of patients were identified as having at least *possible memory impairment* [$\chi^2(1) = 62.97$, $p < 0.001$; OR = 54.3 (95% CI = 15.5–186.3); PPV = 0.84 (95% CI = 0.80–0.86); NPV = 0.91 (95% CI = 0.80–0.97)] and 73.5% were identified as having *probable memory impairment* [$\chi^2(1) = 65.07$, $p < 0.001$; OR = 43.5 (95% CI = 14.4–131.3); PPV = 0.94 (95% CI = 0.90–0.97); NPV = 0.74 (95% CI = 0.62–0.82)] using the criteria stratified by WAIS-III FSIQ. When considering the criteria stratified by estimated premorbid intelligence (i.e., WTAR-Demo FSIQ), 97.1% had at least *possible memory impairment* [$\chi^2(1) = 59.07$, $p < 0.001$; OR = 117.0 (95% CI = 18.9–705.2); PPV = 0.78 (95% CI = 0.75–0.79); NPV = 0.97 (95% CI = 0.87–0.99)] and 94.1% were identified as having *probable memory im-*

Table 4. Number of low memory scores found in sample with possible or probable AD

Age years	Sex	Ethnicity	Intelligence (WAIS-III FSIQ)	Estimated premorbid intelligence (WTAR-Demo FSIQ)	Education, years	Cutoff: ≤16th percentile				Cutoff: ≤5th percentile			
						number of low scores	memory classification, based on level of functioning			number of low scores	memory classification, based on level of functioning		
							WAIS-III criteria	WTAR-Demo FSIQ criteria	education criteria		WAIS-III criteria	WTAR-Demo FSIQ criteria	education criteria
56	M	BL	73	85	14	8	probable	probable	probable	6	probable	probable	probable
61	F	WH	79	95	12	8	probable	probable	probable	6	probable	probable	probable
61	F	WH	76	102	12	8	probable	probable	probable	5	*possible*	probable	probable
62	F	WH	79	88	8	8	probable	probable	probable	7	probable	probable	probable
64	M	WH	88	110	18	7	probable	probable	probable	5	probable	probable	probable
65	M	WH	103	110	18	8	probable	probable	probable	7	probable	probable	probable
65	F	WH	82	115	16	8	probable	probable	probable	6	probable	probable	probable
66	F	WH	78	107	13	7	*possible*	probable	probable	4	*possible*	probable	probable
66	M	WH	76	116	16	5	BN	probable	probable	4	*possible*	probable	probable
67	M	WH	100	109	14	7	probable	probable	probable	5	probable	probable	probable
67	M	WH	88	113	16	8	probable	probable	probable	6	probable	probable	probable
69	M	WH	72	108	14	8	probable	probable	probable	7	probable	probable	probable
70	F	WH	93	106	16	6	probable	probable	probable	3	probable	probable	probable
70	M	WH	113	122	18	8	probable	probable	probable	5	probable	probable	probable
72	M	WH	70	85	12	8	probable	probable	probable	6	probable	probable	probable
74	M	BL	77	76	11	5	BN	BN	*possible*	4	*possible*	probable	probable
74	F	WH	67	88	7	7	*possible*	probable	probable	6	probable	probable	probable
74	M	WH	94	119	18	5	probable	probable	probable	2	probable	probable	probable
75	M	WH	81	105	16	8	probable	probable	probable	6	probable	probable	probable
75	M	WH	93	107	18	8	probable	probable	probable	6	probable	probable	probable
76	M	WH	120	123	16	8	probable	probable	probable	2	probable	probable	probable
77	F	WH	71	87	10	4	BN	BN	BN	2	BN	BN	*possible*
77	F	WH	82	109	12	8	probable	probable	probable	4	probable	probable	probable
78	F	BL	67	87	12	8	probable	probable	probable	5	*possible*	probable	probable
78	F	WH	88	101	16	7	probable	probable	probable	5	probable	probable	probable
78	F	WH	84	108	14	7	probable	probable	probable	6	probable	probable	probable
78	M	WH	89	111	15	6	probable	probable	probable	2	BN	probable	probable
80	M	WH	100	109	18	5	probable	probable	probable	1	*possible*	*possible*	*possible*
80	F	WH	95	116	16	7	probable	probable	probable	6	probable	probable	probable
81	M	WH	87	120	18	8	probable	probable	probable	5	probable	probable	probable
82	F	OT	77	113	14	7	*possible*	probable	probable	3	BN	probable	probable
82	M	WH	108	114	16	6	probable	probable	probable	4	probable	probable	probable
84	F	WH	97	112	14	7	probable	probable	probable	4	probable	probable	probable
88	M	WH	101	117	18	8	probable	probable	probable	2	probable	probable	probable

M = Male; F = female; WH = white, not Hispanic; BL = black; OT = other. Predicted intelligence is based on the WTAR-Demo FSIQ prediction model. These scores are equivalent to Index scores with a mean = 100 (SD = 15). ≤16th percentile (i.e., ≤1 SD) is equal to a scaled score of 7 (mean = 10, SD = 3). ≤5th percentile is equal to a scaled score of 5 (mean = 10, SD = 3). BN = Broadly normal; possible = possible memory impairment; probable = probable memory impairment.

pairment [$\chi^2(1) = 80.44$, $p < 0.001$; OR = 129.5 (95% CI = 29.5–550.3); PPV = 0.89 (95% CI = 0.85–0.90); NPV = 0.94 (95% CI = 0.84–0.98)]. When considering the criteria stratified by years of education, 100% had at least *possible memory impairment* [$\chi^2(1) = 61.54$, $p < 0.001$; PPV = 0.77 (95% CI = 0.74–0.77); NPV = 1.0 (95% CI = 0.91–1.0)] and 94.1% were identified as having *probable memory impairment* [$\chi^2(1) = 80.44$, $p < 0.001$; OR = 129.5 (95% CI = 29.5–550.3); PPV = 0.89 (95% CI = 0.85–0.90); NPV = 0.94 (95% CI = 0.84–0.98)].

Discussion

The purpose of this study was to derive and undertake an initial validation of psychometrically derived criteria for determining possible or probable memory impairment on a commonly used battery of memory measures from the WMS-III. The new criteria were developed using a large sample of healthy older adults, and then validated in healthy older adults and in a sample of patients with possible or probable AD. These psychometric crite-

ria have the potential to be applicable to any neurological or psychiatric disorder that has a known or suspected impact on memory abilities in older adults, although validation in different clinical samples would be necessary before this potential is known. Although each set of diagnostic criteria for dementia and AD requires the presence of memory impairment (i.e., DSM-IV, ICD-10, NINCDS-ADRDA, Dubois Working Group Criteria), they have not attempted to define the psychometric impairment necessary to be considered to have such impaired memory. By deriving a set of empirical criteria, this study may improve the approach to determining the presence of memory impairment in older adults.

When assessing a patient's cognition for the purpose of identifying memory impairment, clinicians will generally administer several measures and simultaneously interpret the results of multiple scores. It is critical to appreciate that the base rate of obtaining a low score on a single measure does not correspond to the results of a battery of tests. Without knowledge of the prevalence of low scores across multiple measures, it becomes easy to overinterpret an isolated low memory score [20, 21, 23] and more likely to misclassify a healthy person as having significant memory impairment or the early stage of a dementia [19, 22]. For example, in the healthy validation sample, 30% would meet criteria for MCI based on having at least one memory score <1.5 SDs. However, when applying the newly proposed criteria in this study, the false-positive rate for memory impairment ranged from 6% when using the intelligence-stratified criteria to 11% when using the estimated-intelligence- and education-stratified criteria.

There are several important aspects to the new criteria for determining memory impairment presented in tables 2 and 3. Beyond being empirically derived based on the performance of a large sample of healthy older adults, the criteria have a known false-positive rate when simultaneously interpreting several memory tests (i.e., this was validated in the healthy sample of 100 older adults). The same information is not transferable to individual memory tests that are not conormalized. Second, criteria for identifying both possible and probable memory impairment are provided, which might be important for identifying subtle (i.e. possible) memory problems associated with mild cognitive impairment, as well as frank (i.e. probable) memory impairment found in patients with fully expressed dementia. Finally, the new criteria are designed to minimize misclassification of memory impairment in those who have lesser intellectual abilities or fewer years of education. Performance on cognitive tests, including memory tests, needs to be interpreted in light of a person's level of functioning [39].

In addition to the risk of misdiagnosing memory problems in lower-functioning persons, there is a substantial risk of missing a diagnosis of memory problems in a person who is higher functioning. If clinicians and researchers employ a universal cutoff for memory impairment, then a person with superior intellectual abilities would have to experience a much greater decline to meet the criterion than a person with low average intellectual abilities. This can have major implications on outcome research. For example, in clinical trials for treatments aimed at the prodromal stage of AD, utilizing a universal cutoff might result in the inclusion of higher-functioning persons who have experienced a decline in memory that is beyond prodromal. A distinct advantage of the newly proposed criteria for memory impairment is that they can also minimize missed classification in those with higher intellectual abilities or more years of education.

The importance of interpreting a person's cognitive performance in light of demographic factors and/or intellectual abilities is further highlighted when considering performance in a subsample of the study participants. For example, in those with average estimated intelligence (n = 255; based on WTAR-Demographics estimated IQ scores), the participants who constituted the *probable memory impairment* group (n = 27) for the ≤5th percentile cutoff were compared to those who did not have probable memory impairment (n = 228). There was no significant difference on age ($p = 0.50$; Cohen's effect size, $d = 0.13$), education ($p = 0.52$; $d = 0.12$), or WMS-III Information/Orientation ($p = 0.49$; $d = 0.13$). Those who met criteria for probable memory impairment had significantly lower intelligence on the WAIS-III ($p < 0.001$; $d = 0.96$). In other words, as illustrated with the primary analyses in this study, persons with lower intelligence are more likely to have low scores compared to those with higher intelligence. Moreover, there appeared to be more ethnic minorities who were identified as having memory impairment compared to those without memory impairment, although the sample sizes were too small for statistical analyses or to draw any definitive conclusions.

A sample of 34 patients with probable AD, and therefore having known memory problems, was used for the preliminary examination of the sensitivity of the criteria. Some important conclusions can be drawn from this examination of the validity of the criteria in this clinical sample. First, the overall classification accuracy was nearly equivalent when using either cutoff score (i.e., 16th percentile vs. 5th percentile). Despite the similarities,

there is a possible advantage to using a more lenient cutoff score (i.e., 16th percentile) for those individuals who are higher functioning because it requires a less significant decline before a clinician will deem that there is an objective cognitive marker of memory problems. Second, in these patients who have experienced a substantial decline in overall functioning (i.e., in this sample of patients with possible or probable AD, there is a 19-point difference between mean premorbid intelligence and mean current intelligence), it becomes less appropriate to use criteria based on current functioning and more appropriate to use criteria based on estimated premorbid functioning. As a result, it is recommended that premorbid functioning (i.e., premorbid intelligence and/or highest level of education) be considered in those patients who have experienced a substantial decline in cognitive abilities. Third, when considering performance on the education-stratified criteria, none of the patients with possible or probable AD were identified as having broadly normal memory ability impairment across the battery of memory tests (i.e., 100% were identified as having possible and 94.1% were identified as having probable memory impairment). When comparing the AD sample to the validation sample, patients with possible or probable AD were 129 times more likely to meet criteria for probable memory impairment. This is important information, albeit perhaps not overly surprising, as part of the clinical validation of these criteria. Clinicians can have reasonable confidence in these criteria for possible or probable memory impairment when assessing older adults for the presence of possible or probable AD, because the criteria are rooted in psychometrics. Of course, these criteria are only one tool used in the diagnostic process of identifying memory impairment and will be most beneficial when used to supplement clinical judgment.

This newly proposed methodology (i.e., simultaneously interpreting all memory test scores) has the distinct benefit of lower false-positive rates compared to interpreting each test from a battery in isolation. For example, the false-positive rate for an individual test score, when considering at or below the 5th percentile as the cutoff, is 5%. However, when considering 8 test scores using at or below the 5th percentile as the cutoff, the false-positive rate in a healthy standardization sample for 1 low score increases to 28%. Several studies have demonstrated that healthy older adults obtain some low memory scores when a battery of tests is administered [19–21, 23] and the chances of making a false-positive diagnosis increase with lesser education and intellectual abilities [21, 23]. It is well established that a substantial minority of patients identified as having MCI do not progress to AD within 3–5 years [see ref. 23 for an overview]. Their memory impairment normalizes when they are retested at a later date. Thus, the patients either had a reversible form of memory impairment, or they were psychometrically misdiagnosed (due to measurement error, failure to use the most appropriate normative data, or faulty psychometric decision rules). The potential factors contributing to the low rates of progression to AD and the failure of medications to delay the time to diagnosis of AD in several large-scale trials remain uncertain; however, the emphasis on a single test performance without broader psychometric considerations may have had an impact [40–45].

The recently published consensus-based research criteria for diagnosing probable AD include self- or informant-reported decline in memory abilities plus 'significant episodic memory impairment' as the core diagnostic criteria [1]. The authors emphasized the importance of measuring delayed recall utilizing a memory test paradigm that includes encoding specificity (i.e., cueing and/or recognition paradigms). The 3 memory tests used in this study have immediate recall, delayed recall, and recognition memory components (although normative scores are available for only 2 of the recognition measures). Memory tests for stories [46–48], designs [46, 49], and word lists [47, 50–52] are widely used in AD research. Given that these WMS-III tests are co-normalized, they seem particularly useful in clinical practice and research involving patients with prodromal AD.

There are clear limitations to this study. First, it is not known whether some of the participants in the standardization sample had prodromal AD. Although these people are screened for such medical problems, this method of exclusion is not infallible. Having some people with prodromal AD in the normative data would, unfortunately, dilute the effectiveness of the criteria [53]. This problem is, of course, universal to all normative samples for cognitive tests. Second, the AD sample is archival and additional information (e.g., biological markers, structural imaging, and CSF analyses) is not available. Additional research with those that have suspected prodromal AD will be needed to determine if the subtle changes in memory abilities can be identified using the possible memory impairment criteria. Third, it is not knowable, with these archival data, how early in the disease process memory impairment can be identified using these new criteria. Fourth, the level of premorbid functioning for the AD sample is somewhat high (i.e., only 8 patients with AD had premorbid intelligence below the population mean and only 3 had less than high-school education),

which limits the generalizability of these preliminary findings to lower-functioning patients with suspected AD. Finally, the sample size in the healthy older adults with superior/very superior predicted premorbid intellectual abilities is small (n = 16), which might have an impact on the stability of the findings if a larger group was available for these analyses.

The new consensus-based research criteria for AD emphasized the need, through future research, to operationalize the proposed diagnostic criteria for memory impairment, structural imaging, molecular imaging, and CSF markers [1]. More recently, Petersen and Negash [54, p. 51] indicated that 'further refinements of the criteria [for MCI] and prediction techniques may be necessary for prognosticating the outcomes'. As part of operationalizing and refining the criteria for memory impairment, several steps must be undertaken. First, 'normal' memory performance in healthy older adults across a battery of measures must be studied and well understood. Second, the criteria for memory impairment must be derived from a large sample of healthy older adults. Third, the new criteria must be validated in patients with known memory impairment. Finally, the criteria can be applied to older adults with suspected prodromal AD or amnestic MCI. The present study represents the first three steps. Future longitudinal research is needed to examine the predictive accuracy of the proposed memory impairment criteria for identifying true prodromal AD.

Acknowledgements

The authors thank Jennifer Bernardo and Tatiana Ramirez-Aponte for assistance with manuscript preparation. Portions of this work were presented at the annual conference of the National Academy of Neuropsychology in Scottsdale, Ariz., USA (November 2007). The data in tables 1–4 are original and produced by special permission of the publisher, Pearson Assessment. Statistical analyses for this paper were conceptualized by Drs. Brooks and Iverson, and were conducted by Dr. Holdnack.

Disclosure Statement

J.A.H. is a Senior Research Director with Pearson Assessment, which is the publisher of the WAIS-III, WMS-III, and WTAR.

References

1 Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rosser M, Salloway S, Stern Y, Visser PJ, Scheltens P: Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria – a position paper. Lancet Neurol 2007;6:734–746.

2 Iverson GL, Brooks BL: Improving accuracy for identifying cognitive impairment; in Schoenberg MR, Scott JG (eds): The Black Book of Neuropsychology: A Syndrome-Based Approach. New York, Springer, in press.

3 Heaton RK, Miller SW, Taylor MJ, Grant I: Revised comprehensive norms for an expanded Halstead-Reitan battery: demographically adjusted neuropsychological norms for African American and Caucasian adults professional manual. Lutz, Psychological Assessment Resources, 2004.

4 Heaton RK, Grant I, Matthews CG: Comprehensive norms for an extended Halstead-Reitan battery: demographic corrections, research findings, and clinical applications. Odessa, Psychological Assessment Resources, 1991.

5 Iverson GL, Brooks BL, White T, Stern RA: Neuropsychological assessment battery (NAB): introduction and advanced interpretation; in Horton AM Jr, Wedding D (eds): The Neuropsychology Handbook. New York, Springer, 2008, pp 279–343.

6 Crawford JR, Garthwaite PH, Gault CB: Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: a generic method with applications. Neuropsychology 2007;21:419–430.

7 Axelrod BN, Wall JR: Expectancy of impaired neuropsychological test scores in a non-clinical sample. Int J Neurosci 2007;117:1591–1602.

8 Schretlen DJ, Testa SM, Winicki JM, Pearlson GD, Gordon B: Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. J Int Neuropsychol Soc 2008;14:436–445.

9 Binder LM, Iverson GL, Brooks BL: To err is human: 'abnormal' neuropsychological scores and variability are common in healthy adults. Arch Clin Neuropsychol 2009;24:31–46.

10 Stern RA, White T: Neuropsychological assessment battery. Lutz, Psychological Assessment Resources, 2003.

11 Horton AM Jr: Above-average intelligence and neuropsychological test score performance. Int J Neurosci 1999;99:221–231.

12 Warner MH, Ernst J, Townes BD, Peel J, Preston M: Relationships between IQ and neuropsychological measures in neuropsychiatric populations: within-laboratory and cross-cultural replications using WAIS and WAIS-R. J Clin Exp Neuropsychol 1987;9:545–562.

13 Tremont G, Hoffman RG, Scott JG, Adams RL: Effect of intellectual level on neuropsychological test performance: a response to Dodrill (1997). Clin Neuropsychol 1998;12:560–567.

14 Steinberg BA, Bieliauskas LA, Smith GE, Ivnik RJ, Malec JF: Mayo's older Americans normative studies: age- and IQ-adjusted norms for the Auditory Verbal Learning Test and the Visual Spatial Learning Test. Clin Neuropsychol 2005;19:464–523.

15 Steinberg BA, Bieliauskas LA, Smith GE, Ivnik RJ: Mayo's older Americans normative studies: age- and IQ-adjusted norms for the Trail-Making Test, the Stroop test, and MAE Controlled Oral Word Association Test. Clin Neuropsychol 2005;19:329–377.

16 Wechsler D: Wechsler Memory Scale, ed. 3 San Antonio, Psychological Corporation, 1997.

17 Wechsler D: Wechsler Adult Intelligence Scale, ed 3. San Antonio, Psychological Corporation, 1997.
18 Blackford RC, LaRue A: Criteria for diagnosing age-associated memory impairment: proposed improvement from the field. Dev Neuropsychol 1989;5:295–306.
19 de Rotrou J, Wenisch E, Chausson C, Dray F, Faucounau V, Rigaud AS: Accidental MCI in healthy subjects: a prospective longitudinal study. Eur J Neurol 2005;12:879–885.
20 Palmer BW, Boone KB, Lesser IM, Wohl MA: Base rates of 'impaired' neuropsychological test performance among healthy older adults. Arch Clin Neuropsychol 1998;13:503–511.
21 Brooks BL, Iverson GL, White T: Substantial risk of 'accidental MCI' in healthy older adults: base rates of low memory scores in neuropsychological assessment. J Int Neuropsychol Soc 2007;13:490–500.
22 Loewenstein DA, Acevedo A, Ownby R, Agron J, Barker WW, Isaacson R, Strauman S, Duara R: Using different memory cutoffs to assess mild cognitive impairment. Am J Geriatr Psychiatry 2006;14:911–919.
23 Brooks BL, Iverson GL, Holdnack JA, Feldman HH: The potential for misclassification of mild cognitive impairment: a study of memory scores on the Wechsler Memory Scale-III in healthy older adults. J Int Neuropsychol Soc 2008;14:463–478.
24 Petersen RC, Smith GE, Ivnik RJ, Kokmen E, Tangalos EG: Memory function in very early Alzheimer's disease. Neurology 1994;44:867–872.
25 Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E: Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–308.
26 Tulving E, Thompson D: Encoding specificity and retrieval processes in episodic memory. Psychol Rev 1973;80:352–373.
27 Grober E, Buschke H: Genuine memory deficits in dementia. Dev Neuropsychol 1987;3:13–36.
28 Grober E, Kawas C: Learning and retention in preclinical and early Alzheimer's disease. Psychol Aging 1997;12:183–188.
29 The Psychological Corporation: WAIS-III/WMS-III technical manual. San Antonio, Psychological Corporation, 1997.
30 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984;34:939–944.
31 Rabin LA, Barr WB, Burton LA: Assessment practices of clinical neuropsychologists in the United States and Canada: a survey of INS, NAN, and APA division 40 members. Arch Clin Neuropsychol 2005;20:33–65.
32 The Psychological Corporation: Wechsler Test of Adult Reading manual. San Antonio, Psychological Corporation, 2001.
33 Strauss E, Sherman EMS, Spreen O: A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, ed 3. New York, Oxford University Press, 2006.
34 Bright P, Jaldow E, Kopelman MD: The national adult reading test as a measure of premorbid intelligence: a comparison with estimates derived from demographic variables. J Int Neuropsychol Soc 2002;8:847–854.
35 Maddrey AM, Cullum CM, Weiner MF, Filley CM: Premorbid intelligence estimation and level of dementia in Alzheimer's disease. J Int Neuropsychol Soc 1996;2:551–555.
36 McFarlane J, Welch J, Rodgers J: Severity of Alzheimer's disease and effect on premorbid measures of intelligence. Br J Clin Psychol 2006;45:453–463.
37 O'Carroll RE, Prentice N, Murray C, van Beck M, Ebmeier KP, Goodwin GM: Further evidence that reading ability is not preserved in Alzheimer's disease. Br J Psychiatry 1995;167:659–662.
38 Fromm D, Holland AL, Nebes RD, Oakley MA: A longitudinal study of word-reading ability in Alzheimer's disease: evidence from the national adult reading test. Cortex 1991;27:367–376.
39 Steinberg BA, Bieliauskas LA, Smith GE, Ivnik RJ: Mayo's older Americans normative studies: age- and IQ-adjusted norms for the Wechsler Memory Scale – Revised. Clin Neuropsychol 2005;19:378–463.
40 Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, Galasko D, Jin S, Kaye J, Levey A, Pfeiffer E, Sano M, van Dyck CH, Thal LJ: Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 2005;352:2379–2388.
41 Thal LJ, Ferris SH, Kirby L, Block GA, Lines CR, Yuen E, Assaid C, Nessly ML, Norman BA, Baranak CC, Reines SA: A randomized, double-blind, study of rofecoxib in patients with mild cognitive impairment. Neuropsychopharmacology 2005;30:1204–1215.
42 Feldman HH, Ferris S, Winblad B, Sfikas N, Mancione L, He Y, Tekin S, Burns A, Cummings J, del Ser T, Inzitari D, Orgogozo JM, Sauer H, Scheltens P, Scarpini E, Herrmann N, Farlow M, Potkin S, Charles HC, Fox NC, Lane R: Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. Lancet Neurol 2007;6:501–512.
43 Birks J, Flicker L: Donepezil for mild cognitive impairment. Cochrane Database Syst Rev 2006;3:CD006104.
44 Birks J, Flicker L: Selegiline for Alzheimer's disease. Cochrane Database Syst Rev 2003;1:CD000442.
45 Salloway S, Ferris S, Kluger A, Goldman R, Griesing T, Kumar D, Richardson S: Efficacy of donepezil in mild cognitive impairment: a randomized placebo-controlled trial. Neurology 2004;63:651–657.
46 Hutchinson AD, Mathias JL: Neuropsychological deficits in frontotemporal dementia and Alzheimer's disease: a meta-analytic review. J Neurol Neurosurg Psychiatry 2007;78:917–928.
47 Hamilton JM, Salmon DP, Galasko D, Delis DC, Hansen LA, Masliah E, Thomas RG, Thal LJ: A comparison of episodic memory deficits in neuropathologically confirmed dementia with Lewy bodies and Alzheimer's disease. J Int Neuropsychol Soc 2004;10:689–697.
48 O'Donnell BF, Drachman DA, Lew RA, Swearer JM: Measuring dementia: assessment of multiple deficit domains. J Clin Psychol 1988;44:916–923.
49 Griffith HR, Netson KL, Harrell LE, Zamrini EY, Brockington JC, Marson DC: Amnestic mild cognitive impairment: diagnostic outcomes and clinical prediction over a two-year time period. J Int Neuropsychol Soc 2006;12:166–175.
50 Greenaway MC, Lacritz LH, Binegar D, Weiner MF, Lipton A, Munro Cullum C: Patterns of verbal memory performance in mild cognitive impairment, Alzheimer disease, and normal aging. Cogn Behav Neurol 2006;19:79–84.
51 Delis DC, Wetter SR, Jacobson MW, Peavy G, Hamilton J, Gongvatana A, Kramer JH, Bondi MW, Corey-Bloom J, Salmon DP: Recall discriminability: utility of a new CVLT-II measure in the differential diagnosis of dementia. J Int Neuropsychol Soc 2005;11:708–715.
52 Lacritz LH, Cullum CM, Weiner MF, Rosenberg RN: Comparison of the Hopkins Verbal Learning Test-Revised to the California Verbal Learning Test in Alzheimer's disease. Appl Neuropsychol 2001;8:180–184.
53 Ritchie LJ, Frerichs RJ, Tuokko H: Effective normative samples for the detection of cognitive impairment in older adults. Clin Neuropsychol 2007;21:863–874.
54 Petersen RC, Negash S: Mild cognitive impairment: an overview. CNS Spectr 2008;13:45–53.