Performance on the RI-48 Cued Recall Test Best Predicts Conversion to Dementia at the 5- and 10-Year Follow-Ups

Bernard Hanseeuw  Adrian Ivanoiu

Department of Neurology, Saint Luc University Hospital, and Institute of Neuroscience, Catholic University of Louvain, Brussels, Belgium

Key Words
Alzheimer's disease · Encoding specificity · Follow-up study · Mild cognitive impairment · RI-48 cued recall test

Abstract
Background: The RI-48 cued recall test was devised to discriminate between healthy elderly and patients with mild cognitive impairment who are at risk of developing Alzheimer’s disease (AD). However, no long-term follow-up studies have been conducted using this test. Methods: We analyzed the predictive power of the RI-48 test for determining the patients who will convert to AD dementia within the decade after testing. During 10 years, we followed up 40 non-demented patients who attended our Memory Clinic and underwent complete neuropsychological evaluation including the RI-48. Results: Of the 40 patients, 21 developed dementia (converters, CO) and 19 remained stable patients (SP). Of the tests performed at inclusion, only the RI-48 (p < 0.0001) and semantic fluency (p = 0.030) tests gave significantly different results between CO and SP. Conclusion: The RI-48 had the best overall diagnostic accuracy at 5- and at 10-year follow-ups.

Introduction

It is now widely acknowledged that Alzheimer’s disease (AD) develops progressively before reaching the dementia stage. This progression to dementia can take many years, if not decades. Autopsy studies have revealed that some non-demented patients have typical AD...
lesions [1, 2]. Predicting AD dementia years before its occurrence is, therefore, theoretically possible but remains a challenge in clinical practice.

The concepts of mild cognitive impairment (MCI) and amnestic MCI (aMCI) have been suggested as indicative of this pre-dementia phase of AD [3, 4]. Patients with aMCI complain about their memory and suffer from episodic memory impairment on cognitive testing; however, this does not disturb their daily activities, and other cognitive functions are well preserved, so they are not demented. Although constituting a risk factor for AD, only about 70% of patients with aMCI will convert to dementia within 5 years [5] and a significant percentage of patients with aMCI will regain normal memory performance within 2 years [6].

Subjective cognitive impairment (SCI), i.e. subjective memory complaints not confirmed by cognitive testing, may also constitute a risk factor for future conversion to dementia [7], although much less so than MCI (9.2% within 3 years [8]). Memory complaints are very common during healthy aging [9], and particularly in elderly people suffering from depression [10]. Hence, neither memory complaints (SCI) nor episodic memory deficits (aMCI) can be considered as being 100% reliable as predictors of conversion to dementia in clinical practice. Episodic memory therefore needs to be better characterized in order to identify those individuals who will later convert to AD [11]. Some memory tests may be better than others in the differentiation between healthy elderly individuals and patients with SCI, MCI or mild AD [12].

Most cross-sectional studies that focus on memory testing suffer from circular reasoning. Because diagnostic labels such as MCI or SCI are defined by memory testing, studying memory tests by comparing these groups is somewhat flawed (although study tests are different from those used for diagnosis, they will always be correlated to some extent). By definition, patients with MCI have a worse memory than those with SCI and a better memory than patients with AD. This does not, however, exclude the possibility that some individuals with SCI may progress to AD and some MCI patients may remain stable. We believe the best way to assess the clinical usefulness of cognitive testing is to follow patients regardless of their initial diagnostic label (SCI or MCI). Nevertheless, longitudinal studies remain rare; most have only short follow-up periods (≤ 3 years), and some also use diagnostic labels as part of their inclusion criteria (in order to increase conversion rates; but again this may lead to circularity).

Moreover, although several longitudinal studies have been conducted, only a few have included memory tests controlling effective encoding of information. In a 5-year follow-up study, executive function (the trail-making test, TMT) and episodic memory [the California Verbal Learning Test (CVLT)] appeared to be the cognitive functions most predictive of future conversion to dementia [13]. However, although the CVLT uses the cued recall technique, it does not control for the effectiveness of encoding. Indeed, the category cued recall (which includes an encoding control by immediate recall of each item category pair) allows very good separation between healthy elderly and mild AD individuals [14]. A French adaptation of the category cued recall by our team (the RI-48 test) had the best sensitivity and specificity for differentiating MCI from SCI and healthy elderly [12]. Moreover, the effectively encoded items separated groups best [15]. However, as explained above, this good discriminative power to distinguish between clinical entities (AD, MCI and SCI) in cross-sectional studies does not imply the same power for predicting dementia conversion. Indeed, in a 12- to 18-month follow-up study, we showed that visual memory tests tended to have better discriminative power than the RI-48 to separate evolving from stable MCI [12].

The predictive power of verbal cued recall tests with encoding specificity control has been investigated by Sarazin et al. [16] and by Dierckx et al. [17]. However, these studies had relatively short follow-up times (36 and 18 months, respectively) and used short lists (16 and
6 items to remember, respectively, instead of 48 items in the RI-48). This small number of items per list can result in ceiling effects in the healthy population and in better performing patients [14].

In the present study, to give a clinical perspective, we followed all patients who attended the Memory Clinic and expressed memory complaints seriously enough to justify a neuropsychological examination. We included 50 non-demented patients with memory complaints (regardless of whether they were SCI or MCI), in whom we performed neuropsychological testing including the RI-48 task. We followed most of these patients for ≥10 years. The main goal of the study was to evaluate the diagnostic accuracy of the RI-48 cued recall test and other neuropsychological tasks to predict future conversion to dementia. We also wanted to compare the predictive power of these tasks in the long (5 years) and in the very long (10 years) run.

**Patients and Methods**

**Subjects**

**Initial Population.** Fifty non-demented patients with memory complaints who attended the Memory Clinic of the Saint Luc University Hospital between 1998 and 2002 were included in this study. We excluded patients who had other neurological or psychiatric conditions, particularly patients suffering from dementia (according to the NINCDS-ADRDA criteria [18]) or from major depression (according to the DSM IV criteria). All patients had a computed tomography scan or magnetic resonance imaging to rule out vascular or other focal causes of cognitive impairment. The Mini-Mental State Examination (MMSE) score at inclusion was ≥24/30 (mean ± SD: 27.3 ± 1.8) in all subjects [19]. The initial population included 20 males and 30 females with an average age of 68.4 ± 7.6 years.

**Final Population.** Twenty-five patients were regularly followed up at the Memory Clinic starting from their initial evaluation until 2009 or death. Six of them died during the follow-up, and 5 of the 6 had dementia. In 2010, the remaining 25 patients were contacted by telephone about their fate. Fifteen additional patients were also assessed this way. Three patients died without follow-up, and 7 patients could not be contacted. In total, 40 patients were thus available for the study and had their final follow-up after a median of 10 years (range: 1–13, mean: 8.5 ± 3.5 years). Patients (n = 10) who dropped out from the study were no different from those who were followed up (n = 40) in any measure. Among the 40 patients who were finally assessed, 21 were demented [= converters (CO)] and 19 were not [= stable patients (SP)]. Hence, 52.5% of the initial population converted to dementia during the study period. All SP except for 1 patient (who died during the follow-up without dementia) were followed for ≥7 years in order to ensure that none had been misdiagnosed.

**The RI-48 Task**

This task has already been described in detail previously [12, 15]. The RI-48 task includes 48 items belonging to 12 semantic categories. Items are presented to participants as written words on 12 consecutive cards, each card containing 4 items from different categories (e.g. the first card contained the French words for an insect – ‘ladybird’, a fruit – ‘raspberry’, a tree – ‘palm’ and a garment – ‘jacket’). Patients are asked to encode these items with the category given as a semantic cue. On completion of each card, an immediate cued recall test is performed and the patient’s performance is recorded. After showing the last card, participants are asked to count backwards for 20 s. Participants then perform a cued recall task, using the categories as cues, e.g. ‘what were the flowers, insects, etc.’. The patient’s performance and incorrect answers are recorded.
Neuropsychological Testing
All patients had an extensive neuropsychological assessment at the time of inclusion. This testing evaluated episodic memory with a verbal (the RI-48) and a non-verbal (the door test from the ‘Doors and People’ battery [20]) task. The doors test is a recognition task involving the presentation of two series of color photographs of doors, which subsequently have to be selected from four alternatives. The first set of 12 doors (set A) is easier than the second one (set B).

Visuospatial processing was assessed by the command and copy condition of the ‘Clock Drawing Test’ (CDT) [21] and by the ‘Praxis’ part of the CERAD (Consortium to Establish a Registry for Alzheimer’s Disease) battery [22]. Language and semantic memory were assessed with the LEXIS naming test [23] and with the category animal fluency, i.e. the number of different animals listed in 2 min.

Executive function was evaluated with the letter fluency test for the letter P over 2 min [23], the TMT and the ‘Stroop’ test. Executive indices were computed for the latter two. The TMT index was the time necessary to perform part ‘B’ (tracking number and letter alternatively) minus the time for part ‘A’ (simple tracking). The Stroop index was the time necessary to perform the interference condition divided by the average time to perform the reading and the color naming conditions. Similar indices were calculated for errors.

Global cognitive functioning was evaluated at inclusion and at follow-up. This was assessed clinically and using the MMSE or a variant of this test adapted for telephone administration (TICS-30 = modified version of the Telephone Interview for Clinical Status [24]). We converted the results of the TICS-30 into an MMSE score according to Fong et al. [25] in order to compute the decrease in the MMSE score per year of follow-up for each patient.

Statistical Analysis
All analyses were performed using Statistica (version 9). We first performed between-group analyses comparing demographic data (age and sex) of SP and CO. Significantly different demographic data were introduced as covariates in subsequent analyses. We then performed between-group analyses for every measure of the neuropsychological assessment including MMSE at inclusion and at follow-up and decrease per year in MMSE score.

Second, we introduced all measures that showed at least a trend towards a significant difference between groups [(two-tailed) \( p < 0.0999 \)] into a multiple regression analysis in order to determine which variables best predicted conversion to dementia. Third, we computed sensitivity, specificity and overall diagnostic accuracy for these cognitive tests. We performed this analysis twice: once using our final diagnosis (10 years of median follow-up) and once considering patients as CO only when they converted within 5 years of inclusion (17 of the 21 CO) in order to assess the predictive power of neuropsychological tasks in the long (5 years) and very long (± 10 years) run.

Missing Data
Five SP did not do the doors test and 3 did not have the visuospatial assessment, 2 patients did not do the TMT (1 CO and 1 SP), and 7 did not do the Stroop task (2 CO and 5 SP).

Results

Between-Group Analyses
Demographic Data. Average ages at inclusion were 65.0 ± 7.9 years for the SP and 71.8 ± 6.4 years for the CO group; this was significantly different between groups (\( F_{(1,39)} = 8.94; p = 0.0049 \)). Consequently, we decided to introduce age as a covariate in all subsequent
analyses. There were 8 males and 11 females in the SP group, and 8 males and 13 females in the CO group; this difference was not statistically significant ($\chi^2 = 0.07; p = 0.7960$).

**Neuropsychological Data.** At study inclusion, only results from the RI-48 delayed recall ($p = 0.00008$) and animal fluency ($p = 0.0307$) were significantly different between SP and CO. The scores for the MMSE ($p = 0.0637$) and the copy condition of the CDT ($p = 0.0753$) tended to be significantly lower in CO than in SP (table 1). At follow-up, MMSE was 27.3 ± 1.7 in SP and 17.5 ± 6.9 in CO. This difference was highly significant ($F(1, 38) = 25.34; p = 0.00001$). Similarly, the decrease in MMSE score per year was 0.1 ± 0.3 in SP and 1.6 ± 0.9 in CO, which was also significant ($F(1, 38) = 33.88; p = 0.00001$).

**Multiple Regression**

We then performed a multiple regression with clinical follow-up outcome (SP vs. CO) as the dependent variable and age, inclusion MMSE, RI-48 delayed recall, CDT copy and animal fluency introduced as continuous predictors. This model was predictive of clinical follow-up outcome (adjusted $R^2 = 0.47; p = 0.0001$). In particular, the RI-48 delayed recall test ($\beta = -0.41; p = 0.026$) and age ($\beta = 0.29; p = 0.030$) were significant predictors, whereas animal fluency ($\beta = -0.06; p = 0.703$) and inclusion MMSE ($\beta = -0.15; p = 0.352$) were not. The CDT showed a tendency towards being a significant predictor ($\beta = -0.27; p = 0.066$).

### Table 1. Neuropsychological data at inclusion of SP (n = 19) and CO (n = 21)

<table>
<thead>
<tr>
<th>Data at inclusion</th>
<th>SP mean</th>
<th>SP SD</th>
<th>CO mean</th>
<th>CO SD</th>
<th>$F_{(1, 38)}$</th>
<th>p (two tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE, n/30</td>
<td>27.8</td>
<td>1.9</td>
<td>26.5</td>
<td>1.7</td>
<td>3.65</td>
<td>0.0637</td>
</tr>
<tr>
<td>RI-48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate recall, n/48</td>
<td>35.4</td>
<td>6.6</td>
<td>31.2</td>
<td>5.7</td>
<td>2.17</td>
<td>0.1494</td>
</tr>
<tr>
<td>Delayed recall, n/48</td>
<td>21.4</td>
<td>4.5</td>
<td>14.6</td>
<td>6.1</td>
<td>13.08</td>
<td>0.00008</td>
</tr>
<tr>
<td>Intrusions, n</td>
<td>3.5</td>
<td>3.1</td>
<td>6.2</td>
<td>8.7</td>
<td>1.98</td>
<td>0.1673</td>
</tr>
<tr>
<td>Doors test, n/12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set A</td>
<td>9.5</td>
<td>2.0</td>
<td>8.7</td>
<td>2.9</td>
<td>0.22</td>
<td>0.6440</td>
</tr>
<tr>
<td>Set B</td>
<td>5.6</td>
<td>3.0</td>
<td>4.7</td>
<td>1.9</td>
<td>1.22</td>
<td>0.2777</td>
</tr>
<tr>
<td>Clock drawing (command), n/10</td>
<td>8.9</td>
<td>0.9</td>
<td>8.1</td>
<td>2.1</td>
<td>1.61</td>
<td>0.2129</td>
</tr>
<tr>
<td>CDT (copy), n/10</td>
<td>9.6</td>
<td>0.7</td>
<td>8.9</td>
<td>1.2</td>
<td>3.36</td>
<td>0.0753</td>
</tr>
<tr>
<td>Praxis part of the CERAD, n/11</td>
<td>9.8</td>
<td>1.2</td>
<td>9.3</td>
<td>1.6</td>
<td>0.17</td>
<td>0.6791</td>
</tr>
<tr>
<td>LEXIS naming test, %</td>
<td>85.3</td>
<td>7.3</td>
<td>82.5</td>
<td>6.8</td>
<td>0.45</td>
<td>0.5074</td>
</tr>
<tr>
<td>Fluency tests for 2 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animals</td>
<td>28.4</td>
<td>8.1</td>
<td>21.6</td>
<td>7.1</td>
<td>5.04</td>
<td>0.0307</td>
</tr>
<tr>
<td>Letter P</td>
<td>19.7</td>
<td>6.1</td>
<td>19.1</td>
<td>8.0</td>
<td>0.03</td>
<td>0.8620</td>
</tr>
<tr>
<td>TMT (B – A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seconds</td>
<td>70.5</td>
<td>60.0</td>
<td>107.2</td>
<td>62.1</td>
<td>1.59</td>
<td>0.2162</td>
</tr>
<tr>
<td>Errors</td>
<td>0.8</td>
<td>1.4</td>
<td>0.9</td>
<td>1.0</td>
<td>0.28</td>
<td>0.6007</td>
</tr>
<tr>
<td>Stroop index</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time$^a$</td>
<td>2.3</td>
<td>0.5</td>
<td>2.5</td>
<td>0.6</td>
<td>0.30</td>
<td>0.5877</td>
</tr>
<tr>
<td>Errors</td>
<td>1.0</td>
<td>2.2</td>
<td>5.5</td>
<td>10.5</td>
<td>0.88</td>
<td>0.3559</td>
</tr>
</tbody>
</table>

Significant differences are shown in bold. Age was introduced as a covariate in all analyses. The number of subjects and degrees of freedom were adapted according to missing data (see Patients and Methods/Missing Data in the text).

*a Time to perform the interference condition divided by the average time to perform the reading and the color naming conditions.
Diagnostic Accuracy

The best cutoff for the RI-48 delayed recall was 16/48. The best animal fluency cutoff was 23 animals in 2 min. The best cutoff for inclusion MMSE was 27/30 and 9/10 for CDT copy. Sensitivity, specificity and overall diagnostic accuracy are shown in Table 2: the RI-48 had the best overall diagnostic accuracy at the 5- (85%) and 10-year (75%) follow-ups. Of note, when looking at the wrongly classified CO (i.e. RI-48 $\leq$ 16/48 at inclusion = patients complaining of memory loss without objective memory impairment = SCI), they converted to dementia in 6.2 years on average whereas CO with low RI-48 scores ($<_{16/48}$ = aMCI) at inclusion converted in 2.6 years.

Discussion

This study confirmed that verbal cued recall with effective encoding control and, in particular, the RI-48 task had good predictive power for conversion to dementia in patients attending a memory clinic. Previous studies using verbal cued recall have shown similar results but with shorter-term follow-ups and using different tasks [16, 17, 26]. Dierckx et al. [17] reported an overall diagnostic accuracy of 87% over 18 months with 7 CO from 31 patients with MCI. Sarazin et al. [16] included more patients with MCI (251) and proved the effectiveness of the Free and Cued Selective Reminding Test (FCSRT) to predict conversion of MCI to dementia (59 conversions in 3 years). However, the FCSRT may not be appropriate when considering patients attending Memory Clinics who do not suffer from episodic impairment (SCI) since this task has ceiling effects [14]. In a 5-year follow-up study, Dickerson et al. [26] found that the cued recall tests (FCSRT and CVLT) were good predictors of conversion of MCI to AD, but less predictive for what they called ‘very mild cognitive impairment’ (a concept close to what we call SCI). Moreover, good predictive power – as has been shown for FCSRT – in relatively short-term follow-up periods may not remain in the longer term, as demonstrated by the decrease in the predictive power of the neuropsychological tasks at the later follow-up period. This could also be the explanation for the smaller predictive power observed for FCSRT in patients with SCI, which, if evolving, will only convert after a long follow-up time (6.2 years in our study).

Table 2. Sensitivity, specificity and overall diagnostic accuracy (ODA) of neuropsychological tests that showed at least a trend towards a significant difference between CO and SP at the 5-/10-year follow-up

<table>
<thead>
<tr>
<th>Neuropsychological tests</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>ODA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI-48: delayed recall</td>
<td>76.5%</td>
<td>91.3%</td>
<td>85.0%</td>
</tr>
<tr>
<td>Animal fluency (2 min)</td>
<td>70.6%</td>
<td>69.6%</td>
<td>70.0%</td>
</tr>
<tr>
<td>Inclusion MMSE &lt;27/30</td>
<td>88.2%</td>
<td>65.2%</td>
<td>75.0%</td>
</tr>
<tr>
<td>CDT (copy) &lt;9/10</td>
<td>76.5%</td>
<td>56.5%</td>
<td>65.0%</td>
</tr>
<tr>
<td>10-year follow-up*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI-48: delayed recall</td>
<td>61.9%</td>
<td>89.5%</td>
<td>75.0%</td>
</tr>
<tr>
<td>Animal fluency (2 min)</td>
<td>66.7%</td>
<td>73.7%</td>
<td>70.0%</td>
</tr>
<tr>
<td>Inclusion MMSE &lt;27/30</td>
<td>76.2%</td>
<td>63.2%</td>
<td>70.0%</td>
</tr>
<tr>
<td>CDT (copy) &lt;9/10</td>
<td>71.4%</td>
<td>57.9%</td>
<td>65.0%</td>
</tr>
</tbody>
</table>

* Median – see text for details.
The RI-48 had an overall diagnostic accuracy of 85% at 5 years and of 75% at 10 years. Because we included a group of consecutive patients attending our Memory Clinic, this study is representative of a typical population of patients with memory complaints who are not demented. We would thus recommend the use of the RI-48 to predict dementia because it had the best predictive power at 10 years and does not have a ceiling effect in the healthy population (in SCI or in better-performing MCI).

New criteria for AD have recently been proposed [11] and are employed in various memory clinics [27]. These criteria do not require that patients have reached a dementia stage in order to be diagnosed as AD, but that they have a positive biomarker and an objective memory impairment defined as a 'recall deficit that does not normalize with cueing (…) and after effective encoding of information has been previously controlled'. As shown in this study, the RI-48 fulfills these requirements.

Among the non-memory tests, the inclusion MMSE showed the best sensitivity (88% at 5 years) for detecting future conversion to dementia, but at the cost of a high threshold (<27/30), which reduces its specificity (65% at 5 years). This confirms the clinical use of the MMSE as a screening tool that allows the determination of patients who should undergo complete neuropsychological evaluation (including tests with higher specificity). In agreement with other studies [28], the animal fluency test was a sensitive test to predict dementia. Two long-term studies showed that the animal fluency test was predictive of dementia at 6 [29] and 9 [30] years. The present study nevertheless showed that the predictive power of the animal fluency test was lower than that of the RI-48 over similar follow-up periods.

Conclusion

This long-term follow-up study showed that a verbal cued recall test with controlled encoding, the RI-48 task, can predict dementia 5 and even 10 years before its occurrence. The advantage of this study was that we included patients according to their memory complaints rather than their memory performance, which allowed us to avoid any circular reasoning. The absence of a ceiling effect for the RI-48 task makes it particularly suitable as an assessment tool for SCI patients and better-performing MCI patients in whom more traditional cued recall tests such as the FCSRT have been shown to be of limited value.

Acknowledgments

B.H. has been supported by FNRS (Belgian Foundation for Scientific Research). We are grateful to Charlotte Rensonnet for helping us with data collection.

References


