ELSEVIER

Alzheimer's & Dementia 5 (2009) 295–306

Alzheimer's & Dementia

Development and validation of the Memory Performance Index: Reducing measurement error in recall tests

William R. Shankle*, Tushar Mangrola, Timothy Chan, Junko Hara

Medical Care Corporation, Irvine, CA, USA

Abstract

Background: The Memory Performance Index (MPI) quantifies the pattern of recalled and nonrecalled words of the Consortium to Establish a Registry for Alzheimer's Disease Wordlist (CWL) onto a 0 to 100 scale and distinguishes normal from mild cognitive impairment with 96% to 97% accuracy.

Methods: In group A, 121,481 independently living individuals, 18 to 106 years old, were assessed with the CWL and classified as cognitively impaired (N = 5,971) or normal (N = 115,510). The MPI and CWL immediate free recall (IFR), delayed free recall (DFR), and total free recall (TFR) scores (the outcome measures) were each regressed against predictors of age, gender, race, education, test administration method (in-person or telephone), and wordlist used. Predictor effect sizes (Cohen's f^2) were computed for each outcome. In addition, CWL plus Functional Assessment Staging Tests (FAST) were administered to 441 normal to moderately severely demented (FAST stages 1 to 6) patients (group B). Median MPI scores were tested for significant differences across FAST stage.

Results: For group A, the variance explained by all predictors combined was MPI = 55.0%, IFR = 24.9%, DFR = 23.4%, and TFR = 26.9%. The age effect size on MPI score was large, but it was small on IFR, DFR, and TFR. The other predictors all had negligible (<0.02) or small effect sizes (0.02 to 0.15). For group B, median MPI scores progressively declined across all FAST stages ($P < .0002$).

Conclusions: MPI score progressively declines with increasing dementia severity. Also, MPI score explains 2 to 3 times more variance than total scores, which improves ability to detect treatment effects.

Keywords: Alzheimer's disease; Mild cognitive impairment; Dementia; Accuracy; Normal aging; Episodic memory; Declarative memory; Short-term memory; Hippocampus; Working memory; Correspondence analysis; Receiver operating characteristic; Logistic regression

1. Introduction

1.1. Consortium to Establish a Registry for Alzheimer's Disease Wordlist Recall Test

The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Wordlist Recall Test (CWL) is a standardized, well-validated assessment of immediate free recall (IFR) and delayed free recall (DFR) developed in the 1980s by the National Institute of Aging Alzheimer's Disease Centers [1,2]. The traditional scoring uses a cutoff score based on the number of words recalled during the three learning trials or during the DFR trial, for which adjustment might or might not be made to account for the age of the subject. A study of community-based versus university subject samples showed that normal aging subjects recalled a mean of 19.5 ± 5.0 to 21.4 ± 4.4 words during the three CWL learning trials and recalled a mean of 6.0 ± 2.8 to 7.5 ± 2.0 words on DFR [3]. One of the few studies that included CWL cutoff scores for patients with mild cognitive impairment (MCI) was one involving the Finnish version of the CERAD neuropsychological test battery [4]. In that study, the CERAD battery was administered to patients with normal aging, amnestic MCI (aMCI), and mild dementia caused by Alzheimer's disease (AD). The authors found that the CWL test gave the best discrimination for these patient groups and reported optimal cutoff scores for the sum of the three learning trials (16/30) and for DFR trial (6/10), which gave respective sensitivities of 0.33 and 0.33 for aMCI and of 0.6 and 0.86 for mild AD

*Corresponding author. Tel.: 949-838-0154; Fax: 949-838-0153.
E-mail address: rshankle@mccare.com

1552-5260/09/$ – see front matter © 2009 The Alzheimer's Association. All rights reserved.
doi:10.1016/j.jalz.2008.11.001

dementia, with specificities of 0.93 and 1.0 for normal aging. The authors suggested increasing the number of list words or the time between learning and DFR to further improve detection of aMCI cases.

1.2. Improving the CWL scoring with recall pattern analyses

An alternative way to improve detection of MCI is to use more of the available information in the CWL test. When one considers that there are at least 1 trillion possible patterns for recalling 10 words across four trials, the reliance on total scores for learning and DFR ignores almost all of the available information.

We have previously introduced a mathematical algorithm that measures the pattern of both recalled and nonrecalled words across the four CWL trials and classifies the pattern as cognitively normal or impaired on the basis of a cut point that characterizes sensitivity and specificity levels appropriate to the requirements of the clinical setting [5]. On the basis of nonparametric receiver operating characteristic curve determination of overall accuracy, this algorithm discriminates normal aging from MCI by 96% to 97% and discriminates normal aging from mild dementia with 99% accuracy [5–7]. The inter-rater reliability was 0.83, and the diagnostic validity for patients in Functional Assessment Staging Tests (FAST) stages 1 to 4 has a kappa value of 0.92 ± 0.09 [6].

The algorithm's parameters were originally derived from an analysis of a sample of 471 well-characterized subjects who had no cognitive or functional impairment, had MCI, or had mild dementia [5]. For the majority of the impaired subjects, underlying causes included AD, Lewy body disease, frontotemporal lobe disease, and cerebrovascular disease. A small number of subjects had other dementing disorders.

The score produced by this algorithm measures characteristics that are not captured by total scores of the numbers of correctly recalled words [5]. These characteristics include differential effects on recall difficulty as a function of (1) a word's position in the learning list, (2) the number of times a subject has been exposed to the word, (3) the delay between learning the list and recalling its words, and (4) the patterns of recall across the learning and testing trials that are unique to persons with no cognitive impairment, MCI, or mild dementia. In addition, the algorithm's score also measures the effect of not recalling a word in a given trial. This effect is also influenced by word position, frequency of exposure, and delay between exposure and recall.

For example, Fig. 1 in our previous publication [5] shows that the first (w1) and seventh to tenth words (w7 to w10) in the learning list are easier to recall than words second to sixth (w2 to w6), which is consistent with well-known effects of primacy and recency. Also, for a given list word such as *actor*, immediate recall is hard on trial 1, easier on trial 2, easiest on trial 3, and then is hardest after a several minute delay. Consider a real example of similarly aged subjects who recalled 6 of 10 on IFR trial 3 and 4 of 10 on DFR trial 4. The aforementioned algorithm classified one subject as impaired and the other as normal. The orders in which they recalled list words w1 to w10 on trial 4 were the following:

w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	Classification
0	0	0	0	0	0	2	4	1	3	Impaired
1	0	4	3	0	2	0	0	0	0	Normal

In spite of identical recall totals on trials 3 and 4, the patterns of recall for the normal and impaired subjects differed, with the impaired subject recalling the most recently exposed words from the end of the list and the normal subject recalling words more closely to the order in which they were presented. Howard and Kahana [8] have shown that recall order approximates the order of presented stimuli when the stimuli cannot be easily associated. This low associability among words is the case for the wordlists used in the CWL and in the MCI Screen (MCIS), a web-based implementation of the CWL that uses the aforementioned scoring algorithm.

1.3. Quantifying wordlist recall pattern

Although a classification of the pattern of recalled and not recalled words is useful to discriminate between healthy subjects and those with some underlying cause of progressive memory loss, a quantification of such a pattern offers a more useful and intuitive understanding of overall memory function and might provide more precise measurement of longitudinal change.

The present article presents the Memory Performance Index (MPI), a scale from 0 to 100 with cut point centered at 50, which quantifies a subject's pattern of recalled and not recalled words. Through a monotonic transformation, the MPI scale provides a more useful interpretation of the results produced by the CWL scoring algorithm. The present article also analyzed the amount of variance of the MPI score and of the total numbers of words correctly recalled immediately or after a delay that can be accounted for by typical sources of variability such as age, gender, race, education, test administration method, and test stimuli. For this purpose, a sample of 121,481 long-term care (LTC) insurance applicants aged 18 to 106 years old were analyzed. Finally, the relation between MPI score and dementia severity was analyzed by using a well-characterized clinical sample of 441 cognitively normal to moderately severely demented patients.

2. Methods

2.1. Development of the MPI

The MCIS is a web-based assessment tool that guides examiners to reliably administer the CWL test plus additional measures of executive function [6]. The CWL scoring algorithm has been adopted as a cognitive measure in academic, clinical, disease management, and insurance settings.