

Evaluation of the Usability of a Brief Computerized Cognitive Screening Test in Older People for Epidemiological Studies

J. Fredrickson^a P. Maruff^{a,c} M. Woodward^b L. Moore^d A. Fredrickson^a
J. Sach^d D. Darby^{a–c}

^aCogState Ltd., ^bAged and Residential Care/Neurosciences Clinical Service Unit, Austin Health, and ^cCentre for Neuroscience, University of Melbourne, Melbourne, Vic., and ^dAlzheimer's Australia Vic, Hawthorn, Vic., Australia

Key Words

Computerized testing · Elderly population, community-based · Cognitive screening · Usability

Abstract

Background: Computerized cognitive testing has the potential to be an effective way to assess and monitor cognition in large neuroepidemiological studies. CogState is a game-like computerized test with demonstrated validity and reliability that has shown sensitivity to decline in older individuals over time. This study aimed to evaluate the serial usability of the test specifically within an older community cohort. **Methods:** The test battery was administered to healthy volunteers aged 50 years and above at 3-month intervals over 12 months in a community setting. Test usability was examined in terms of acceptability, efficiency and stability. **Results:** Of 301 subjects (age: 61.9 ± 7.2 years), 87% completed the study. In addition, 85% completed the first test within the allowed time and passed integrity criteria with their performance improving and stabilizing at subsequent visits. The computerized battery required 15 min for administration on average, allowing 263 patients to be assessed on 5 occasions by 2 assessors. All tasks showed stability and a high test-retest reliability with serial administration. **Conclusions:** This computerized test was shown to have good ac-

ceptability, efficiency and stability for the repeated assessment of cognitive function in older people. Together with its demonstrated sensitivity to cognitive impairment and cognitive change, these data suggest that it would be a useful tool for application in neuroepidemiological studies.

Copyright © 2009 S. Karger AG, Basel

Introduction

The detection of cognitive impairment is a central aim of neurological epidemiology, particularly in the area of dementia [1, 2]. In neuroepidemiological studies, cognitive function has been assessed in large samples measured on a single occasion and in 'at risk' groups followed prospectively [3–6]. In general, cognitive assessments require that each subject be assessed individually, by a specifically trained assessor in a relatively lengthy assessment period. These factors can conspire not only to limit the potential size of neuroepidemiological samples or to increase the cost of studies, but also to limit the potential for repeated assessment, especially over the relatively

J.F., P.M., A.F. and D.D. are full-time employees of CogState Ltd. The company supplies the test battery used in this study.

short re-retest intervals that are optimal for monitoring change in the cognition of individuals over time [7]. While some large studies have incurred the resource cost of using formal assessments of cognition [2, 8, 9], others have defined cognitive function using only single instruments (e.g. the Mini-Mental State Examination) [10] or small sets of brief cognitive tests (e.g. the Mini-Cog) [11], or by using questionnaires to have subjects rate their cognitive function subjectively [12].

Another limitation related to the use of standardized cognitive assessment in prospective studies is that often the tests themselves have not been designed for repeated application, especially at short re-retest intervals [13]. For example, the stimuli used in some memory tests can be learned and then remembered between administrations, diminishing the sensitivity of subsequent assessments to true change. Alternatively, some tasks define cognitive functions in terms of the abilities of an individual to develop strategies for successful performance. However, once developed, these strategies are not forgotten and can be used immediately when the same task is presented on a second occasion [13]. Thus, the dimensions of cognitive function measured by standardized tests can change with repeated exposure to these tests and this reduces the comparability of results between successive testing occasions. This problem increases as retest intervals decrease and can obscure sensitivity to the cognitive deterioration that characterizes early neurodegenerative disease [7].

Computerized cognitive testing offers one possible solution for the problem of assessing and monitoring cognition in large samples [14]. We have developed a computerized system for the assessment of cognitive function that can be used for the single [15] or serial [7, 16, 17] evaluation of cognitive function in large samples [18–21]. The cognitive tasks themselves in this system were derived from cognitive neuroscientific and experimental psychological paradigms of psychomotor function, attention, learning and working memory [22]. The tasks developed are simple and can be repeated without giving rise to practice effects [13, 23]. Performance measures from the tasks have been shown to be sensitive to deterioration and improvement in cognitive function in both healthy and disordered populations [24, 25], and to decline in otherwise healthy older people prior to the onset of dementia [26] and in older people who have undergone cardiac surgery [27]. The validity, reliability, stability and sensitivity of the CogState tasks in experimental samples have been detailed previously [13, 15, 28–30].

One area in which the CogState test battery has been used extensively is in the monitoring of cognitive chang-

es associated with concussion in elite contact sports [18, 20]. Sensitivity to concussion-related central nervous system (CNS) change requires that the cognitive tasks be able to detect deterioration in performance from a preinjury baseline, and then also to detect improvement in performance with the resolution of an injury [18, 31, 32]. Consequently, concussion management systems that utilize cognitive tests require that all of the individuals at risk are assessed in the test prior to being exposed to the playing environment. The benefit to concussion management from the increased sensitivity of computerized cognitive tests has been discussed previously [20, 33]. One factor important to the successful implementation of concussion management systems has been their high level of efficiency and acceptability as well as their stability and sensitivity to cognitive change associated with mild head injury [18, 28, 34]. It is these same factors that we believe make computerized cognitive assessments potentially useful for neuroepidemiological studies, particularly those prospective studies which aim to understand the development of CNS disease in older people through monitoring change in cognitive function via repeated assessments.

There are methodological challenges associated with the use of computers to measure cognitive function in older people that must be overcome before any potential benefits can be realized. For example, older people generally have less experience with computers than younger adults and can therefore require more time and assistance before they feel comfortable completing computerized tasks [35]. Consequently, errors in performance can reflect a lack of computer experience instead of, or as well as, true cognitive difficulties. A good interface design [36] and the use of simple cognitive tests does improve the usability of computer-based assessments in older people and, in turn, this increases the reliability and validity of cognitive performance data [37]. These same principles evolved in response to the use of the CogState battery in concussion management systems, and therefore it is likely that they will also apply to the study of cognitive function in older people.

The primary aim of the current study was to evaluate the usability of the CogState battery for the assessment of cognitive function in large groups of older people. This was evaluated by studying the acceptability, efficiency and stability of performance of the tests in the CogState battery in a large community sample of older adults. The acceptability of the CogState battery was determined by measuring the extent to which older people could understand and complete the tests across the 5 visits. The effi-

ciency of the computerized battery was determined by the length of time required for the assessment of individuals and the resource use associated with the repeated measurement of cognitive function in a large group of older adults assessed 5 times within 1 year. The stability of the battery was determined by comparing performance between visits as well as estimating the within-individual variation for each test across the study period. The primary hypothesis was that the usability of the CogState battery would be high in healthy older people. Thus, we proposed that the tests would have a high acceptability in older people, that the battery could be applied with a high efficiency, and that the performance data generated would be stable across the study period.

Materials and Methods

Study Design

This is a prospective community-based longitudinal observational study. Testing was conducted at 2 sites within the suburbs of Melbourne. The first site was in a community setting at the state headquarters of a disease advocacy group (Alzheimer's Australia), whilst the second was in a clinical research unit in the memory clinic of a large metropolitan hospital. The subjects selected the site of their assessment by personal choice, then repeated assessments remained at that site. A single assessor was responsible for all testing at both sites using staggered appointments. At the baseline session, an interview obtained information about subject demographics, medical history and expectancies (Appendix). Computerized testing was then performed, followed by scheduling the next visit. The test was presented only once on each visit. A questionnaire concerning the experiences of each subject with the testing was provided to each subject with a pre-paid return envelope. There were 5 testing visits, occurring at 3-month intervals over a 12-month period. At the baseline visit, each subject's testing was supervised individually.

Following each visit, a report was sent to each subject and to their medical practitioner. The report was generated automatically by the computer system and compared the performance of the subject to normative data. A simple authored interpretation was added by a study physician (D.D.). If subjects expressed concern about their results, they were offered initial telephone support and discussion by the study assessor, and if their concern remained, they were offered counselling by a study investigator or the disease advocacy group, or a medical evaluation at the memory clinic of the metropolitan medical centre.

At the completion of the 12-month visit, each subject completed a subject expectation questionnaire (Appendix), and a final visit was scheduled to discuss their results. The final visit was conducted by a study physician (D.D.), preferably in person or by telephone. This involved giving feedback on the study overall as well as on their own performance in serial cognitive testing. Subjects with ongoing concerns, or in whom impairment or serial decline was detected, were counselled as well as offered further

counselling by the disease advocacy group, medical evaluation by their medical practitioner or referral to the memory clinic. In addition, their medical practitioners were sent a copy of the final report and a questionnaire (Appendix) to assess their experiences with the study. Subjects who withdrew early from the study were contacted to elicit their reasons and were sent the subject expectation questionnaire to complete and return.

Subjects and Recruitment

Subjects were recruited via an advertising campaign that included: (1) an announcement on a radio talk show, in which one of the authors (L.M.) explained the purpose of the study; (2) articles in the advocacy group newsletter, and (3) a news release from the advocacy group, which was reproduced in a number of local newspapers. In all the media, a telephone number was given as the initial point of contact, with people who registered subsequently contacted by the researchers to assess their eligibility. Inclusion criteria were: (1) age of 50 years or over; (2) willingness to nominate a current medical practitioner to be kept informed of the study participation and results; (3) willingness to be informed of their results during the study and to accept the potential risk that participation in the study may demonstrate their cognitive performance to be impaired or declining; (4) willingness to attend all 6 study visits at one of the two testing sites, and (5) ability and willingness to provide informed consent. The exclusion criteria were based on self-report and included: (1) any known significant cognitive impairment due to neurological or medical disease; (2) any other condition that might make it difficult for them to complete 12 months of testing; (3) uncorrectably impaired visual acuity preventing the discrimination of visual changes on a computer screen; (4) physical handicap or condition preventing effective use of a computer keyboard or 2-button mouse, and (5) unwillingness to undergo testing using a computer. All subjects had sufficient conversational English language skills to complete the surveys. The study design was approved by the institutional ethics committee, and all subjects gave their written informed consent to participate.

Apparatus

The computerized cognitive battery was presented on a laptop or tablet computer, with 2 of them available at each testing site. The subjects were given 30 min to complete the battery, which consisted of 5 initial tasks displaying playing card stimuli and has been described before [7, 13]. Each task was presented in 2 parts, with an initial familiarization practice before the scored component. Simple textual instructions appeared prior to each task to remind the subject of the task requirements. The assessor was on hand to answer any additional questions during testing.

In each task, a playing card is presented face-down in the centre of the screen on a green background. After an interval that varied randomly between 2.5 and 3.5 s, the card instantly turned face-up, and the subjects were required to respond 'Yes' or 'No' based on simple questions which varied for each task. The 'Yes' response was indicated by a keyboard 'K' key, and 'No' by a 'D' key, or alternatively by the right and left mouse buttons, respectively. Visual feedback differentiated correct and incorrect responses, with an additional error sound after incorrect responses. Anticipatory responses occurring before the card turned face-up triggered a distinctive error sound. The speed and accuracy of each response was recorded. Each task is described below.

Detection Task

The detection (DET) task is a simple reaction time task that measures psychomotor function. In this task, the subject is required to press the 'Yes' key as quickly as possible when the central card turns face-up (constituting 1 trial). Correct responses following an anticipatory response are ignored. The face-up card displayed is always the same joker card. The task ends when 35 correct trials have been recorded. This task is repeated after the one back (OBK) task (see below).

Identification Task

The identification (IDN) task is a choice reaction time task that measures visual attention. This task is presented similarly to the DET task, with instructions indicating the subject should respond 'Yes' if the face-up card is red, or 'No' if it is not red. The cards displayed are red or black joker cards, and 30 correct (non-anticipatory) trials are required. Joker cards are used to ensure that playing cards presented in the next task were not previously seen in the same testing session.

One Card Learning

The one card learning (OCL) task is a continuous visual recognition learning task that assesses visual recognition memory and attention. This task is similar in presentation to the IDN task, with instructions indicating the subject should respond 'Yes' if the face-up card has appeared in the task before, and 'No' if it has not yet appeared. Normal playing cards are displayed (without joker cards). There are 6 cards that repeat and 42 trials without rescheduling for postanticipatory correct trials.

One Back Task

This is an OBK task that assesses working memory and attention. It is similar in presentation to the OCL task, with instructions indicating the subject should respond 'Yes' if the face-up card is exactly the same as the immediately previous card, or 'No' if it is not the same as the previous card. Normal playing cards are again used, and 30 correct trials required for completion.

For each task, the accuracy of performance was defined by the number of correct responses made (i.e. true positive and true negative), expressed as a proportion of the total trials attempted. An arcsine transformation was then applied to normalize the distribution. The speed of performance was defined in terms of the average reaction time (RT; milliseconds) for correct responses. A base 10 logarithmic transformation was then applied to normalize the distributions of mean RT. On the basis of prior studies [38–40], the performance in each task was defined by a single outcome measure. Speed of performance was used for the DET, IDN and OBK tasks, and accuracy of performance for the OCL task.

Subjective Rating Scales

Several questionnaires were administered at each testing session prior to the core battery. The Penn State Worry Questionnaire-Past Week [41, 42] assesses anxiety during the past week, using subjective ratings from 'Never' (0) to 'Almost always' (6) for 15 simple statements. The maximum score is 90. The Patient Health Questionnaire [43] requires a subjective rating of 9 short statements covering mood and somatic symptoms of depression over the past 2 weeks using 'Not at all' (0) to 'Nearly every day' (3). The maximum score is 27, suggestive of severe depression, with

mild depression if the score is 5 or higher. Subjective fatigue was also rated using a scale from 'Fully awake' (0) to 'Asleep' (9).

Following the computerized tests, the subjects rated their performance in the core battery. They rated their ability to make decisions, concentrate, respond quickly and learn new information on a scale from 'Very much better than usual' (0) to 'Very much worse than usual' (10).

Computer Experience

The computer experience of each subject was classified at the baseline visit as: (1) 'none' if they required general help with the basic hardware, for example being shown how to use the mouse, or (2) 'some' if the subject navigated the hardware with no help, or only needed help specific to that particular computer.

Data Analysis

The acceptability of the tasks was determined in 2 stages. First, acceptability was defined in terms of the retention of subjects in the study. Retention was evaluated using the proportion of subjects who: (1) failed to attend any of the 3-, 6- or 9-month visits; (2) withdrew from the study before the 12-month visit, and (3) completed the 12-month visit. The second aspect of acceptability was the extent to which subjects could understand and adhere to the requirements of each task at each visit, and was determined on the basis of task completion and performance integrity. For each task, task completion was defined as completion of 75% or more of the trials for that task. In addition, for each task, performance integrity was defined using a score on a measure different to the main outcome. Specifically, the 4 integrity criteria were that: (1) the accuracy of performance in the DET task was >1.107 (i.e. 80%); (2) the accuracy of performance in the IDN task was >1.107 (i.e. 80%); (3) the accuracy of performance in the OBK task was >0.99 (i.e. 70%), and (4) the speed of performance in the OCL task was $<3.24 \log_{10}(\text{RT})$ (i.e. 1,749 ms). If a subject's performance in a task failed both the task completion and integrity criteria, it was classified as a failure of task completion.

In order to control for self-exclusion due to subjectively assessed poor initial performance [44], calculations on task completion and integrity criteria compared the performance of the primary outcome variables in those who withdrew to those completing the 12-month visit by using χ^2 tests. A comparison of computer experience rating with test performance was also made to evaluate the effect of computer experience on test acceptability. For these analyses, non-parametric tests were used when the size of subgroups became small, or where there was heterogeneity of variance.

We also sought to determine how demographic and mood variables influenced performance in the CogState tasks as an additional measure of acceptability. The subjective ratings of depression and anxiety before each test were compared over time using non-parametric tests. Multiple regression was used to determine the relationship between mood and performance, and also to examine the relationship between age and performance for the individual tasks.

The efficiency of the CogState battery was defined according to the length of time required for the administration of each battery on each visit, the number of subjects that could be assessed

Table 1. Performance and attendance at each visit

	Baseline	At 3 months	At 6 months	At 9 months	At 12 months
Subjects	301	271	253	258	263 (87%)
Passed	255 (85%)	250 (92%)	239 (94%)	247 (96%)	245 (93%)
Failed to attend	0	5	14	8	0
Withdrawals	0	25	34	35	38
Failed to complete within 30 min	18 (6%)	6 (2%)	1 (0%)	0	0
Failed 1 or more integrity check(s)	28 (9%)	15 (6%)	13 (5%)	11 (4%)	18 (7%)
Battery duration, min	16.3 ± 2.3	16.02 ± 2.19	14.81 ± 1.9	14.47 ± 2.17	14.1 ± 2.4

Table 2. Comparison of subjects who withdrew to those who completed the 12-month visit – performance at baseline

	Withdrawn	Completed 12-month visit	Significance
Subjects	38	263	
Failed to complete within 30 min	8 (21%)	10 (4%)	$\chi^2 = 17.57$, d.f. = 1, $p < 0.0001$
Failed 1 or more integrity check(s)	5 (13%)	16 (6%)	$\chi^2 = 2.11$, d.f. = 1, $p < 0.15$
No computer experience	15 (40%)	34 (13%)	$\chi^2 = 17.17$, d.f. = 1, $p < 0.0001$
Anxiety, PSWQ score	31.6 ± 13.6	24.9 ± 15.8	Kruskal-Wallis $\chi^2 = 38.3$, d.f. = 1, $p = 0.0001$

PSWQ = Penn State Worry Questionnaire-Past Week.

on 1 day, and the number of assessors required to supervise the computer systems in the study.

The stability of the main outcome measures for each test was determined by first excluding data that failed the task completion or integrity criteria. For each of the 5 outcome measures, the group mean performance was compared between adjacent assessments by planned comparisons within a one-way ANOVA. Measures of effect size [45] were used to express the magnitude of any change in performance over time. The within-subject SD was also computed for each measure by computing the square root of the mean square error term from each of these ANOVA. The test-retest reliability of each measure across the 5 visits was then established using intraclass correlations.

Results

In total, of 394 people who registered interest, 338 were screened by the researchers and scheduled for a baseline visit. Of these, 37 withdrew from the study before testing commenced. The remaining 301 subjects (212 females, 89 males) were aged between 50 and 86 years (mean age: 61.9 ± 7.2 years) and commenced the baseline visit of the study. A summary of the data pertaining to acceptability is presented in table 1, which shows the classification of subject performance and attendance at each visit. Of the

301 subjects who commenced a baseline session, 263 subjects (87%) completed the 12-month study visit. The majority of subjects (85%) completed the test at the baseline visit within the allotted time and passed all of the integrity criteria. On subsequent visits, between 92 and 96% of the subjects satisfied all the integrity criteria. Overall, 13% (38) of the subjects who enrolled in the study failed to attend the 12-month visit. The majority of subjects who withdrew left the study after the baseline or the 3-month visit. Thereafter, retention was high. Of those who completed the 12-month visit, 236 (90%) attended all 5 visits, with 26 subjects attending 4 visits, and 1 attending 3 visits. Of the 38 subjects who withdrew from the study, 8 (21%) failed to complete the test within the allowed time at baseline, and a further 5 (13%) failed 1 or more integrity criteria at baseline. Their reasons for withdrawal included: personal illness in 5 subjects (13%); family illness or death in 6 (16%); inability to travel to testing sites or scheduling difficulties in 10 (26%); loss of interest in 6 (16%); 8 (21%) could not be contacted, and 3 subjects (8%) cited a perceived difficulty with testing or the computer. Table 2 compares the characteristics of the subjects who withdrew and of those that completed the 12-month visit. It can be seen that failure on 1 or more integrity crite-

Table 3. Proportion of failed integrity criteria by task

	Baseline	At 3 months	At 6 months	At 9 months	At 12 months
Subjects	301	271	253	258	263
DET	12 (4%)	6 (2%)	2 (1%)	3 (1%)	3 (1%)
IDN	4 (1%)	1 (0%)	1 (0%)	4 (2%)	4 (2%)
OBK	21 (7%)	6 (2%)	6 (2%)	4 (2%)	1 (0%)
OCL	17 (6%)	4 (1%)	5 (2%)	2 (1%)	9 (3%)

ria, or failure of task completion at baseline, was 3 times more common in those withdrawing than in those who were retained (37% compared to 12%). Compared to the subjects retained in the study, a classification of no computer experience was also 3 times more common in those withdrawing from the study, and the level of anxiety symptoms was also significantly higher in this group.

There were 49 subjects (16%) who were classified as having no computer experience. Of these, 15 (30%) failed to complete the test within the allowed time at the baseline visit compared to 3 (1%) of those with some computer experience ($\chi^2 = 63.2$; d.f. = 1; $p < 0.0001$). However, there was no significant difference by level of computer experience for failing integrity criteria, with 2 subjects with no computer experience (4%) and 26 (10%) with some computer experience failing. From the 6-month visit to the 12-month visit, all of the remaining 34 subjects with no computer experience completed the test battery within the allotted time. In addition, the number of participants with no computer experience who failed on 1 or more integrity criteria was between 2 (6%) and 4 (12%) from the 6-month visit onwards. A comparison of 29 subjects with no computer experience with an age- and gender-matched subgroup of 29 subjects with some computer experience regarding DET, IDN and OBK speed and OCL accuracy revealed no significant differences in speed or accuracy at either the baseline visit (χ^2 ; d.f. = 1; DET: $p = 0.8$; IDN: $p = 0.3$; OBK: $p = 0.4$; OCL: $p = 0.6$) or the 12-month visit (χ^2 ; d.f. = 1; DET: $p = 0.1$; IDN: $p = 0.2$; OBK: $p = 0.3$; OCL: $p = 0.9$) in any of the tasks. The group classified as having no computer experience also included significantly more subjects that were 75 years or older (24%), compared to the group with some computer experience (4%; $\chi^2 = 18.7$; d.f. = 1; $p < 0.0001$).

In addition to analysing performance by subject, acceptability was also analysed according to task (table 3). Less than 8% of subjects failed the integrity criteria associated with any of the tasks on any visit. However, where

a failure did occur, it was most often at the first visit and for the OBK and OCL tasks. The influence of mood and demographic variables was examined, and table 4 shows that the subjective ratings of anxiety and depression decreased across the study period. In order to assess the relationship between subjective rating scales and task performance at the 12-month visit, a multiple regression analysis was performed. This analysis showed no significant relationship for the DET ($r = 0.03$; $p = 0.14$) or IDN ($r = 0.10$; $p = 0.18$) tasks; however, there were significant but small magnitude relationships found for the OCL and OBK tasks, with higher anxiety and depression scores associated with lower accuracy in the OCL ($r = 0.17$; $p < 0.05$) and slower speed in the OBK ($r = 0.14$; $p < 0.05$) task. A regression for age against task performance at the 12-month visit showed a small but significant relationship for all tasks with increasing age associated with slower speed (DET: $r = 0.21$, $p < 0.001$; IDN: $r = 0.24$, $p < 0.001$; OBK: $r = 0.25$, $p < 0.001$) and for OCL with reduced accuracy (OCL: $r = 0.24$, $p < 0.001$).

The data for the efficiency of the computerized tasks are summarized in the time for test completion in table 1. Overall, the average time required for test administration was 15.14 ± 2.6 min. At the baseline and 3-month visits, the test duration was around 16 min and was between 15 and 14 min at subsequent visits. At the baseline visit, the subjects were scheduled half an hour apart as the full attention of the supervisor was required during at least half of the visit. At subsequent visits, the subjects were scheduled 15 min apart as computerized testing was mostly self-administered and 2 subjects were able to be assessed simultaneously in adjacent rooms by a single assessor using staggered starting times. Thus, a maximum of 14 people were tested each day during baseline testing (mean: 10.6 ± 3.6 people), and up to 20 subjects were tested per day at subsequent visits (mean: 13.6 ± 3.6 subjects).

The stability of the main performance measure for each task at each visit with the exclusion of individual results failing integrity criteria is displayed in table 5. For each measure, the coefficients of variation were low and the intraclass correlations high. There was a small improvement in group performance from the baseline to the 12-month visits for the accuracy of performance in the OCL task, and a decrease in the speed of performance for the OBK task. The effect sizes show that the magnitudes of these changes were small. An analysis to determine at which visit improvement occurred showed that practice effects were not evident between the first and second visit for any of the tasks. However, for OCL and OBK an improvement in performance of moderate magnitude oc-

Table 4. Self-rated anxiety and depression of subjects at each visit

	Baseline	At 3 months	At 6 months	At 9 months	At 12 months	Kruskal-Wallis
Anxiety	25.7 (15.6%)	23.6 (17.2%)	23.4 (18.5%)	21.3 (17.6%)	21.5 (17.5%)	$\chi^2 = 109.8$, d.f. = 1, $p < 0.05$
Depression	2.8 (3.4%)	2.5 (3.2%)	2.7 (3.4%)	2.4 (3.2%)	2.5 (3.1%)	$\chi^2 = 29.0$, d.f. = 4, $p < 0.05$

Table 5. Means \pm SD by task over time for all subjects

	DET ¹	IDN ¹	OCL ²	OBK ¹
Baseline	2.52 \pm 0.11 (289)	2.72 \pm 0.07 (297)	1.01 \pm 0.11 (284)	2.94 \pm 0.09 (280)
3-month visit	2.54 \pm 0.12 (264)	2.74 \pm 0.07 (270)	1.04 \pm 0.11 (267)	2.93 \pm 0.09 (264)
6-month visit	2.50 \pm 0.11 (251)	2.71 \pm 0.07 (252)	1.03 \pm 0.11 (248)	2.92 \pm 0.09 (247)
9-month visit	2.49 \pm 0.10 (255)	2.71 \pm 0.06 (254)	1.04 \pm 0.10 (256)	2.91 \pm 0.09 (254)
12-month visit	2.52 \pm 0.11 (260)	2.72 \pm 0.07 (259)	1.04 \pm 0.10 (254)	2.91 \pm 0.09 (262)
Overall	2.52	2.72	1.03	2.92
Effect size ³	-0.03	0.07	0.27	0.4
WSD	0.08	0.05	0.09	0.04
CoV	0.03	0.02	0.09	0.01
ICC	0.79	0.86	0.65	0.91

Numbers of subjects for each task and visit are in brackets. WSD = Within-subjects standard deviation; CoV = coefficient of variation; ICC = intraclass correlation coefficient.

¹ Speed measure. ² Accuracy measure. ³ Effect size (calculated using Dunlap's *d*) is based on the difference between baseline and 12-month visits, a negative effect size indicates poorer performance at the 12-month visit.

curred between the baseline and 9-month visits (0.26 and 0.38, respectively), and these advantages remained present at the 12-month visit.

Discussion

The purpose of this study was to evaluate the usability of the computerized test battery in older community-dwelling individuals for a prospective screening of cognition. The tasks had good acceptability in this older group of people. Initial interest in the study was high when the study was advertised, and there was a very low rate of withdrawal across the study period. People who did withdraw from the study did so mainly for reasons unrelated to the use of the computerized test. Hence, the high level of interest was maintained across the study period despite people having to visit the assessment centres 5 times within 12 months. Furthermore, the majority of the sample (85%) were able to understand and adhere to all of the task requirements on their first visit. Of the individuals who had some initial difficulty with the computerized

test, the majority then provided test data that met all task completion and performance integrity criteria. From the 3-month to the 12-month visits, between 92 and 96% of the study group were able to complete the test in the allowed time without failing integrity checks. The efficiency of the test was high, since all of the testing could be conducted with a minimal supervision of individuals and the relatively short administration time allowed up to 20 assessments to be conducted on each study day. Thus, these data suggest that the computerized battery has had a high acceptability and efficiency when used to assess cognition in older people.

Each performance measure from the computerized battery had a high stability across the 12-month study period, indicated by the low coefficients of variation and the strong test-retest reliability correlations over the course of the study year. For the measures of working memory and visual recognition memory, there was a small improvement in performance across the study period. Compared to the baseline visit, this improvement became evident only the fourth time the battery was performed (i.e. at the 9-month visit). Despite this subtle im-

provement, the estimates of stability and test-retest reliability were high for each outcome measure. These data suggest that the computerized test battery provides reliable estimates of cognition over relatively long periods of time, and that the variability in the performance within older individuals on each of the cognitive measures is very small. Taken together, the efficiency and acceptability of the computerized test battery as well as the strong test-retest reliability and high stability of the performance measures indicate that the CogState test battery has a good usability for studies of cognitive function in older people, particularly for those studies using a neuroepidemiological methodology.

The impact of experience with computers is particularly relevant when testing older people as it is a potential moderator of age-related performance differences in computer-based tasks [46]. There was a range of ability levels represented in the current group, and consistent with previous studies, older age was associated with no computer experience [47]. Although the absence of experience with computers was initially associated with slower performance, this association did not persist throughout the study visits. Of those who failed to complete the test in the allotted time at the initial visit, but who continued in the study, all completed the tests successfully at the 12-month visit. This indicates that the initial slowness of performance observed in some older people may reflect their unfamiliarity with computers rather than impaired cognitive function. This unfamiliarity may be amenable to simpler instruction or more careful supervision. Importantly, by the 6-month visit the extent of an individual's experience with computers was no longer related to any of the cognitive performance measures. These data suggest that, with some supervision or practice, the computerized battery can be completed successfully by older people with very limited computer experience at a level that is similar to that of people who have some computer experience.

The majority of people who withdrew from the study did so in the first 3 months, and the rates of withdrawal did not increase over the study period. This suggests that withdrawal was not related to inconvenience or to time demands but rather to the initial experience of assessment. This is supported by the generally poorer performance, higher anxiety levels and more failures of task completion and performance integrity at baseline in the individuals who later withdrew from the study. This is consistent with reports of selective withdrawal in which an initial poor performance may have discouraged continued participation [44]. This is a concern for epidemio-

logical studies as it may result in a sample that does not reflect the population at large. Indeed, research has found that subjects who withdraw early are more likely to be dissatisfied with the study, disappointed with their performance or even cognitively impaired [48, 49]. Thus, even when computer-administered tests can be applied to large samples with low supervision rates, there is still a need to identify individuals experiencing difficulty and offer them greater support or assistance in the initial assessment sessions. This strategy has been successful in a context where computerized testing was alien to people for cultural reasons, where the acceptance of the test was boosted when time was allowed for familiarization [25]. The benefits of familiarization could also be observed in the current study among older people with no computer experience, in whom their initially poor performance improved with further exposure to the tests. Hence, a lower initial rate of withdrawal should be possible if subjects who are having difficulty are allowed more time to practice, understand and familiarize themselves with the tasks, aiming to reduce test anxiety. A brief practice test or more practice items are possible test improvements, which would not affect the detection of decline as practice does not alter performance in this test [13]. If this is successful in enabling people of all ability levels to tolerate the test, it may result in retaining a more representative epidemiological sample.

The presence of minimal practice effects in the tasks of this battery is consistent with the design of this test as an instrument for repeated use and with its previous use in serial testing [13, 23]. We have argued that the sensitivity of a test to true change in cognitive function in individuals is highly dependent on the extent to which performance remains stable in the absence of true CNS change [38, 39]. Given that the performance measures studied here were stable and that these same measures have been shown previously to be sensitive to subtle change in cognition over time [39, 50, 51], this strongly suggests that the CogState battery has a good potential for use as a cognitive screening tool for older people. On the basis of the criteria for completion and task understanding, older persons appeared to be as adept with the tasks as younger populations, with similar proportions failing such checks [18]. It is possible that the features of the test that were designed for resistance to practice effects in younger populations, such as the simple familiar stimuli of the playing cards and the constancy of visuomotor paradigms, aid understanding in older people. Furthermore, some of these features including the simple response mapping and the absence of complicated instruc-

tions have been shown previously to enhance older people's performance in computerized tasks [36]. Consistent with previous research, we also observed on a single visit significant relationships between age and performance in the different cognitive measures, where performance generally declined as age increased [46, 52].

There are several limitations on generalizing the findings of this study to the wider population. The low patient-to-assessor ratio in this study was necessary to determine the usability of the battery in neuroepidemiological contexts because the sample studied here was smaller than that typically recruited in epidemiological studies. Also, the subjects in this study were clearly highly motivated, suggesting that they were likely to tolerate more testing difficulties than less enthusiastic cohorts. The low rate of withdrawal and apparent satisfaction are likely to be lower in the general community. A voluntary participation in studies investigating memory decline is also less likely to attract persons without subjective cognitive complaints but who are nevertheless still at risk of dementia [53]. Furthermore, the proportion of subjects with no computer experience (16%) is lower than reported in similar general community studies, with up to 50% of those aged over 55 years reported to have never used a computer before [54]. In addition, given previous research showing a slow uptake of computer-based programs in older sections of the community [55], a screening program using this test may not cover the majority of those older persons at risk of cognitive decline and may experience a higher rate of withdrawal from the program. Nevertheless, the proportion of older persons who have computer experience is likely to increase over time as the currently younger cohorts with computer experience age. In addition, future epidemiological studies in which the screening of cognition will be important will also be likely to include motivated volunteers for whom access to, and confidence with, using computerized testing can be a condition of participation.

This study suggests that computerized testing has a good usability for older people, with performance levels consistent with those observed in younger groups. Where this has been observed in previous studies of other computerized tests on a single testing occasion [14, 56], this study demonstrated the acceptability, efficiency and stability of this computerized test battery over repeated testing visits, indicating its potential for use in longitudinal studies of cognitive decline. There is no doubt that the current frequency of testing and numbers involved could not have been delivered as efficiently by using conventional neuropsychological techniques. There was also no

requirement for the assessor to be an expert neuropsychologist. Provided that people that display initial difficulty with the test can be identified and guided through the process, the ease of application demonstrated in this study supports a further evaluation of computerized testing as a viable means of frequent cognitive screening in large epidemiological studies including older community-dwelling persons.

Acknowledgement

The study was conducted under an unrestricted educational grant from Pfizer Australia under the auspices of Alzheimer's Australia Vic.

Appendix

Demographic information

- a. Year of birth
 - b. Gender
 - c. Handedness
 - d. Preferred language
 - e. Highest educational achievement
 - f. Vocational history
 - g. Current medication (OTC, herbal or prescription)
 - h. Medical and psychiatric history of significant conditions
-

Subject expectation questionnaire

1. What interested you most about this study?
 2. Are you concerned about your memory?
 3. Are you worried about any particular causes of memory difficulty?
 4. How often do you worry about your memory?
 5. Are there other changes in your thinking, mood or behaviour which you have noticed in the past year which are concerning you?
 6. Do you have a family member or close friend who has or has had Alzheimer's disease or other forms of dementia?
 7. What do you understand about the testing procedures?
 8. What do you think the tests will show?
 9. What are you likely to feel if your tests find something is wrong?
 10. What would you want to happen if your tests are found to have something wrong?
 11. What do you think your doctor's attitude is towards screening for memory problems?
 12. Do you have any other expectations of this study?
-

Subject experience questionnaire

1. Did the study meet your expectations? Please explain your reasons.
2. What bad experiences did you have during the study?
3. What good experiences did you have during the study?
4. Did you feel that there was sufficient information provided during the study?

5. What additional information would have helped?
6. To what extent did you consider it worthwhile to participate in the study?
7. How acceptable was the 3-monthly testing to you?
8. What interval would be most acceptable to you?
9. To what extent was the study of value to you?
10. Did the study satisfy your original motivation for agreeing to participate?
11. Do you feel more or less anxious having taken part in the study?
12. What did you think of the regular reports?
13. What did you think of the frequency of testing sessions?
14. What do you think your final results will show?
15. What feedback did you get from your general practitioner about the study, if any?
16. If testing were available through community centres or medical practitioners, would you be likely to continue with 1- to 2-yearly screening tests?
17. If you were found to be declining, what would you want to happen?
18. If you were found to be declining, what therapies would you consider trying?

19. If you were found to be declining, would you want to participate in research trials of possible therapies?
20. Would you recommend screening to family, friends or associates?
21. What would you suggest might improve the screening procedures?
22. What else could have been done better?
23. Do you have any other comments about the study?

Post-final-visit doctor questionnaire

1. What expectations did you have when you first heard about the study?
 2. Were your expectations confirmed?
 3. Did you feel the study was beneficial in any way?
 4. Did you feel the study was detrimental in any way?
 5. Do you have any other concerns about such screening?
 6. Would you refer patients to community centres for regular screening visits?
 7. Do you have any suggestions on how the screening process could be improved?
 8. Do you have any other comments on the study?
-

References

- 1 Brookmeyer R, Gray S, Kawas C: Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am J Public Health* 1998;88:1337-1342.
- 2 Fish M, Bayer AJ, Gallacher JE, Bell T, Pickering J, Pedro S, Dunstan FD, Ben-Shlomo Y, Ebrahim S: Prevalence and pattern of cognitive impairment in a community cohort of men in South Wales: methodology and findings from the Caerphilly Prospective Study. *Neuroepidemiology* 2008;30:25-33.
- 3 Christensen H, Dear KB, Anstey KJ, Parslow RA, Sachdev P, Jorm AF: Within-occasion intraindividual variability and preclinical diagnostic status: is intraindividual variability an indicator of mild cognitive impairment? *Neuropsychology* 2005;19:309-317.
- 4 Dufouil C, Fuhrer R, Alperovitch A: Subjective cognitive complaints and cognitive decline: consequence or predictor? *The Epidemiology of Vascular Aging Study. J Am Geriatr Soc* 2005;53:616-621.
- 5 Finkel D, Reynolds CA, McArdle JJ, Pedersen NL: The longitudinal relationship between processing speed and cognitive ability: genetic and environmental influences. *Behav Genet* 2005;35:535-549.
- 6 Jorm AF, Masaki KH, Petrovitch H, Ross GW, White LR: Cognitive deficits 3 to 6 years before dementia onset in a population sample: the Honolulu-Asia Aging Study. *J Am Geriatr Soc* 2005;53:452-455.
- 7 Darby D, Maruff P, Collie A, McStephen M: Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology* 2002;59:1042-1046.
- 8 Farmer ME, White LR, Kittner SJ, Kaplan E, Moes E, McNamara P, Wolz MM, Wolf PA, Feinleib M: Neuropsychological test performance in Framingham: a descriptive study. *Psychol Rep* 1987;60:1023-1040.
- 9 McDowell I, Hill G, Lindsay J: An overview of the Canadian Study of Health and Aging. *Int Psychogeriatr* 2001;13(suppl 1):7-18.
- 10 Gascón-Bayarri J, Reñé R, del Barrio JL, de Pedro-Cuesta J, Ramón JM, Manubens JM, Sánchez C, Hernández M, Estela J, Juncadella M, Rubio FR: Prevalence of dementia subtypes in El Prat de Llobregat, Catalonia, Spain: the PRATICON Study. *Neuroepidemiology* 2007;28:224-234.
- 11 Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A: The Mini-Cog: A cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry* 2000;15:1021-1027.
- 12 Reid LM, MacLullich AM: Subjective memory complaints and cognitive impairment in older people. *Dement Geriatr Cogn Disord* 2006;22:471-485.
- 13 Falsetti MG, Maruff P, Collie A, Darby DG: Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-min, one-week and one-month test-retest intervals. *J Clin Exp Neuropsychol* 2006;28:1095-1112.
- 14 Tornatore JB, Hill E, Laboff JA, McGann ME: Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery. *J Neuropsychiatry Clin Neurosci* 2005;17:98-105.
- 15 Cysique LA, Maruff P, Darby D, Brew BJ: The assessment of cognitive function in advanced HIV-1 infection and AIDS dementia complex using a new computerised cognitive test battery. *Arch Clin Neuropsychol* 2006;21:185-194.
- 16 Falsetti MG, Maruff P, Collie A, Darby DG, McStephen M: Qualitative similarities in cognitive impairment associated with 24 h of sustained wakefulness and a blood alcohol concentration of 0.05%. *J Sleep Res* 2003;12:265-274.
- 17 Maruff P, Collie A, Darby D, Weaver-Cargin J, Masters C, Currie J: Subtle memory decline over 12 months in mild cognitive impairment. *Dement Geriatr Cogn Disord* 2004;18:342-348.
- 18 Collie A, Makdissi M, Maruff P, Bennell K, McCrory P: Cognition in the days following concussion: comparison of symptomatic versus asymptomatic athletes. *J Neurol Neurosurg Psychiatry* 2006;77:241-245.
- 19 Cysique LA, Maruff P, Brew BJ: Prevalence and pattern of neuropsychological impairment in human immunodeficiency virus-infected/acquired immunodeficiency syndrome (HIV/AIDS) patients across pre- and post-highly active antiretroviral therapy eras: a combined study of two cohorts. *J Neurovirol* 2004;10:350-357.
- 20 Makdissi M, Collie A, Maruff P, Darby DG, Bush A, McCrory P, Bennell K: Computerised cognitive assessment of concussed Australian Rules footballers. *Br J Sports Med* 2001;35:354-360.

- 21 Straume-Naesheim TM, Andersen TE, Bahr R: Reproducibility of computer-based neuropsychological testing among Norwegian elite football players. *Br J Sports Med* 2005; 39(suppl 1):64–69.
- 22 Maruff P: Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Arch Clin Neuropsychol* 2009;24:165–178.
- 23 Collie A, Maruff P, Darby DG, McStephen M: The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc* 2003;9:419–428.
- 24 Collie A, Maruff P, Darby D, Masters C, Currie J: The Melbourne Aging Study; in Tuokko HA, Hultsch DF (eds): *Mild Cognitive Impairment*. New York, Taylor and Francis, 2006, pp 93–113.
- 25 Cairney S, Clough A, Jaragba M, Maruff P: Cognitive impairment in Aboriginal people with heavy episodic patterns of alcohol use. *Addiction* 2007;102:909–915.
- 26 Meguro K: Clinical features of mild cognitive impairment and dementia in a community: an update of the Osaki-Tajiri Project. *Tohoku J Exp Med* 2008;215:125–131.
- 27 Lewis MS, Maruff P, Silbert BS, Evered LA, Scott DA: Detection of postoperative cognitive decline after coronary artery bypass graft surgery is affected by the number of neuropsychological tests in the assessment battery. *Ann Thorac Surg* 2006;81:2097–2104.
- 28 Collie A, Maruff P, Makdissi M, McCrory P, McStephen M, Darby D: CogSport: reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clin J Sport Med* 2003;13:28–32.
- 29 Maruff P, Falletti MG, Collie A, Darby D, McStephen M: Fatigue-related impairment in the speed, accuracy and variability of psychomotor performance: comparison with blood alcohol levels. *J Sleep Res* 2005;14:21–27.
- 30 Silbert BS, Maruff P, Evered LA, Scott DA, Kalpokas M, Martin KJ, Lewis MS, Myles PS: Detection of cognitive decline after coronary surgery: a comparison of computerized and conventional tests. *Br J Anaesth* 2004;92: 814–820.
- 31 Hinton-Bayre AD, Geffen GM, Geffen LB, McFarland KA, Frijs P: Concussion in contact sports: reliable change indices of impairment and recovery. *J Clin Exp Neuropsychol* 1999;21:70–86.
- 32 Maddocks D, Saling M: Neuropsychological deficits following concussion. *Brain Injury* 1996;10:99–104.
- 33 Collie A, Darby D, Maruff P: Computerised cognitive assessment of athletes with sports-related head injury. *Br J Sports Med* 2001;35: 297–302.
- 34 Straume-Naesheim TM, Andersen TE, Dvorak J, Bahr R: Effects of heading exposure and previous concussions on neuropsychological performance among Norwegian elite footballers. *Br J Sports Med* 2005;39 (suppl 1):70–77.
- 35 Morrell RW, Park DC, Mayhorn CB, Kelley CL: Effects of age and instructions on teaching older adults to use ELDERCOMM, an electronic bulletin board system. *Educ Gerontol* 2000;26:221–235.
- 36 van Gerven PW, Paas F, Tabbers HK: Cognitive aging and computer-based instructional design: where do we go from here? *Educ Psychol Rev* 2006;18:141–157.
- 37 Acevedo A, Loewenstein DA: Nonpharmacological cognitive interventions in aging and dementia. *J Geriatr Psychiatry Neurol* 2007;20:239–249.
- 38 Collie A, Darekar A, Weissgerber G, Toh MK, Snyder PJ, Maruff P, Huggins JP: Cognitive testing in early-phase clinical trials: development of a rapid computerized test battery and application in a simulated phase I study. *Contemp Clin Trials* 2007;28:391–400.
- 39 Maruff P, Werth J, Giordani B, Caveney AF, Feltner D, Snyder PJ: A statistical approach for classifying change in cognitive function in individuals following pharmacologic challenge: an example with alprazolam. *Psychopharmacology (Berl)* 2006;186:7–17.
- 40 Weaver Cargin J, Collie A, Masters C, Maruff P: The nature of cognitive complaints in healthy older adults with and without objective memory decline. *J Clin Exp Neuropsychol* 2008;30:245–257.
- 41 Meyer TJ, Miller ML, Metzger RL, Borkovec TD: Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther* 1990;28:487–495.
- 42 Stöber J, Bittencourt J: Weekly assessment of worry: an adaptation of the Penn State Worry Questionnaire for monitoring changes during treatment. *Behav Res Ther* 1998;36: 645–656.
- 43 Spitzer RL, Kroenke K, Williams JB: Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire*. *JAMA* 1999;282:1737–1744.
- 44 Hultsch DF, Hertzog C, Small BJ, McDonald-Miszczak L, Dixon RA: Short-term longitudinal change in cognitive performance in later life. *Psychol Aging* 1992;7:571–584.
- 45 Dunlap WP, Cortina JM, Vaslow JB, Burke MJ: Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol Methods* 1996;1:170–177.
- 46 Salthouse TA: The role of experience in cognitive aging. *Annu Rev Gerontol Geriatr* 1987;7:135–158.
- 47 Ellis D, Allaire JC: Modeling computer interest in older adults: the role of age, education, computer knowledge, and computer anxiety. *Hum Factors* 1999;41:345–355.
- 48 Anstey KJ, Luszcz MA: Selective non-response to clinical assessment in the longitudinal study of aging: implications for estimating population levels of cognitive function and dementia. *Int J Geriatr Psychiatry* 2002;17:704–709.
- 49 von Strauss E, Fratiglioni L, Jorm AF, Viitanen M, Winblad B: Attitudes and participation of the elderly in population surveys: data from a longitudinal study on aging and dementia in Stockholm. *J Clin Epidemiol* 1998;51:181–187.
- 50 Collie A, Maruff P, Snyder PJ, Darekar MA, Huggins JP: Cognitive testing in early phase clinical trials: outcome according to adverse event profile in a phase I study. *Hum Psychopharmacol* 2006;21:481–488.
- 51 Silbert BS, Scott DA, Evered LA, Lewis MS, Kalpokas M, Maruff P, Myles PS, Jamrozik K: A comparison of the effect of high- and low-dose fentanyl on the incidence of postoperative cognitive dysfunction after coronary artery bypass surgery in the elderly. *Anesthesiology* 2006;104:1137–1145.
- 52 Weintraub S, Powell D, Whitla D: Successful cognitive aging: individual differences among physicians on a computerized test of mental state. *J Geriatr Psychiatry* 1994;28: 15–34.
- 53 Palmer K, Backman L, Winblad B, Fratiglioni L: Early symptoms and signs of cognitive deficits might not always be detectable in persons who develop Alzheimer's disease. *Int Psychogeriatr* 2008;20:252–258.
- 54 Kiel JM: The digital divide: Internet and e-mail use by the elderly. *Med Inform Internet Med* 2005;30:19–23.
- 55 Gagliardi C, Mazarini G, Papa R, Giuli C, Marcellini F: Designing a learning program to link old and disabled people to computers. *Educ Gerontol* 2008;34:15–29.
- 56 Fillit HM, Simon ES, Doniger GM, Cummings JL: Practicality of a computerized system for cognitive assessment in the elderly. *Alzheimers Dement* 2008;4:14–21.