

# Sensitivity and specificity of neuropsychological tests for dementia and mild cognitive impairment in a sample of residential elderly in South Africa

S Ramlall,<sup>1</sup> MB ChB, FCPsych (SA), PhD; J Chipps,<sup>2</sup> BSc Hons Nurs, BSc Hons Psych, MPH, PhD;  
A I Bhigjee,<sup>3</sup> MB ChB, MRCP, FCP (SA), MMed, MD, FRCP; B J Pillay,<sup>4</sup> BA Hons Psych, MA Clin Psych, PhD

<sup>1</sup> Department of Psychiatry, School of Clinical Medicine, University of KwaZulu-Natal, Durban, South Africa

<sup>2</sup> School of Nursing, Faculty of Community and Health Science, University of the Western Cape, Cape Town, South Africa

<sup>3</sup> Department of Neurology, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

<sup>4</sup> Department of Behavioural Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

Corresponding author: S Ramlall (ramlalls4@ukzn.ac.za)

**Background.** Neuropsychological tests can successfully distinguish between healthy elderly persons and those with clinically significant cognitive impairment.

**Objectives.** A battery of neuropsychological tests was evaluated for their discrimination validity of cognitive impairment in a group of elderly persons in Durban, South Africa.

**Method.** A sample of 117 English-speaking participants of different race groups (9 with dementia, 30 with mild cognitive impairment (MCI) and 78 controls) from a group of residential homes for the elderly was administered a battery of 11 neuropsychological tests. Kruskal-Wallis independent sample tests were used to compare performance of tests in the groups. Sensitivity and specificity of the tests for dementia and MCI were determined using random operating curve (ROC) analysis.

**Results.** Most tests were able to discriminate between participants with dementia or MCI, and controls ( $p < 0.05$ ). Area under the curve (AUC) values for dementia v. non-dementia participants ranged from 0.519 for the digit span (forward) to 0.828 for the digit symbol (90 s), with 14 of the 29 test scores achieving significance ( $p < 0.05$ ). AUC values for MCI participants ranged from 0.754 for controlled oral word association test (COWAT) Animal to 0.507 for the Rey complex figure test copy, with 17 of the 29 scores achieving significance ( $p < 0.05$ ).

**Conclusions.** Several measures from the neuropsychological battery had discrimination validity for the differential diagnosis of cognitive disturbances in the elderly. Further studies are needed to assess the effect of culture and language on the appropriateness of the tests for different populations.

*S Afr J Psychiatr* 2014;20(4):153-159. DOI:10.7196/SAJP.558



A growing number of people are surviving into old age, with an associated increase in the prevalence of dementia.<sup>[1]</sup> It has been postulated that up to 80% of cases remain undiagnosed,<sup>[2]</sup> which has resulted in a call for dementia to be regarded as a global health priority.<sup>[3]</sup>

The accurate detection of cognitive deficits due to dementia and mild cognitive impairment (MCI), a prodementia stage, is important to distinguish these from normal age-associated cognitive deficits. Delaying the progression from MCI to dementia by even 1 year has been shown to result in significant cost savings.<sup>[4]</sup>

The absence of reliable, universally acceptable biological and radiological markers for dementia necessitates the reliance on clinical assessments for a diagnosis,<sup>[5]</sup> supported by the assessment of cognitive disturbances using a range of screening tests. The use of neuropsychological screening tests allows for the assessment of specific cognitive domains, can distinguish age-related cognitive deficits from those due to MCI or dementia, and is superior to brief cognitive tools for which floor and ceiling effects threaten their

validity.<sup>[6,7]</sup> There is currently no universally accepted battery of tests for MCI<sup>[8,9]</sup> and the functioning of neuropsychological tests in low-resource residential settings in South Africa (SA) has not been widely tested.

The objective of this study was to determine the sensitivity and specificity of a battery of neuropsychological tests in a sample of elderly persons living in a residential setting in SA.

## Methods

The research was conducted from 2010 to 2011 on residents aged  $\geq 60$  years ( $N=1\ 371$ ) in a group of residential homes in Durban, KwaZulu-Natal Province, SA. The homes are administered by a non-governmental organisation and cater for those needing frail care, assisted living and independent living. Inclusion criteria for the study were: randomly selected residents  $\geq 60$  years; with a minimum of 8 years of formal schooling; the ability to speak, read and write in English; and the ability to give written, informed consent. Exclusion criteria were severe physical, mental or sensory handicaps that precluded

engagement in the assessment procedures. The study comprised three stages of cognitive assessments, conducted sequentially. In stage 1, a random sample ( $n=302$ ) of residents aged  $\geq 60$  years was selected based on a statistical calculation from all the homes. The Mini-Mental State Examination (MMSE),<sup>[10]</sup> a subjective memory rating scale,<sup>[11]</sup> and an informant questionnaire, the Deterioration Cognitive Observée (DECO),<sup>[12]</sup> were used to screen for dementia.<sup>[13]</sup> The sample size was calculated on a sensitivity and specificity of 85% for the MMSE<sup>[14]</sup> and a conservative estimate of 20% prevalence for dementia in residential homes, based on the reported ranges of 16<sup>[15]</sup> - 75%.<sup>[16]</sup> In stage 2, 140 participants from stage one, including MMSE screen positives and a random selection of MMSE screen negatives, were clinically assessed by psychiatrists and categorised as having dementia, MCI or not being clinically cognitively impaired.<sup>[17]</sup> *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (text revised) criteria A and B for Alzheimer's and vascular dementia were used to assign a general diagnosis of dementia without reference to aetiology.<sup>[18]</sup> MCI was diagnosed using the criteria contained in the report of the International Working Group on mild cognitive impairment.<sup>[19]</sup> In stage 3, the 140 participants from stage 2 were available and consented to the administration of a neuropsychological battery of tests. Of these, 2 participants died during the study and 20 either refused or were unavailable to participate. One person was unable to complete any of the neuropsychological tests and was excluded from the dataset. Of the remaining 117 participants, 9 were participants with a diagnosis of dementia, 30 with MCI and 78 participants did not meet diagnostic criteria and were classified as controls.

The overall study was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal, with residents providing written, informed consent to participate.

### Neuropsychological tests

Eleven neuropsychological measures (including 29 tests) were administered by clinical psychologists in English in a single session at the participants' residences. The tests were: Rey auditory verbal learning tests (RAVLTs),<sup>[20]</sup> digit span<sup>[20]</sup> and digit symbol;<sup>[20]</sup> controlled oral word association test (COWAT F-A-S and

Animal);<sup>[20]</sup> short story comprehension and recall (using an SA adaptation of the cowboy story, i.e. 'A Farmer from Transkei');<sup>[20]</sup> the token test (short version);<sup>[20]</sup> Rey complex figure (RCF);<sup>[20]</sup> trail making tests A and B (TMT-A and -B);<sup>[20]</sup> the clock drawing test<sup>[21]</sup> (the free-drawing version with the '10 past 11' time setting instruction using Rouleau's 10-point scoring system);<sup>[22]</sup> the Luria hand sequence;<sup>[20]</sup> and the Maze test.<sup>[20]</sup> The psychologists were blind to the participants' cognitive performances in the previous phases of the study.

### Data analysis

Data were analysed using IBM SPSS (version 21.0) and MedCalc (version 12.5.0). Descriptive statistics were calculated for demographic variables, and mean neuropsychological scores and 95% confidence intervals (CIs) were calculated for each of the classified groups. Between-group comparisons were undertaken using non-parametric Kruskal-Wallis independent samples tests and Pearson's  $\chi^2$  tests or Fisher's exact tests where appropriate. Between-group tests were not analysed by race due to the low number of black participants ( $n=4$ ). Individual test random operating curve (ROC) analysis was conducted for discrimination validity and diagnostic accuracy of the different tests for dementia ( $n=9$ ) v. non-dementia participants (controls and MCI cases;  $n=108$ ) and for controls ( $n=78$ ) and MCI participants ( $n=30$ ), respectively. For each test, the area under the curve (AUC) was calculated with 95% CIs. Swet's interpretation of AUC scores was used in this study: 0.5 = non-informative;  $0.5 < \text{AUC} \leq 0.7$  =

less accurate;  $0.7 < \text{AUC} \leq 0.9$  = moderately accurate;  $0.9 < \text{AUC} < 1$  = highly accurate; and the perfect test has an AUC = 1.<sup>[23]</sup> Optimal cut-off scores based on the Youden index and the associated sensitivity and specificity values were generated and ranked as significant tests for dementia and for MCI identification, respectively.

### Results

The average years of education of the 117 participants was 10.1 (standard deviation (SD) 2.2) years, with 87 (74.4%) reporting 8 - 11 years of formal education (Table 1). English was the first language for 103 (88.0%) of the participants, followed by Afrikaans ( $n=6$ ; 5.1%), isiZulu ( $n=4$ ; 3.4%), and other languages ( $n=4$ ; 3.4%). Most of the participants were female ( $n=82$ ; 70.1%). The mean (SD) age was 74.2 (7.5) years, with significant differences between the three groups by race and age (Table 1).

### Cognitive deficits by diagnostic groups and controls

Across the three classification categories, the differences in score means of the RAVLT (total), digit symbol (90 s), and COWAT Animal were highly significant, with  $p < 0.001$ . With the exception of the digit span (forward), digit span total, COWAT A, narrative memory test (delayed recall), token test and the Luria hand sequence test, there were significant mean differences between all other tests in the three groups (Table 2). The mean score on most tests demonstrated a progressive declining pattern in cognitive performance

**Table 1. Demographic data per diagnostic group**

Demographics	Controls ( $n=78$ )	MCI ( $n=30$ )	Dementia ( $n=9$ )	Statistic	<i>p</i> -value
Age (years), mean (SD)*	72.1 (6.7)	76.4 (8.4)	79.0 (7.5)	$K=10.1$	0.006 <sup>†</sup>
Education (years), mean (SD)*	10.3 (2.2)	9.3 (1.6)	10.1 (2.2)	$K=5.8$	0.055
Gender, <i>n</i> (%) <sup>‡</sup>					
Female	51 (65.4)	23 (76.7)	8 (88.9)	$\chi^2=2.6$	0.260
Male	27 (34.6)	7 (23.3)	1 (11.1)		
Race, <i>n</i> (%) <sup>‡</sup>					
Asian	18 (23.1)	5 (16.7)	0 (0)	$\chi^2=12.5$	0.015*
Black	2 (2.6)	1 (3.3)	1 (11.1)		
Coloured	22 (28.2)	13 (43.3)	0 (0)		
White	36 (46.2)	11 (36.7)	8 (88.9)		

\*Age and years of education were compared using independent samples (Kruskal-Wallis test).

<sup>†</sup>Significance was set at  $p < 0.05$ .

<sup>‡</sup>Gender and race were compared using Pearson  $\chi^2$  and Fisher's exact tests where appropriate.

**Table 2. Mean test scores per diagnostic groups and controls**

Test	Control mean (95% CI) (n=78)	MCI mean (95% CI) (n=30)	Dementia mean (95% CI) (n=9)	K-statistic*	p-value
RAVLT					
Trial I	4.8 (4.4 - 5.2)	4.2 (3.4 - 4.9)	3.3 (1.9 - 4.8)	6.6	0.037 <sup>†</sup>
Trial II	6.8 (6.2 - 7.4)	5.8 (5.1 - 6.6)	4.6 (3.4 - 5.7)	10.2	0.006 <sup>†</sup>
Trial III	7.9 (7.3 - 8.5)	6.9 (5.9 - 7.8)	5.2 (4.0 - 6.5)	11.2	0.004 <sup>†</sup>
Trial IV	8.6 (8.0 - 9.3)	7.0 (6.1 - 8.0)	5.6 (3.8 - 7.3)	14.2	0.001 <sup>†</sup>
Trial V	9.1 (8.5 - 9.8)	7.7 (6.7 - 8.8)	5.4 (3.6 - 7.30)	14.5	0.001 <sup>†</sup>
Trials I - V	37.3 (34.8 - 39.8)	31.6 (27.6 - 35.6)	24.1 (18.8 - 29.5)	15.4	<0.001 <sup>†</sup>
Immediate recall	6.6 (5.8 - 7.5)	5.5 (4.4 - 6.7)	3.1 (0.9 - 5.3)	9.6	0.008 <sup>†</sup>
20-min recall	6.0 (5.1 - 6.0)	5.0 (3.9 - 6.1)	2.6 (0.5 - 4.70)	6.9	0.032 <sup>†</sup>
Digit span					
Forward	9.0 (8.5 - 9.5)	7.8 (7.0 - 8.6)	8.4 (7.0 - 9.9)	5.9	0.052
Backward	5.0 (4.6 - 5.5)	4.2 (3.6 - 4.9)	3.8 (2.9 - 4.7)	7.2	0.028 <sup>†</sup>
Total	14.0 (13.2 - 14.9)	12.2 (10.9 - 13.5)	12.2 (10.0 - 14.5)	5.4	0.066
Digit symbol					
90 s	25.9 (23.5 - 28.2)	20.0 (17.2 - 22.8)	12.9 (7.3 - 18.5)	17.0	<0.001 <sup>†</sup>
120 s	31.9 (28.3 - 35.5)	25.5 (21.9 - 29.1)	13.9 (3.5 - 24.3)	13.4	0.001 <sup>†</sup>
COWAT					
F	6.8 (5.9 - 7.7)	4.6 (3.4 - 5.7)	5.1 (2.8 - 7.5)	9.3	0.010 <sup>†</sup>
A	4.2 (3.5 - 4.9)	3.0 (2.0 - 4.0)	3.8 (1.8 - 5.7)	4.5	0.102
S	7.6 (6.7 - 8.4)	5.4 (4.1 - 6.8)*	5.5 (3.2 - 7.8)	9.1	0.011 <sup>†</sup>
Total	18.5 (16.4 - 20.7)	12.4 (9.5 - 15.4)	14.4 (8.6 - 20.1)	12.2	0.002 <sup>†</sup>
Animal	11.9 (10.9 - 12.8)	8.1 (6.7 - 9.6)	6.5 (2.8 - 10.2)	22.2	<0.001 <sup>†</sup>
Narrative recall	8.1 (7.0 - 9.2)	7.1 (5.9 - 8.3)	4.4 (1.4 - 7.5)	6.2	0.045 <sup>†</sup>
Narrative delayed recall	6.8 (5.8 - 7.9)	5.8 (4.4 - 7.3)	3.6 (0.3 - 6.8)	4.8	0.090
Token test	162.7 (159.2 - 166.2)	153.0 (143.9 - 162.1)	147.1 (124.5 - 169.7)	5.5	0.064
RCF					
Copy	30.7 (29.2 - 32.2)	40.4 (19.1 - 61.7)	21.2 (12.6 - 29.7)	8.2	0.017 <sup>†</sup>
Recall	13.5 (11.8 - 15.3)	10.7 (8.7 - 12.7)	7.7 (2.3 - 13.1)	7.1	0.028 <sup>†</sup>
Delayed	14.1 (12.5 - 15.8)	11.0 (8.9 - 13.2)	6.3 (0.7 - 11.9)	11.6	0.003 <sup>†</sup>
Clock drawing	8.7 (8.4 - 9.1)	8.5 (7.7 - 9.3)	6.7 (4.5 - 8.9)	6.1	0.047 <sup>†</sup>
Luria hand sequence	2.5 (2.2 - 2.8)	2.3 (1.8 - 2.9)	1.9 (0.2 - 3.5)	1.3	0.530
TMT					
A	70.8 (62.2 - 79.4)	82.4 (70.8 - 4.1)	86.3 (60.6 - 11.9)	8.4	0.015 <sup>†</sup>
B	154.6 (138.2 - 171.0)	195.1 (167.6 - 222.6)	290.1 (105.1 - 475.1)	10.4	0.005 <sup>†</sup>
Maze total	385.1 (344.1 - 426.1)	521.7 (440.7 - 602.7)	571.1 (312.3 - 830.0)	10.5	0.005 <sup>†</sup>

CI = confidence interval; RAVLTs = Rey auditory verbal learning tests; COWAT = controlled oral word association test; RCF = Rey complex figure; TMT = trail making test.

\*Tests were compared using independent sample Kruskal-Wallis (K) tests.

<sup>†</sup>Significance level was set at  $p < 0.05$  and 95% CIs.

from the control group to MCI to dementia subjects. The exceptions to this pattern were on digit span total (no difference in mean scores between MCI (12.2) and dementia (12.2) ( $p=0.066$ ) and the COWAT group, where the mean dementia group scores were slightly better than the MCI group scores, with COWAT F (5.5 v. 4.6;  $p=0.010$ ), COWAT A (3.8 v. 3.0;  $p=0.102$ ), COWAT S (5.5 v. 5.4;  $p=0.011$ ) and COWAT Total (14.4 v. 12.4;  $p=0.002$ ). Similarly, for the RCF (copy), the MCI score (40.4) was better than that of the controls (30.7) ( $p=0.017$ ).

#### Discrimination validity: Dementia v. non-dementia participants

Using individual ROC curves to measure how well the tests discriminated between participants with dementia and without dementia (participants with MCI and controls), AUCs ranged from 0.519 for digit span (forward) to 0.828 for digit symbol (90 s). Fifteen of the 29 tests achieved a significant difference, with  $p < 0.001$  demonstrated on the RAVLT trial II (AUC = 0.753), RAVLT

**Table 3. ROC analysis for tests for dementia v. non-dementia**

Test	AUC (95% CI)	p-value	Cut-off	Sensitivity	Specificity
<b>RAVLT</b>					
Trial I	0.670 (0.577 - 0.754)	0.095	≤2	44.4	88.9
Trial II	0.753 (0.665 - 0.828)	<0.001*	≤5	77.8	61.1
Trial III	0.775 (0.688 - 0.847)	<0.001*	≤7	100	50.9
Trial IV	0.757 (0.669 - 0.832)	<0.001*	≤7	88.0	57.6
Trial V	0.812 (0.729 - 0.878)	<0.001*	≤7	88.9	61.1
Total trials I - V	0.805 (0.721 - 0.872)	<0.001*	≤34	100	50.0
Immediate recall	0.770 (0.683 - 0.842)	<0.001*	≤4	77.8	68.5
20-min recall	0.741 (0.652 - 0.817)	<0.001*	≤5	88.9	51.9
<b>Digit span</b>					
Forward	0.519 (0.424 - 0.612)	0.835	≤8	66.7	53.7
Backward	0.646 (0.552 - 0.732)	0.094	≤5	100.0	24.1
Total	0.549 (0.454 - 0.641)	0.600	≤15	100.0	25.0
<b>Digit symbol</b>					
90 s	0.828 (0.746 - 0.891)	<0.001*	≤15	87.5	80.6
120 s	0.804 (0.720 - 0.872)	<0.001*	≤20	87.5	75.0
<b>COWAT</b>					
F	0.561 (0.466 - 0.654)	0.557	≤4	50.0	56.4
A	0.530 (0.434 - 0.624)	0.787	>4	50.0	69.2
S	0.595 (0.500 - 0.685)	0.308	≤9	100	26.9
Total	0.552 (0.456 - 0.644)	0.622	≤23	100	20.4
Animal	0.763 (0.675 - 0.837)	0.017*	≤4	50.0	96.3
Narrative recall	0.736 (0.646 - 0.813)	0.011*	≤6.0	88.9	62.0
Narrative delayed recall	0.693 (0.601 - 0.775)	0.050	≤0	44.4	88.0
Token test total	0.650 (0.557 - 0.736)	0.180	≥162	77.8	63.0
<b>RCF</b>					
Copy	0.783 (0.698 - 0.854)	<0.001*	≤21	55.6	89.8
Recall	0.709 (0.618 - 0.790)	0.039*	≤7	66.7	75.9
Delayed	0.781 (0.696 - 0.852)	0.005*	≤8	77.8	76.9
Clock drawing	0.732 (0.642 - 0.810)	0.012*	≤5	44.4	91.7
Luria hand sequence	0.600 (0.505 - 0.690)	0.448	≤1	50.0	75.0
<b>TMT</b>					
A	0.669 (0.575 - 0.754)	0.141	>72	75.0	62.6
B	0.669 (0.575 - 0.754)	0.143	>210	62.5	76.4
Maze total	0.641 (0.546 - 0.728)	0.130	>329	100.0	34.3

ROC = random operating curve; AUC = area under the curve; CI = confidence interval; RAVLT = Rey auditory verbal learning test; COWAT = controlled oral word association test; RCF = Rey complex figure; TMT = trail making test.  
\*Significance level set as  $p < 0.05$  and 95% CIs.

trial III (AUC = 0.775), RAVLT trial V (AUC = 0.812), RAVLT total (AUC = 0.805), RAVLT immediate recall (AUC = 0.770), RAVLT 20 min recall (AUC = 0.741), digit symbol (90 s) (AUC = 0.828), digit symbol (120 s) (AUC = 0.804) and RCF copy (AUC = 0.783) (Table 3).

The 15 significant tests were ranked, based on optimum balance between sensitivity and

specificity (Fig. 1). The most balanced test was the digit symbol, with a sensitivity of 87.5% and specificities of 80.6% at 90 s and 75.0% at 120 s, with cut-off scores of <15 and <20, respectively. Although the RAVLT trial III and the RAVLT total had the best discrimination sensitivity (100% at cut-off scores <7 and <34), this was at the expense of their specificities, which were 50.0% and 50.9%, respectively.

### Discrimination validity: MCI v. control participants

Using individual ROC curves to measure how well the tests discriminated between participants with MCI and controls, 17 of the 29 tests achieved a significant AUC ( $p < 0.05$ ) (Table 4). AUCs ranged from 0.621 to 0.754, with significance levels of  $p < 0.001$  on the digit symbol (90 s), COWAT F, COWAT Total, and COWAT Animal.

The 17 significant tests were ranked, based on optimum balance between sensitivity and specificity (Fig. 2). The most balanced subtest score, the RCF (delayed recall), had a sensitivity of 77.8% and specificity of 76.9%, at a cut-off score of ≤14. The highest sensitivity reported was 93.3%, which was for digit span (backwards) and digit symbol (120 s), at cut-off scores of ≤5 and ≤36, respectively, with the highest specificity being 83.3% (Maze total), at a cut-off score of >544.

### Discussion

This research sought to determine the sensitivity and specificity of a battery of neuropsychological tests in elderly participants from a low-resource residential setting, who were diagnosed with MCI and dementia. While the tests used have been widely researched, to our knowledge, this is the first time that their diagnostic discriminability was evaluated in a heterogeneous elderly SA population. The findings are discussed in terms of screening for overall cognitive decline, dementia and MCI.

### Screening for overall pathological cognitive decline

With the exception of recall on the narrative memory test and the token test, all the tests were able to significantly discriminate between controls and those with clinically significant cognitive impairment (dementia or MCI). The token test was found to be of little value, in keeping with previous research that found that the token test ceiling effects limited its utility.<sup>[24]</sup>

The data from this research supported the conclusion that the clock drawing test has value for screening moderate-to-severe cognitive impairment, but is relatively poor at detecting milder forms of cognitive impairment.<sup>[25]</sup> The clock drawing test has been described as the ideal cognitive screening test owing to its ease of administration

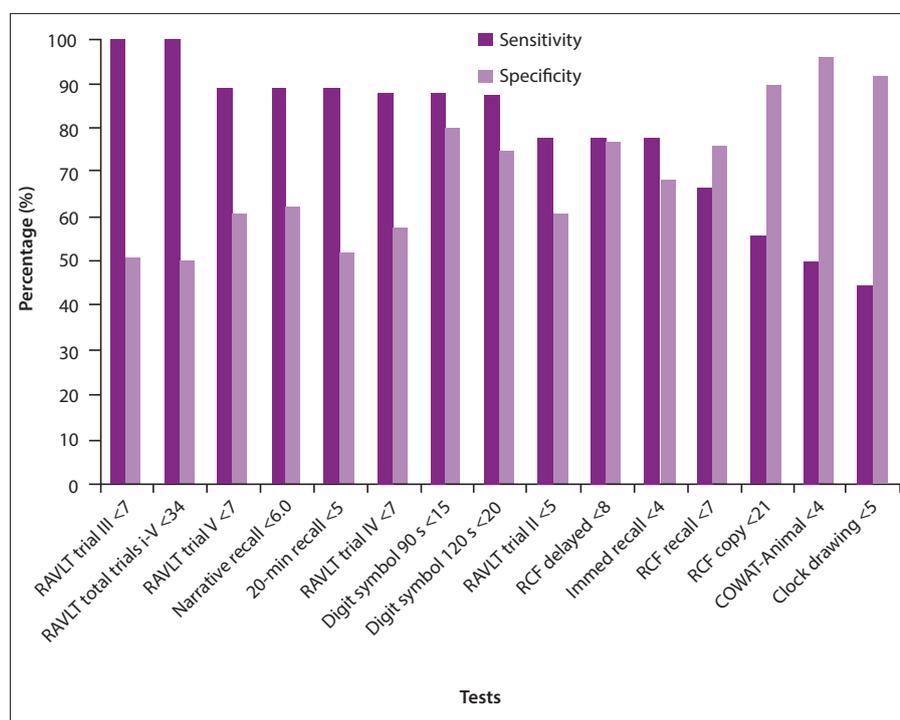


Fig. 1. Optimal sensitivity and specificity of tests for dementia v. non-dementia. (RAVLT = Rey auditory verbal learning tests; RCF = Rey complex figure; COWAT = controlled oral word association test.)

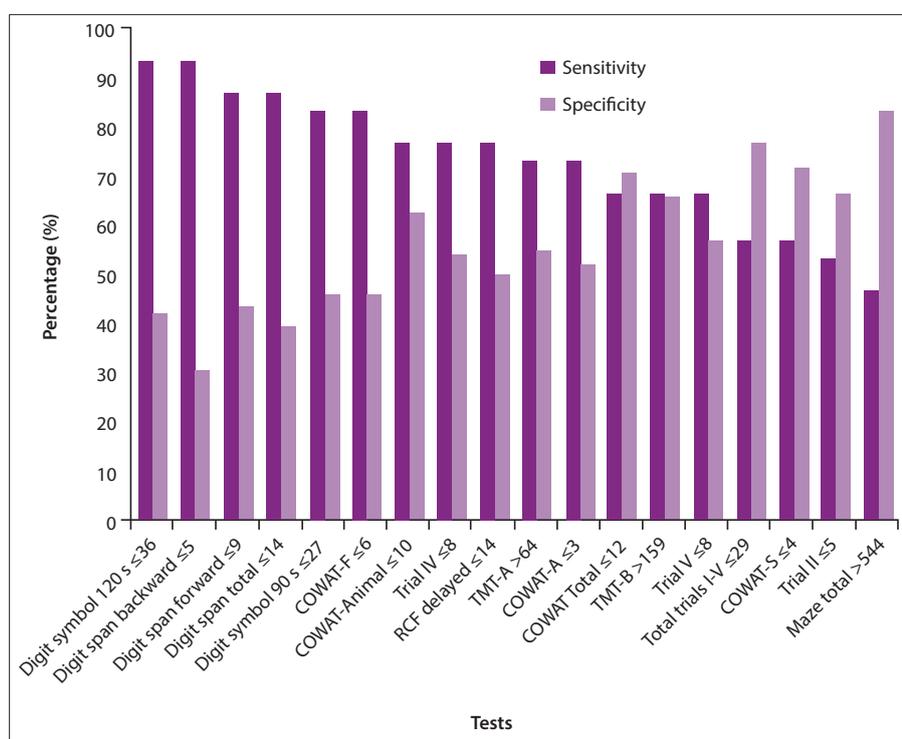


Fig. 2. Optimal sensitivity and specificity of tests for MCI v. controls. (COWAT = controlled oral word association test; RCF = Rey complex figure; TMT = trail making test; MCI = mild cognitive impairment.)

and scoring, its ability to assess a range of cognitive abilities and good psychometric properties.<sup>[26,27]</sup> While it has shown good correlation with the MMSE,<sup>[26]</sup> it has also

shown moderate sensitivity and specificity in detecting executive dysfunction in patients who have a normal MMSE.<sup>[28]</sup> In this study, the clock drawing test, with an AUC of

0.732, supported its utility as a dementia assessment<sup>[27]</sup> and confirmed earlier findings of its ability to differentiate normal from pathological cognitive decline.<sup>[29]</sup> However, the sensitivity of 44.4% in this study was much lower than the mean of 85% reported in the literature for dementia screening, while the specificity of 91.7% compared favourably with the specificity of 85% reported in the literature.<sup>[26]</sup>

### Screening for dementia

Memory disturbances are one of the most common cognitive complaints in the elderly, and can be attributed to either normal decline associated with ageing or dementia.<sup>[30]</sup> In this research, the RAVLTs demonstrated the best discrimination validity for dementia, thus confirming their utility in diagnosing the disorder. With the exception of trial I of the RAVLTs, all the RAVLT subtest measures displayed significance at  $p < 0.001$ . These tests are brief tests of memory function that are easy to administer and sensitive to encoding, storage and retrieval of memory.<sup>[31]</sup> This finding is consistent with those of previous studies, showing that the RAVLTs are useful in distinguishing participants with dementia from those in a control group.<sup>[32]</sup> Some studies have also shown that the RAVLTs are useful in distinguishing normal participants from those with dementia associated with Alzheimer's disease and vascular dementia,<sup>[32,33]</sup> and that the RAVLTs are also able to predict the conversion to dementia in those individuals with subjective memory complaints<sup>[32]</sup> and MCI.<sup>[33]</sup>

However, the most balanced test was found to be the digit symbol ( $p < 0.001$ ), which measures attention and working memory. The mean scores on the digit symbol (90 s) in the study were lower than those reported by Hart *et al.*,<sup>[33]</sup> which may be owing to the lower mean education level of participants; education level is reported to affect performance on the digit symbol test. These findings suggest that the memory deficits evident by the RAVLT scores may be attributable to storage and retrieval difficulties, as the digit symbol test is more demanding of attention and concentration consistent with the graded deterioration of cognitive functions in dementia.

Two other tests useful for screening of dementia and consistent with other studies were the RCF<sup>[34]</sup> and COWAT Animal.<sup>[35,36]</sup>

**Table 4. ROC analysis tests for MCI v. controls**

Test	AUC (95% CI)	p-value	Cut-off	Sensitivity	Specificity
RAVLT					
Trial I	0.618 (0.519 - 0.709)	0.061	≤4	70.0	53.9
Trial II	0.621 (0.523 - 0.713)	0.037*	≤5	53.3	66.7
Trial III	0.618 (0.520 - 0.710)	0.051	≤7	63.3	56.4
Trial IV	0.671 (0.574 - 0.759)	0.002*	≤8	76.7	53.9
Trial V	0.639 (0.541 - 0.729)	0.019*	≤8	66.7	56.4
Total trials I - V	0.657 (0.559 - 0.745)	0.009*	≤29	56.7	76.9
Immediate recall	0.599 (0.501 - 0.692)	0.097	≤7	76.7	42.3
20-min recall	0.568 (0.470 - 0.663)	0.229	≤7	86.7	37.2
Digital span					
Forward	0.646 (0.549 - 0.736)	0.011*	≤9	86.7	43.6
Backward	0.635 (0.537 - 0.726)	0.014*	≤5	93.3	30.8
Total	0.638 (0.540 - 0.728)	0.016*	≤14	86.7	39.7
Digit symbol					
90 s	0.677 (0.580 - 0.764)	<0.001*	≤27	83.3	46.2
120 s	0.649 (0.551 - 0.738)	0.005*	≤36	93.3	42.3
COWAT					
F	0.685 (0.588 - 0.771)	<0.001*	≤6	83.3	45.5
A	0.631 (0.532 - 0.722)	0.027*	≤3	73.3	52.0
S	0.676 (0.579 - 0.763)	0.002*	≤4	56.7	71.8
Total	0.714 (0.619 - 0.797)	<0.001*	≤12	66.7	70.5
Animal	0.754 (0.662 - 0.832)	<0.001*	≤10	76.7	62.8
Narrative recall	0.557 (0.459 - 0.653)	0.307	≤12	96.7	23.1
Narrative delayed recall	0.566 (0.467 - 0.661)	0.251	≤8	80.0	39.7
Token test total	0.611 (0.512 - 0.703)	0.089	≤156	46.7	79.5
RCF					
Copy	0.507 (0.409 - 0.604)	0.918	>28	73.3	21.8
Recall	0.603 (0.505 - 0.696)	0.062	≤13	76.7	46.1
Delayed	0.631 (0.533 - 0.722)	0.023*	≤14	76.7	50.0
Clock drawing	0.506 (0.408 - 0.604)	0.922	≤5	16.7	94.9
Luria hand sequence	0.537 (0.438 - 0.633)	0.547	≤3	80.0	28.2
TMT					
A	0.655 (0.557 - 0.744)	0.006	>64	73.3	54.6
B	0.679 (0.581 - 0.766)	0.001*	>159	66.7	65.8
Maze total	0.685 (0.588 - 0.771)	0.002*	>544	46.7	83.3

ROC = random operating curve; AUC = area under the curve; CI = confidence interval; RAVLT = Rey auditory verbal learning tests; COWAT = controlled oral word association test; RCF = Rey complex figure; TMT = trail making test.

\*Significance level set at  $p < 0.05$  and 95% CIs.

The RCF findings were similar to a study that showed that it was able to distinguish patients with Alzheimer's disease and vascular dementia from those in the control group.<sup>[34]</sup> The COWAT semantic fluency ('animal') category displayed diagnostic significance for dementia ( $p=0.017$ ), whereas phonological fluency (FAS) did not. The superiority of semantic fluency over other

verbal fluency measures has been shown in SA<sup>[35]</sup> and elsewhere.<sup>[36]</sup> The animal category is also simple to administer and interpret, and could be valuable as a screening and/or diagnostic measure for use in low-income country settings.

Although the TMT is a popular neuropsychological test and is included in most test batteries, in this study the TMT did

not significantly discriminate dementia from non-dementia participants. However, this finding may be owing to the non-clinical sample in this study, as neuropsychological tests are considered to be less accurate in community samples.<sup>[29]</sup>

### Screening for MCI

MCI is widely regarded as an intermediate stage between Alzheimer's disease and normal ageing, and has a heterogeneous cognitive profile.<sup>[37,38]</sup> It is therefore recommended that an extensive range of domains be covered when assessing for MCI, including language, memory, executive functions and attention.<sup>[9,38]</sup> Several tests displayed significant diagnostic accuracy for MCI, namely digit symbol (90 s), COWAT F, FAS Total and Animal, and the TMT-B, with the highest sensitivity reported for digit span (backwards), digit symbol (120 s), and Maze total. The most balanced test was the RCF (delayed recall). The findings showed a similarity to the Goteborg MCI study,<sup>[38]</sup> with significantly low scores on memory, attention and working memory, and visuospatial and executive functions.

Compared with the profile of tests that were significant for dementia, the language tests (COWAT F-A-S, Total and Animal) and the working memory/attention tests (digit span) showed little overlap, suggesting that in this sample at least, MCI and dementia may represent different clinical entities rather than MCI being a milder or earlier stage of dementia. Our findings therefore validate MCI as a distinct clinical entity that is distinguishable, with neuropsychological testing, from age-related and dementia-associated cognitive decline.

### Study limitations

The lack of stratification of the sample per demographic variables reviewed (race, gender, age, education), the small number of dementia cases and the specificity of the residential setting limited the generalisability of our findings. Though the sample reflected the population of the residential homes, it did not represent all race groups in the SA community adequately. In view of the confounding effects of culture on test performance, it is recommended that this research is repeated with a large sample of black African participants. As the ability to speak English was required for eligibility to

participate in the study, the confounding effects of language cannot be excluded, as English was not the first language of all participants. Further, the discrimination capacity of the tests for dementia may have been diminished by the inclusion of MCI participants in the comparison group, inflating the mean test scores of the non-dementia group.

## Conclusion

A recommended neuropsychological test battery can be used effectively either to screen for or discriminate between early and later stages of cognitive impairment in the elderly. However, in line with the Jacova<sup>[7]</sup>'s evidence-based review, it is recommended that neuropsychological tests should not be used alone for diagnostic purposes; they should be part of a clinically integrative process, used selectively to aid in distinguishing normal age-related cognition from MCI and early dementia, for the differential diagnosis of cognitive impairment and possibly for assessing the risk of progression from MCI to dementia.<sup>[7]</sup>

## References

- Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: A systematic review and meta-analysis. *Alzheimers Dement* 2013;9(1):63. [http://dx.doi.org/10.1016/j.jalz.2012.11.007]
- Weimer DL, Sager MA. Early identification and treatment of Alzheimer's disease: Social and fiscal outcomes. *Alzheimers Dement* 2009;5(3):215-226. [http://dx.doi.org/10.1016/j.jalz.2009.01.028]
- World Health Organization. Dementia: A Public Health Priority. Switzerland: World Health Organization Press, 2012. [http://www.who.int/mental\\_health/publications/dementia\\_report\\_2012/en/](http://www.who.int/mental_health/publications/dementia_report_2012/en/) (accessed 10 January 2014).
- Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: Early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001;56(9):1133-1142. [http://dx.doi.org/10.1212/WNL.56.9.1133]
- Robillard A. Clinical diagnosis of dementia. *Alzheimers Dement* 2007;3(4):292-298. [http://dx.doi.org/10.1016/j.jalz.2007.08.002]
- Salmon DP, Bondi MW. Neuropsychological assessment of dementia. *Annu Rev Psychol* 2009;60(1):257-282. [http://dx.doi.org/10.1146/annurev.psych.57.102904.190024]
- Jacova C, Kertesz A, Blair M, Fisk JD, Feldman HH. Neuropsychological testing and assessment for dementia. *Alzheimers Dement* 2007;3(4):299-317. [http://dx.doi.org/10.1016/j.jalz.2007.07.011]
- Seo EH, Yoon JC, Jhoo JH, et al. Total scores of the CERAD neuropsychological assessment battery: Validation for mild cognitive impairment and dementia patients with diverse etiologies. *Am J Geriatr Psychiatry* 2010;18(9):801-809. [http://dx.doi.org/10.1097/JGP.0b013e3181cab764]
- Grundman M, Doody R, Kaye J, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol* 2004;61(1):59-66. [http://dx.doi.org/10.1001/archneur.61.1.59]
- Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12(3):189.
- Wang L, van Belle G, Crane PK, et al. Subjective memory deterioration and future dementia in people aged 65 and older. *J Am Geriatr Soc* 2004;52(12):2045-2051. [http://dx.doi.org/10.1111/j.1532-5415.2004.52568.x]
- Ritchie K, Fuhrer R. The validation of an informant screening test for irreversible cognitive decline in the elderly: Performance characteristics within a general population sample. *Int J Geriatr Psychiatry* 1996;11(2):149-156. [http://dx.doi.org/10.1016/0272-7358(95)00056-9]
- Ramlall S, Chipps J, Bhigjee AI, Pillay BJ. Screening a heterogeneous elderly South African population for cognitive impairment: The utility and performance of the Mini-Mental State Examination, Six Item Screener, Subjective Memory Rating Scale and Deterioration Cognitive Observed. *Afr J Psychiatry* 2013;16(6):445-455. [http://dx.doi.org/0.4314/ajpsy.v16i6.58]
- Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *J Psychiatr Res* 2009;43(4):411-431. [http://dx.doi.org/10.1016/j.jpsychires.2008.04.014]
- Alvarado-Esquivel C, Hernandez-Alvarado AB, Tapia-Rodriguez RO, Guerrero-Iturbe A, Rodriguez-Corral K, Martinez SE. Prevalence of dementia and Alzheimer's disease in elders of nursing homes and a senior center of Durango City, Mexico. *BMC Psychiatry* 2004;4:3. [http://dx.doi.org/10.1186/1471-244X-4-3]
- Matthews FE, Denning T. Prevalence of dementia in institutional care. *Lancet* 2002;360(9328):225-226. [http://dx.doi.org/10.1016/S0140-6736(02)09461-8]
- Ramlall S, Chipps J, Pillay BJ, Bhigjee AI. Mild cognitive impairment and dementia in a heterogeneous elderly population: Prevalence and risk profile. *Afr J Psychiatry* 2013;16(6):456-465.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. (text rev). Washington DC: American Psychiatric Publishing, Inc, 2000.
- Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment--beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* 2004;256(3):240-246. [http://dx.doi.org/10.1111/j.1365-2796.2004.01380.x]
- Lezak MD. *Neuropsychological Assessment*. Oxford: Oxford University Press, 2012.
- Freedman M. *Clock drawing: A neuropsychological analysis*. New York: Oxford University Press, 1994.
- Rouleau I, Salmon DP, Butters N. Longitudinal analysis of clock drawing in Alzheimer's disease patients. *Brain Cogn* 1996;31(1):17-34. [http://dx.doi.org/10.1006/brcg.1996.0022]
- Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240(4857):1285-1293. [http://dx.doi.org/10.1126/science.3287615]
- De Jager CA, Hogervorst E, Combrinck M, Budge MM. Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychol Med* 2003;33(6):1039-1050.
- Nishiwaki Y, Breeze E, Smeeth L, Bulpitt CJ, Peters R, Fletcher AE. Validity of the clock-drawing test as a screening tool for cognitive impairment in the elderly. *Am J Epidemiol* 2004;160(8):797-807. [http://dx.doi.org/10.1093/aje/kwh288]
- Shulman KI. Clock-drawing: is it the ideal cognitive screening test? *Int J Geriatr Psychiatry* 2000;15(6):548-561.
- Royall DR, Cordes JA, Polk M. CLOX: An executive clock drawing task. *J Neurol, Neurosurg Psychiatry* 1998;64(5):588-594.
- Juby A, Tench S, Baker V. The value of clock drawing in identifying executive cognitive dysfunction in people with a normal Mini-Mental State Examination score. *CMAJ* 2002;167(8):859-864.
- Cahn DA, Salmon DP, Monsch AU, et al. Screening for dementia of the Alzheimer type in the community: The utility of the clock drawing test. *Arch Clin Neuropsychol* 1996;11(6):529-539.
- Morris RG. Neuropsychological assessment in older people: Old principles and new directions. *Adv Psychiatr Treat* 2000;6(5):362-370. [http://dx.doi.org/10.1192/apt.6.5.362]
- Lezak MD, Howieson DB, Loring DW, eds. *Neuropsychological Assessment*. 4th ed. New York: Oxford University Press, 2004.
- Estévez-González A, Kuliseky J, Boltes A, Otermin P, García-Sánchez C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *Int J Geriatr Psychiatry* 2003;18(11):1021-1028. [http://dx.doi.org/10.1002/gps.1010]
- Hart RP, Kwentus JA, Wade JB, Hamer RM. Digit symbol performance in mild dementia and depression. *J Consult Clin Psychol* 1987;55(2):236-238. [http://dx.doi.org/10.1037/0022-006X.55.2.236]
- Cherrier MM, Mendez MF, Dave M, Perryman KM. Performance on the Rey-Osterrieth complex figure test in Alzheimer disease and vascular dementia. *Neuropsychiatry Neuropsychol Behav Neurol* 1999;12(2):95.
- Roos A, Calata D, Jonkers L, et al. Normative data for the Tygerberg Cognitive Battery and Mini-Mental Status Examination in a South African population. *Compr Psychiatry* 2010;51(2):207-216. [http://dx.doi.org/10.1016/j.comppsy.2009.03.007]
- Monsch AU, Bondi MW, Butters N, Salmon DP, Katzman R, Thal LJ. Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Arch Neurol* 1992;49(12):1253-1258. [http://dx.doi.org/10.1001/archneur.1992.00530360051017]
- Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Int Med* 2004;256(3):183-194. [http://dx.doi.org/10.1111/j.1365-2796.2004.01388.x]
- Nordlund A, Rolstad S, Hellstrom P, et al. The Goteborg MCI study: Mild cognitive impairment is a heterogeneous condition. *J Neurol Neurosurg Psychiatry* 2005;76(11):1485-1490. [http://dx.doi.org/10.1136/jnnp.2004.050385]