The Diagnostic Value of Controlled Oral Word Association Test-FAS and Category Fluency in Single-Domain Amnestic Mild Cognitive Impairment

Michael Malek-Ahmadi a, c  Brent J. Small b  Ashok Raj a

a University of South Florida Byrd Alzheimer’s Institute and b University of South Florida School of Aging Studies, Tampa, Fla., and c Banner Sun Health Research Institute, Sun City, Ariz., USA

Key Words
Prodromal Alzheimer’s disease · Alzheimer’s disease · Dementia · Neuropsychology · Mild cognitive impairment · Verbal fluency

Abstract
Background: Recent studies have shown that decreases in both letter fluency and category fluency may be present in addition to memory impairment in single-domain amnestic mild cognitive impairment (aMCI). However, the clinical utility of these fluency measures is unclear. The aim of this study was to determine what, if any, diagnostic value letter and category fluency provide in differentiating single-domain aMCI from normal cognition. Methods: Data from 66 individuals [33 cognitively normal (CN) and 33 aMCI] between the ages of 66 and 87 years participating in the Florida Alzheimer’s Disease Research Center were compared on the Controlled Oral Word Association Test (COWAT)-FAS and Category Fluency test, both in terms of raw and scaled scores. Results: Participants were matched on age, education and sex. Two-tailed independent sample t-tests found statistically significant differences between the CN and aMCI groups for both raw and scaled scores of COWAT-FAS and Category Fluency (p < 0.001). Logistic regression analyses found that COWAT-FAS and Category Fluency did not significantly improve diagnostic accuracy when combined with the Hopkins Verbal Learning Test-Revised delayed recall. Conclusion: Although decreased COWAT-FAS and Category Fluency performance may be present in single-domain aMCI, these tests do not improve the ability of the Hopkins Verbal Learning Test-Revised delayed recall to differentiate aMCI from CN individuals.

Introduction
Since amnestic mild cognitive impairment (aMCI) is thought to be prodromal Alzheimer’s disease (AD) [1, 2], accurately identifying these individuals early in the disease process has become a priority in order to achieve better clinical outcomes. Given the increasing interest in aMCI as a therapeutic target [3], utilizing neuropsychological measures with additional discriminatory power might help identify individuals in earlier stages of AD.

As a diagnostic entity, aMCI was first characterized as a syndrome consisting of decreased memory performance at or below 1.5 standard deviations (SD) on age- and education-adjusted normative values on a verbal memory test with the inclusion of subjective memory complaints by the affected individual [1]. However, the
diagnostic criteria for MCI have been refined to differentiate between aMCI and non-amnestic MCI, with the latter showing performance at or below 1.5 SD on one or more tests in one or more domains other than memory. Both entities can be further classified as single- or multiple-domain MCI depending upon the number of cognitive domains that demonstrate test performance at or below 1.5 SD [4]. In addition to these criteria, aMCI is also characterized by a lack of impairment in daily functioning. However, recent studies have suggested that subtle declines in daily activity functioning may be present in individuals with aMCI [5–8].

Recent studies have demonstrated that decreased non-memory domain performance may be characteristic of aMCI despite the fact that this performance may still fall within the currently defined normal limits [9–11]. Other studies have demonstrated that decreases in semantic fluency performance (e.g. animal naming) may be indicative of single-domain aMCI [12–15]. Some of these studies also found statistically significant decreases in letter fluency performance among those categorized as single-domain aMCI [9, 13, 14]. Although statistically significant differences in letter and semantic fluency were found in these studies, the extent to which these measures add diagnostic value in identifying single-domain aMCI was not determined. Lam et al. [15] state that semantic fluency does provide some degree of diagnostic value, but whether or not letter fluency adds diagnostic value in identifying aMCI is yet to be seen.

The Controlled Oral Word Association Test (COWAT)-FAS and Category Fluency test have been used as measures of both language [16, 17] and executive function [18, 19] domains. There is no objective consensus on which interpretation is correct, as cognitive domains are not always mutually exclusive constructs, with the consequence that it is difficult to state that a cognitive test is a pure measure of a particular domain. Previous research [20, 21] has demonstrated that aMCI individuals have decreased semantic processing when compared to cognitively normal (CN) individuals. These findings suggest that cognitive tests which rely more heavily on semantic networks might be able to more accurately differentiate aMCI individuals from CN individuals. This is thought to be a result of AD pathology disrupting neural networks connecting the prefrontal cortex and temporal lobe that help mediate executive functions [21]. The COWAT-FAS and Category Fluency test appear to utilize both semantic and executive function resources, so it is possible that they may be able to elucidate more subtle cognitive differences between aMCI and CN individuals.

The aim of the current study will be to determine what diagnostic value the COWAT-FAS and Category Fluency test may provide in differentiating single-domain aMCI from normal cognition.

Materials and Methods

Study Sample
Data from 66 (33 CN and 33 aMCI) participants between the ages of 65 and 87 years from the Florida Alzheimer’s Disease Research Center were used in the analysis. Each aMCI individual was matched on sex, education and age to a CN individual. Informed consent was obtained from all participants, and research was conducted in compliance with institutional regulations. All participants completed a comprehensive evaluation including full clinical history, neurologic examination, informant-based interview, clinical laboratory tests, magnetic resonance imaging of the brain, and a full neuropsychological battery. Consensus diagnoses with multiple clinicians were performed on all participants. The aMCI group included participants whose memory test scores were at or below an age-adjusted scaled score of 5 on the Wechsler Memory Scale- Revised (WMS-R) Logical Memory test. In addition, aMCI cases all had a Clinical Dementia Rating [22] global score of 0.5. CN participants were diagnosed as such based on an informant interview in which no decline in cognition was reported. Furthermore, all CN participants had a scaled score greater than 5 on all cognitive tests and also had a global score of 0 on the Clinical Dementia Rating.

Individuals with a history of stroke or other cerebrovascular event were excluded from the analysis. In addition, individuals with a significant history of psychiatric illness, such as bipolar disorder and schizophrenia, were excluded. Individuals with a history of significant cognitive problems due to brain injury, tumor or other medical conditions were also excluded, as well as those whose native language was not English.

Neuropsychological Tests
WMS-R Logical Memory. A short fictional story is read to the participants after which they are asked to repeat as much of the story as they can remember. After a 20-min delay, participants are asked to recall the story [23].

Hopkins Verbal Learning Test-Revised. A list of 12 words is read aloud to the participant after which he/she is asked to recall as many of the words as possible. This is done 3 times for the immediate recall section. Delayed recall occurs after a 20- to 25-min delay in which the participant is asked to recall the words from the list. Version 1 of the Hopkins Verbal Learning Test-Revised (HVLT-R) was used for this study [24].

Trails A. The participant is instructed to trace a line that connects circled numbers in consecutive order [25].

Trails B. The participant is asked to trace a line that connects circled numbers and circled letters in consecutive order while alternating between numbers and letters (1 – A – 2 – B – 3 – C, and so on) [25].

WMS-R Digit Symbol. The examinee is shown a series of numbers, each corresponding to a unique symbol. The examinee is then instructed to write down the symbol that corresponds to each number in an empty box placed below the number [26].
COWAT-FAS. This test requires the individual to name as many words as possible that begin with a given letter, i.e. F, A and S. Sixty seconds are allotted for each letter. Individuals cannot use proper names or numbers and cannot use words with different tenses or endings once the root word has been given [27].

Category Fluency Test (Animals, Vegetables, Fruits). The individual is asked to name as many items as possible in a given category (animals, fruits and vegetables). Sixty seconds are allotted for each category [27].

Boston Naming Test. A collection of 60 drawings of objects is shown individually to the participant who is asked to verbally identify what the object is [28].

Age-adjusted normative scores for Trails A, Trails B and the Boston Naming Test were derived from Ivnik et al. [29]. Age-adjusted normative scores for Logical Memory and WAIS-R digit symbol were derived from Ivnik et al. [30]. Category Fluency scaled scores were derived from Lucas et al. [31]. COWAT-FAS z-scores were derived from Loonstra et al. [32]. z-scores for COWAT-FAS were converted into scaled scores using a standard normative score conversion table. The use of metanorms [32] for the COWAT-FAS produced a normal distribution of z-scores when converted from raw scores. The subsequent conversion of z-scores to scaled scores also maintained a normal distribution. This ensured that the COWAT-FAS scaled scores were similar to those of Lucas et al. [31] since their data were also normally distributed and they used scaled scores to classify Category Fluency performance.

Statistical Analysis

Two-sample t-tests were carried out to determine group differences on raw scores and scaled scores for COWAT-FAS and Category Fluency. Bonferroni correction was implemented to adjust for multiple comparisons. Statistical significance was defined by Bonferroni-adjusted p values that were ≤ 0.05. Logistic regression analyses were used to derive the diagnostic value of the individual fluency tests when used in conjunction with the HVLT-R delayed recall as expressed by the area under the curve (AUC) value. AUC values for the logistic models were compared in order determine if the additive effect of COWAT-FAS or Category Fluency significantly improved the discriminatory power of HVLT-R delayed recall. The Nagelkerke R² was used to determine the amount of variance that each logistic model accounted for.

Results

The study sample had a mean age of 74.80 ± 5.77 years, with a mean education level of 14.23 ± 2.27 years, and was comprised of 22 females and 44 males. Results from the neuropsychological tests used in the consensus diagnosis are displayed in table 1. Table 2 displays performance on COWAT-FAS, Category Fluency and the HVLT-R for both groups. Both mean raw and mean scaled scores are reported for COWAT-FAS and Category Fluency while only raw scores are reported for HVLT-R. Statistically significant differences and strong effect sizes between the CN and aMCI groups were found for both the raw and scaled scores for COWAT-FAS and Category Fluency.

Table 1. Diagnostic neuropsychological performance by clinical group

<table>
<thead>
<tr>
<th></th>
<th>CN</th>
<th>aMCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMS-R Logical Memory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate recall</td>
<td>11.82 ± 4.40</td>
<td>8.80 ± 4.20</td>
</tr>
<tr>
<td>Delayed recall</td>
<td>10.70 ± 4.13</td>
<td>7.21 ± 4.58</td>
</tr>
<tr>
<td>Trails A</td>
<td>28.09 ± 8.02</td>
<td>39.97 ± 12.80</td>
</tr>
<tr>
<td>Trails B</td>
<td>79.00 ± 23.20</td>
<td>119.64 ± 48.32</td>
</tr>
<tr>
<td>WAIS-R digit symbol</td>
<td>48.97 ± 9.06</td>
<td>38.39 ± 8.95</td>
</tr>
<tr>
<td>BNT</td>
<td>55.15 ± 3.77</td>
<td>51.73 ± 5.66</td>
</tr>
</tbody>
</table>

Raw scores are reported for all measures. Data are means ± SD. BNT = Boston Naming Test. ¹ Story 1 only.

Three separate logistic regression analyses were carried out to determine the diagnostic value of the verbal fluency measures, with the outcome set as aMCI and CN as the reference. The first model used only the HVLT-R delayed recall raw score as the predictor variable, the second logistic model used HVLT-R delayed recall and Category Fluency, and the third logistic model used HVLT-R delayed recall and COWAT-FAS raw scores as the predictor variables. The respective AUC values for each model were then compared to determine if they were significantly different from the model that used only HVLT-R delayed recall. These results are displayed in table 3.

Discussion

The results of this study show that decreases in both COWAT-FAS and Category Fluency performance can occur in individuals with single-domain aMCI when compared to CN individuals, which is consistent with previous research [9, 13–15]. Despite the statistically significant differences and the impressive effect sizes, it is important to note that these differences were not found on a clinical level. Consequently, it is difficult to determine the extent to which COWAT-FAS and Category Fluency performance are predictive of single-domain aMCI. Although statistically significant differences were noted on the scaled scores for these measures, the scaled scores themselves were within currently defined normal limits of performance. These findings suggest that the COWAT-FAS and Category Fluency tests are not sensitive enough to detect clinically meaningful differences in single-domain aMCI individuals.
However, high premorbid function that is followed by subsequent decline may be indicative of impairment. For example, an individual who obtains a scaled score of 13 on a baseline assessment and later obtains a scaled score of 8 on the same measure could be deemed impaired despite the fact that a scaled score of 8 is still considered to be within normal limits. In addition, neither of the fluency measures improved diagnostic accuracy when combined with the HVLT-R delayed recall. Recent evidence has demonstrated that decreased COWAT-FAS performance is a strong predictor of conversion to aMCI from normal cognition [33]. Oulhaj et al. [34] also demonstrated that longitudinal changes in verbal expression are highly predictive of conversion from normal cognition to aMCI. Lower scores on the expression subtest of the Cambridge Cognitive Examination [35], which includes verbal fluency, comprehension and other verbal tasks, were associated with a significantly shorter time to aMCI conversion from normal cognition. Additional longitudinal evidence is demonstrated by Clark et al. [36] who showed that both category and letter fluency are strong predictors of conversion from normal cognition to preclinical AD. However, they found that category fluency was a stronger predictor of conversion than letter fluency. These studies strongly suggest that longitudinal changes in verbal fluency are predictive of incident aMCI, so it is possible that serial assessments are needed to elucidate the predictive value of these measures in single-domain aMCI.

The findings of this study are prescient, given the newly proposed diagnostic criteria which aim to define the preclinical stages of AD [37]. Sperling et al. [37] state that cognitive decline in non-memory domains often accompanies declines in episodic memory which highlights the need to identify neuropsychological tests that will aid in identifying individuals in the earliest stages of cognitive decline. This statement is supported by longitudinal results from Johnson et al. [38] and Small and Bäckman [39], showing that decline in cognitive domains other than memory is detectable in preclinical AD, and also by Kramer et al. [40] and Ribeiro et al. [41] who also demonstrated that non-memory domains show decline in individuals with MCI.

One weakness of the study is that our sample is ethnically homogenous as the majority of the participants were Caucasians. As a result, it is unknown whether these findings can be generalized to a more ethnically diverse group. Another weakness of the study is the inherent circularity of using the COWAT-FAS and Category Fluency measures to discern group differences. Although both

### Table 2. Independent sample t-test for COWAT-FAS, Category Fluency and HVLT-R

<table>
<thead>
<tr>
<th>Measure</th>
<th>CN</th>
<th>aMCI</th>
<th>t-value</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>COWAT-FAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>raw score</td>
<td>41.70 ± 10.56</td>
<td>28.09 ± 9.95</td>
<td>5.39</td>
<td>1.33</td>
</tr>
<tr>
<td>scaled score</td>
<td>12.82 ± 2.52</td>
<td>9.67 ± 2.46</td>
<td>5.15</td>
<td>1.26</td>
</tr>
<tr>
<td>Category Fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>raw score</td>
<td>45.15 ± 10.44</td>
<td>35.67 ± 7.27</td>
<td>4.28</td>
<td>1.05</td>
</tr>
<tr>
<td>scaled score</td>
<td>11.27 ± 2.61</td>
<td>8.70 ± 1.93</td>
<td>4.56</td>
<td>1.12</td>
</tr>
<tr>
<td>HVLT-R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total recall</td>
<td>24.67 ± 4.49</td>
<td>18.18 ± 4.67</td>
<td>5.75</td>
<td>1.42</td>
</tr>
<tr>
<td>delayed recall</td>
<td>8.30 ± 2.34</td>
<td>5.42 ± 2.69</td>
<td>4.64</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Data are means ± SD. p < 0.001 for all comparisons with Bonferroni correction; degrees of freedom = 64.

### Table 3. Logistic regression model comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>95% CI</th>
<th>R²</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 - HVLT-R delayed recall</td>
<td>0.79</td>
<td>0.67–0.88</td>
<td>0.33</td>
<td>–</td>
</tr>
<tr>
<td>Model 2 - HVLT-R delayed recall and Category Fluency</td>
<td>0.81</td>
<td>0.70–0.90</td>
<td>0.38</td>
<td>0.39¹</td>
</tr>
<tr>
<td>Model 3 - HVLT-R delayed recall and COWAT-FAS</td>
<td>0.88</td>
<td>0.78–0.95</td>
<td>0.56</td>
<td>0.08¹</td>
</tr>
</tbody>
</table>

CI = Confidence interval. ¹ Significance level when the AUC value is compared to model 1.
measures were used to make consensus diagnoses, the issue of circularity is not as damaging in this study as all of the individuals fell within currently defined normal limits on these measures. This problem has been noted previously when these same measures were also used in the diagnostic process [12, 42].

The current study suggests that individuals with single-domain aMCI have decreased performance on both COWAT-FAS and Category Fluency when compared to CN individuals. Despite statistically significant differences accompanied by very large effect sizes, these results did not demonstrate a clinically significant difference in the ability to differentiate aMCI from CN. The results of previous longitudinal studies strongly suggest that serial assessment may be necessary to elucidate the true predictive value of COWAT-FAS and Category Fluency in conversion from normal cognition to aMCI. However, this study demonstrated that COWAT-FAS and Category Fluency do not provide additional diagnostic value when differentiating single-domain aMCI individuals from CN individuals.

Acknowledgment

This work was supported by the National Institute of Health for the Florida Alzheimer’s Disease Research Center (1P50AG025711-01) and the Banner Sun Health Research Institute (P30AG019610).

Disclosure Statement

The authors of this study have no financial ties or conflicts of interest to disclose.

References


