Original Article

A Delayed Recall Battery as a Sensitive Screening for Mild Cognitive Impairment: Follow-up Study of Memory Clinic Patients after 10 years

Yutaka Takayama

Department of Psychiatry, Mita Hospital, International University of Health and Welfare, 1-4-3 Mita, Minato-ku, Tokyo 108-8329, Japan

Objective: I examined the predictive value of the combination of three delayed recall tests to distinguish 1) those with probable Alzheimer’s disease (AD) from those within normal range, and 2) those with Mild cognitive impairment (MCI) from those within normal range. The data from 90 visitors to a memory clinic in Tokyo was used. I first examined patients clinically, neuroradiologically, and excluded the mental and neurological illness. AD was diagnosed according to the NINCDS-ADRDA criteria, MCI according to the criteria of Petersen et al. Normal must be free from any disease examined above. Methods: After the diagnosis, the baseline neuropsychological tests were performed for all participants; the Mini Mental State Examination, Raven’s Colored Progressive Matrices, the Stroop Test, a 10-words list learning and recall test, a story recall test, and the Rey-Osterrieth Complex Figure Test. After 10 years, all patients were reassessed and diagnosed again. Results: Of the MCI patients for follow-up (n=29), 19 were converted to AD, while 5 not. One died. 4 lost. All AD patients (n=30) remained as AD. The combination of 3 delayed recall battery provides clinically useful predictive values for both AD and MCI in a memory clinics and dementia research clinics.

Key words: Mild cognitive impairment (MCI), Alzheimer’s disease (AD), neuropsychological tests, delayed recall, and predictive value.

Introduction

With the increase in life span, the absolute number and proportion of aged individuals has increased. As a consequence, age-related cognitive disorders, such as neurodegenerative dementia, will become even more prevalent in the coming years. Considering that cognitive impairment results in the increased use of health services and increased mortality, and taking into account that new treatment options emerge, it is warranted that screening and diagnostic tools are improved.

At present, Alzheimer’s disease (AD) is the most common cause of dementia, therefore the identification of early clinical AD has become very important as new treatments are being developed. Patients might benefit from present cognition-enhancing drugs (donepezil and others) or future disease-modifying drugs.

In all neuropsychological measures, delayed recall of the word list was the most successful discriminator to classify correctly 96% of normal subjects and 86% of mild AD subjects with the Mini Mental State Examination (MMSE) scores of 24 or greater. On the other hand, a detailed review found that the MMSE classified only 44-68% of mild AD using the standard 23/24 cutoff points especially when subjects are those with mild dementia (MMSE>20). This is the reason why the MMSE nowadays is not used as a screening procedure any more but for assessment of dementia severity.

In Canada, a study to determine whether neuropsychological test accurately predict incident Alzheimer’s disease after 5 and 10 years in participants of Canadian Study of Health and Aging who are initially nonde-
mented. The 10-year prediction study included those who completed neuropsychological testing then. In the 10-year follow-up study, only one test (short delayed verbal recall) emerged from the forward regression analyses. The model was significant, \( \chi^2 = 36.61 \), \( p < 0.0001 \) (sensitivity = 74\%, specificity = 83\%). It means that in a large epidemiologic sample of nondemented participants, neuropsychological tests accurately predicted conversion to Alzheimer’s disease after 10 years\(^6\). Currently, Mild Cognitive Impairment\(^7, 8\) (MCI) become widely used term to describe the intermediate between normal ageing and dementia. With this concept, a new impetus to develop strategies to identify individuals at high risk for having future dementia has emerged. MCI is associated with an increased risk of developing dementia: patients develop dementia at the rate of 10-15\%/year\(^1, 6, 8, 9\) compared with the healthy seniors who develop dementia at the rate of 1.2\%/year\(^1, 6\). This is why MCI patients have been suspected to have some sort of fragility in the limbic region, so that a new therapy might be developed. However, not all patients\(^8, 9\) will undergo dementia in the same way, as some will remain stable for years or rarely improve scores of neuropsychological tests. In such cases, doctors or psychologists could provide an effective first screening procedure for their rehabilitations.

Memory tasks appear to have the predictive power for indicating early AD as the episodic memory function is anatomically located within the medial temporal lobe structures, including the hippocampus\(^10\) which is also the location with the highest densities of AD neuropathological lesions. Most studies addressed the issue of neuropsychological prediction of the conversion to AD in MCI patients using several types of memory tests and proved that delayed recall tasks are very sensitive to the early detection of and the conversion to AD\(^8, 11\). There are two study designs in agnostic MCI research: i.e. one is a cross-sectional design looking for features that optimally distinguish MCI patients from those who are cognitively normal and the other is a longitudinal design to investigate which features at the baseline of a group of MCI patients optimally predict the conversion to dementia.

Objective;

Although MCI has a high risk that will convert to AD rapidly, there are a few studies that investigate the probable speed at which MCI will convert to AD and what sort of conditions accelerate the conversion, and what sort of cognitive decline is most harmful.

This longitudinal study investigated what sort of neuropsychological test is able to distinguish AD or MCI from the normally ageing senior in the very early stage. We paid special attention to modalities of delayed recall memory tests because each type of memory (word, story, visual) has been memorized in somewhat different way\(^11, 12\). This study is also to ascertain that a follow-up diagnosis after 10 years is the same as those for the first assessment, especially for MCI and AD.

Materials and Methods:

Subjects; The Subjects consisted of 90 patients who were recruited from patients of a memory clinic in Tokyo. They were explained that the examination in this study was the same as what they had already been routinely administered in the memory clinic treatment at a minimum risk. All the participating subjects gave their informed consents and those were documented on the medical record sheets. After the diagnoses are determined, all subjects were investigated using the following screening tests for the baseline assessments:

Screening tests; MMSE\(^3\), RCPM\(^13\), The Stroop test\(^14\) (Scores of these three tests are listed in Table 1) were used for the diagnosis of subjects. A 10-words (list) learning\(^15\), A story recall test\(^16\), and ROCFT\(^16\) (Scores of these three tests are given in Table 2) were also carried out for the detailed assessment of the memory\(^11\). These neuropsychological screening tests were carried out at patient’s first visit after his diagnosis was decided.

Mini Mental State Examination\(^3\) (MMSE) is a well-known brief general cognitive battery that measures orientation, immediate and delayed memory, concentration, language, and praxis. The Highest scores are 30. Raven’s Colored Progressive Matrices\(^13\) (RCPM) is a well-known assessment battery of nonverbal intelligence using the ability of matching appropriate colors and patterns. The Highest scores are 36. The Stroop Test\(^14\) is a well-known battery assessing one’s attention ability. The time necessary to name the colors of 24 dots (color naming) and the time (seconds) necessary to name the printed colors of 24 Japanese KANJIs with irrelevant meanings (reading KANJI) are evaluated. A subject needs a long time to read the colors if his ability of attention declines. A 10-words list learning and recall test\(^15\) (10-words list learning or 10-words learning) is a variation of Rey’s Auditory Verbal Learning Test\(^15\). The modifications included using a word list of 10 words with a matching line drawing card for each word and
repeating a memorizing process 5 times. Immediate free recall words and free delayed recall words after 30 minutes are assessed as the memory scores. The Highest scores are 10. **A Story recall test** is a variation of the logical memory part of the Wechsler Memory Scale. The modifications included using a short story consisting of 15 words and the story is composed of three sentences. Subjects hear the story only once and immediate free recall words and free delayed recall words after 30 minutes are assessed as the story memorizing scores. The Highest scores are 15. **Rey-Osterrieth Complex Figure Test** (ROCFT) is a well-known battery for the assessment of visuospatial constructional cognitive functions (copy) and visuospatial memory (immediate recall and delayed recall). This test was carried out in the original way. The Highest scores are 36. Results of These three memory test (10-words learning, Story recall, and ROCFT) are listed in Table 2.

**Ethical consideration:**
This study started in 1992 when a memory clinic investigated whether neuropsychological tests are able to discriminate among Alzheimer’s disease, when MCI is yet called as Age-Associated Memory Impairment (AAMI). Our memory clinic was one of the first few ones for the cognitively healthy seniors and MCI in Tokyo. At that time there was no Ministerial Ordinance of Good Clinical Practice (GCP) in Japan, which was first promulgated in 1997. Therefore, this study was performed according to the Hospital guidelines for Clinical Research by the Institutional Review Board for clinical research and the principles outlined in the Declaration of Helsinki.

**Statistics:**
All statistical analyses for the investigation of group differences were carried out using the statistics program SPSS (SPSS 11.0J for windows). A 1-way ANOVA with the factor ‘study group’ (AD, MCI, and Normal) was run to test for group differences. For post hoc comparisons, the Scheffe test was used. These results are given in Table 1 and Table 2. (The sex difference was analyzed using $\chi^2$ tests.)

**Evaluation of Screening Test:**
The effectiveness of a screening test is assessed using several indices. They are **Sensitivity** (Correct detection of cognitively impaired; e.g. AD or MCI in this paper), **Specificity** (Correct detection of normal), **Positive Predictive Value** (PPV); (this is the proportion of those whose score is less than a given cut-off score and are actually AD / MCI), and **Negative Predictive Value** (NPV); (this reflects the probability that an individual whose screening is negative is not AD / MCI).

The sensitivity, specificity, positive predictive value, and negative predictive value were calculated for each screening test (10-words list learning, story recall, and ROCFT) with each appropriate cut-off score of delayed recall. For the combination of three memory (delayed recall) tests scores, appropriate cut-off for each delayed recall score was selected for best performance. It was used in the way that the subject was assessed as cognitively impaired only when all of the three test results were less than each cut-off score. The rationale for this was to give credit to each single cognitive domain, especially each different aspect of memory function, and that normal subjects should exceed a given cutoff score in each and every neuropsychological domain.

**Results:**
The demographic characteristics, education, severity of the dementia syndrome assessed by the MMSE scores, the grade of the nonverbal intelligence assessed by the RCPM scores, and the attention ability assessed by the Stroop Test are listed in Table 1. Education was represented as years of going to any type of school.

We did not find significant differences among AD, MCI, and normal in age, sex and education. However, there was some tendency of differences in education that normal subjects might have been educated 1 year longer ($p=0.059$) and that normal group might be younger about 1-2 years. We did not think these small differences have significant effects on our results about AD, MCI symptoms.

The AD patients had significantly lower scores on the MMSE than the MCI, and the MCI was significantly lower than normal (both $p<0.05$). However, MMSE scores of three groups have one broad overlap. This is given in Table 1.

On the RCPM scores both the AD and MCI patients had significantly lower scores than normal ($p<0.05$). However, no difference was found between the AD and MCI ($p>0.05$). This means that MCI patients already started to decline in the nonverbal intelligence. This is given in Table 1.

On the Stroop test scores the AD was significantly inferior to normal ($p<0.05$) for both conditions. The
The score of MCI on the Stroop test was in the middle between AD and normal. However, there is no significant difference between AD and MCI. There is no difference between MCI and normal (Both comparison; p>0.05). This means the ability of attention declined gradually with the years and from normal to AD. This is given in Table 1.

Of the MCI patients who were for follow-up (n=29), 19 fulfilled the (NINCDS-ADRDA) criteria of probable AD, while 5 did not convert to AD and stayed in the MCI states on the Petersen criteria. One subject died of cerebral apoplexy after 9 years follow-up at our clinic. He also stayed in the MCI state. Four subjects did not continue to visit our memory clinic. After 10 years we asked them and their family members how they live and how their memory complains are and if there are other possible symptoms. Their answers were “There is no severe problem about the daily life”. On the scores of repeated neuropsychological tests, no subject of those who could be assessed improved in their cognitive function. Those who were are converted from MCI to AD showed various cognitive declines other than cerebral apoplexy after 9 years follow-up at our clinic.

### Table 1. Demographic characteristics and neuropsychological data of patients

<table>
<thead>
<tr>
<th></th>
<th>AD</th>
<th>MCI</th>
<th>normal</th>
<th>F (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>30</td>
<td>29</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>70.0±5.8</td>
<td>70.3±5.5</td>
<td>68.1±5.5</td>
<td>1.1514 (p=.249)</td>
</tr>
<tr>
<td>Female/Male</td>
<td>23/7</td>
<td>19/10</td>
<td>23/8</td>
<td>(p=.605)*</td>
</tr>
<tr>
<td>Education</td>
<td>12.1±1.7</td>
<td>12.2±2.5</td>
<td>13.3±2.2</td>
<td>2.92 (p=.059)</td>
</tr>
<tr>
<td>MMSE</td>
<td>24.0±2.5</td>
<td>26.0±2.2</td>
<td>28.5±1.9</td>
<td>32.22 (p&lt;0.001) #†‡</td>
</tr>
<tr>
<td>RCPM</td>
<td>27.4±4.1</td>
<td>29.1±3.3</td>
<td>32.0±2.2</td>
<td>15.09 (p&lt;0.001) #‡</td>
</tr>
<tr>
<td>stroop (color naming)</td>
<td>20.2±5.9</td>
<td>18.2±4.8</td>
<td>15.9±3.7</td>
<td>5.76 (p&lt;0.001) †</td>
</tr>
<tr>
<td>stroop (reading KANJI)</td>
<td>37.8±10.0</td>
<td>33.1±10.3</td>
<td>27.6±7.3</td>
<td>8.89 (p&lt;0.01) †</td>
</tr>
</tbody>
</table>

MMSE: Mini Mental State Examination, RCPM: Raven’s Colored Progressive Matrices, AD: Alzheimer’s disease, MCI: Mild Cognitive Impairment

*: χ²-test
#: MCI group had a significantly lower score than the normal group
†: AD group had a significantly lower score than the normal group
‡: AD group had a significantly lower score than the MCI group
±: the figure next means standard deviation

### Table 2. Results of three memory tests

<table>
<thead>
<tr>
<th></th>
<th>AD</th>
<th>MCI</th>
<th>normal</th>
<th>F (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-words learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>immediate recall</td>
<td>4.8±1.1</td>
<td>6.4±0.9</td>
<td>8.5±0.9</td>
<td>112.21 (p&lt;.001) #†‡</td>
</tr>
<tr>
<td>delayed recall</td>
<td>0.3±0.7</td>
<td>3.8±2.0</td>
<td>8.5±1.4</td>
<td>258.05 (p&lt;.001) #†‡</td>
</tr>
<tr>
<td>story recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>immediate recall</td>
<td>4.8±1.1</td>
<td>6.4±0.9</td>
<td>11.3±2.0</td>
<td>52.98 (p&lt;.001) #†‡</td>
</tr>
<tr>
<td>delayed recall</td>
<td>0.2±0.7</td>
<td>3.3±2.5</td>
<td>10.5±2.3</td>
<td>216.47 (p&lt;.001) #†‡</td>
</tr>
<tr>
<td>ROCFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>copy</td>
<td>34.5±2.2</td>
<td>34.8±1.8</td>
<td>35.3±1.1</td>
<td>1.85 (p=.163)</td>
</tr>
<tr>
<td>immediate recall</td>
<td>7.8±3.2</td>
<td>12.5±6.5</td>
<td>24.3±4.2</td>
<td>97.9 (p&lt;.001) #†‡</td>
</tr>
<tr>
<td>delayed recall</td>
<td>3.5±3.5</td>
<td>10.6±6.4</td>
<td>23.5±5.3</td>
<td>114.31 (p&lt;.001) #†‡</td>
</tr>
</tbody>
</table>

Each figure is the mean value of each condition of each subject group; ±: the figure next means standard deviation

*: the mean value of the mean of immediate recall scores of 5 trials.
AD: Alzheimer’s disease (n=30), MCI: Mild Cognitive Impairment (n=29), normal (n=31)
ROCFT: Rey-Osterrieth’s complex figure test

#: MCI group had a significantly lower score than the normal group
†: AD group had a significantly lower score than the normal group
‡: AD group had a significantly lower score than the MCI group
Figure 1: Frequency distributions of individual test scores of three memory tests for patients of Alzheimer’s disease (AD), Mild cognitive impairment (MCI), and the normal old (normal)

Each frequency represents the number of subjects with the same score or the same range of scores in the Figures A, B, C

AD: Alzheimer’s disease (n=30), MCI: Mild Cognitive Impairment (n=29), normal (n=31), ROCFT: Rey-Osterrieth's complex figure test
memory and lost daily executive functions.

The conversion rate of MCI in this study is 66%~85% within 10 years. This figure is rather large number or compatible with other studies. The result is restricted to only 26 of 30 AD patients; 8 patients continued to visit this memory clinic. 18 patients were hospitalized for worsening dementia with the neuropsychological test results and my letter of introduction. Four died of various physical diseases after 3-8 years from first visit.

Normal senior were most uncertain in this type of memory clinic.

Of 31 normal subjects, one was converted to AD in spite of good test scores at first visit. 12 continued to visit this memory clinic, and no one is converted to either AD or MCI. 18 normal subjects did not continue to visit our memory clinic. So we were not able to obtain additional reliable information.

The results of detailed memory tests revealed some characteristics of cognitive changes of those three groups of AD, MCI, and normal senior. Those are given in Table 2.

For all three memory tests (10-words list learning, Story recall, ROCFT) carried out at first visit in the manner with both immediate recall and delayed recall, there were significant differences between AD, MCI, and normal subjects (p<0.05). Only the copy test of ROCFT showed no significant difference in the Table 2. This copy test is to assess the visuospatial constructional cognitive functions. So this means that there is no serious decline about the function in the stage of probable AD during the progression of the disease.

The result of three delayed recall was especially interesting because not only the mean of the test score...
of each group is remarkably separated, but also the frequency (number of subjects of the same score) distribution of each group is fairly separated. This is given in Figure 1. For all three tests, the distribution of AD and normal are separated and there is no overlap. Using this condition, all three delayed recall test score can discriminate AD and normal with high sensitivity (100%) and specificity (100%). This also leads to the high positive predictive value (PPV) and negative predictive value (NPV) (both 100%) at the base rate (the prevalence rate of AD) of 5-20% for all three tests. This is given in Table 3.

This tendency is the same between MCI and normal. As shown in Figure 1, the frequency distribution of MCI and normal has a considerable overlap between them. Therefore, each test has relatively low sensitivity and specificity. Therefore, PPV and NPV are varied according to the kind of test and the base rate.

However, using a combination of three tests score, there is a considerable improvement of those values18. This is given in Table 4.

This combination realized the high specificity (100%), so that PPV is very high(100%) at the base rate of 5-20%.

Discussion

If we use these three delayed recall tests that we introduced in this article, the screening power to distinguish AD from senior normal is considerably high compared with the other short screening battery such as A 7-Minute Neurocognitive Screening Battery19 that shows that PPV (55%) and NPV (99%) at the base rate of 5% and PPV (85%) and NPV (98%) at the base rate of 20%. We introduced 100% PPV and NPV in Table 3. If we try to distinguish MCI from normal, the screening power to distinguish is 100% PPV at the base rate of 5-20% with the combination battery of 3 delayed recall tests.

We can clinically obtain more useful screening battery for MCI with high positive predictive values. And NPV values is less than PPV in this combination, NPV value is more than 95%. This is given in Table 4. Those are clinically useful tools if the base rate of MCI or AD varies from 5% to 20%. These results seem to be the best ones when compared to most other screening tests already reported18, 19, 20.

The base rate of 5% means the general senior population. This case meets the condition of a general population screening. Clinicians can easily detect suspicious candidates of MCI, and guide them to the appropriate medical treatment. The base rate of 20% meets the need of doctors whose specialty is dementia and who daily treat dementia patients. This test might provide relief to those who only appear demented or those who are anxious about becoming demented. The clinician might give that patient peace of mind with a high figure of negative predictive value. On the other hand, those specialists are able to detect patients with other better examinations for specialists in addition to these screening tests, even in such similar situations.

Although this study gave very high predictive values, there are several limitations in this study and results. First, at the first diagnosis a very rigid exclusion criterion was adopted for the examinations, participants have no mental and neurological illness, and passed the extensively laboratory examination, and there were no problematic cerebrovasucular lesions in the neuroradiological investigation. In that condition, we might miss many patients with high risks of dementia who had various neurological disease, hypertension, diabetes or depression. So we might see only brain degradation itself with age. The number of sample subjects is limited to only 90 is the second limitation. However, this result provided a physically normal senior who is free from any disease progress his brain ageing in such a way. We can examine many phenomena from him about the pure ageing process of brain.

In addition, this neuropsychological test battery requires more than 1 hour. We must try to develop a screening battery with less time to use.

Acknowledgements

I wish to thank Megumi Ueda, Faculty of Medical Science for Health, Teikyo Heisei University, and Yoshie Koyama, Department of Communication Science and Welfare, Faculty of Health and Welfare, Prefectural University of Hiroshima, for helping with the 10-year neuropsychological testing of so many patients in the Memory Clinic

References

3. Tombaugh TN, McIntyre NJ. The mini-mental state