American Journal of Alzheimer's Disease & Other Dementias®
26(4) 326-333

DOI: 10.1177/1533317511411907
http://aja.sagepub.com
SAGE

Reliability of Repeated Cognitive Assessment of Dementia Using a Brief Computerized Battery

Dustin Hammers, PhD[1], Elizabeth Spurgeon, BS[1], Kelly Ryan, PhD[1], Carol Persad, PhD[1], Judith Heidebrink, MD[2,3], Nancy Barbas, MD, MSW[2,3], Roger Albin, MD[2,4], Kirk Frey, MD, PhD[2,5], David Darby, MD[6,7,8], and Bruno Giordani, PhD[1]

Abstract

Objective: The aim of this study was to evaluate the short-term stability and reliability of a brief computerized cognitive battery in established dementia types. Method: Patients were administered the computerized battery twice with administrations approximately 2 hours apart, with intervening conventional neuropsychological tests. Patients were classified clinically, via consensus conference, as healthy controls (n = 23), mild cognitive impairment (n = 20), Alzheimer's disease (n = 52), dementia with Lewy Bodies ([DLB], n = 10), or frontotemporal dementia (n = 9). Results: Minimal practice effects were evident across CogState test administrations. Small magnitude improvements were seen across all groups on a working memory task, and healthy controls showed a mild practice effect on the accuracy of associative learning. Conclusions: In established dementia, administration of the CogState tasks appears sensitive to cognitive impairment in dementia. Repeat administration also provided acceptable stability and test-retest reliability with minimal practice effects at short test-retest intervals despite intervening cognitive challenges.

Keywords

neuropsychology, assessment, test-retest, CogState, dementia

Introduction

Repeated administration of neuropsychological test batteries has become common practice for evaluating change over time in patient populations. Due to the progressive nature of dementia,[1] serial assessments aid in the delineation of disease course and are important for evaluating the efficacy of novel therapeutic interventions. Identifying techniques for reliable measurement of change in cognitive impairment in patients with dementia is of vital importance for guiding both clinical prognostication and evaluating new therapies at all stages of these diseases. The stability of an instrument is a measurement of consistency, or the ability of the test to give similar results under similar conditions, and is influenced by many factors, including inherent test properties, testing conditions, and features of the disease under evaluation.[2] Test stability can be evaluated over time, as is done via test-retest reliability, or across populations, by assessing the time by population interaction.[2] A common problem of many standard neuropsychological batteries is the occurrence of improvement with repeated exposure to the test (practice effects), which are described variably even a few months later[3] and are significant in conventional tests administered at high frequency.[4] In contrast, the computerized CogState battery is a collection of brief cognitive tests with minimal practice effects when administered repeatedly over a single day in both control populations[5-7] and individuals with psychiatric disease (schizophrenia[8]) and mild cognitive impairment (MCI[9]). It is currently unknown whether this computerized battery displays similar stability across clinical dementia populations, including probable Alzheimer's disease (AD), frontotemporal dementia (FTD), and dementia with Lewy Bodies (DLB).

[1] Neuropsychology Section, Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA
[2] Department of Neurology, University of Michigan, Ann Arbor, MI, USA
[3] Neurology Service, VAAAHS, Ann Arbor, MI, USA
[4] Geriatrics Research, Education, and Clinical Center, VAAAHS, Ann Arbor, MI, USA
[5] Department of Radiology (Nuclear Medicine), University of Michigan, Ann Arbor, MI, USA
[6] CogState Ltd, Melbourne, Victoria, Australia
[7] Florey Neuroscience Institutes, Parkville, Australia
[8] Centre for Neuroscience, University of Melbourne, Parkville, Australia

Corresponding Author:
Dustin Hammers, Neuropsychology Section, Department of Psychiatry, University of Michigan, 2101 Commonwealth Blvd., Ann Arbor, MI, 48105, USA
Email: hammersd@med.umich.edu

The CogState battery used in the current study consisted of 6 subtests designed to rapidly assess psychomotor processing speed, attention, working memory, new learning, divided attention, and associative learning.[10] The purpose of the current study was to determine the stability of this computerized battery over a relatively brief testing interval (2 hours) across subclinical and dementia groups. Similar to previous research with MCI,[9] we anticipated that this battery would display high stability and test-retest reliability across groups. In addition, consistent with prior research,[11] we expected that the dementia groups would perform worse than subclinical groups on this battery.

Methods

Sample and Design

All participants (n = 114) in the current study were enrolled in a longitudinal cohort of the University of Michigan Alzheimer's Disease Research Center (MADRC). Participants were recruited from several avenues including the University of Michigan's Neuropsychology Section in the Department of Psychiatry and the Cognitive Disorders Clinic in the Department of Neurology, as well as through MADRC community outreach. Participants received a neurological evaluation, and individuals with a history of stroke, Traumatic Brain Injury (TBI), and intellectual disability were excluded. Diagnosis of participants was carried out at a consensus meeting consisting of at least 1 neuropsychologist and 2 neurologists, as well as other support staff. Healthy controls (HCs) were determined at consensus to have normal neurological, neuropsychological, and health-related profiles. Participants were classified using the uniform data set (UDS) criteria of the National Alzheimer's Coordinating Center,[12] which include definitions for MCI according to the revised criteria published by Petersen,[13] probable AD according to National Institute of Neurological and Communicative Diseases and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria,[14] and FTD[15] and DLB[16] by consensus conference.

Neuropsychological Measures and PiB Binding

Conventional neuropsychological tests were included to evaluate the domain-specific impairment in the clinical groups and document severity by global scales. In conjunction with their participation in the MADRC, all participants were administered a battery of neuropsychological tests as specified by the UDS test battery,[12] with selected additions. Specific domains were assessed using raw score values from the following measures: for visual and auditory delayed memory, total items remembered from the visual reproduction (VR) and logical memory (LM) subtests from the Wechsler Memory Scale-III were used[17]; working memory and executive functioning was measured using the total raw score for the digit span (DS) subtest from the Wechsler Memory Scale-Revised,[18] the Trail Making Test Part B (TMT[19]) time in seconds, and the total number of correct matches from the Wisconsin Card Sorting Test (WCST[20]); verbal fluency was assessed via the number of words generated on the Controlled Oral Word Association Test (COWA[21]). The Mini-Mental State Examination (MMSE[22]) and the Geriatric Depression scale (GDS[23]) were also administered. Participants performed the computerized CogState battery before and after the completion of conventional neuropsychological testing. Each computerized session lasted approximately 20 to 25 minutes, and the entire testing process took approximately 3 to 4 hours to complete. Testing was restricted to one-half day so as to minimize fatigue and ensure testing was completed efficiently.

To provide supportive evidence of putative AD pathology in the current sample, the majority of participants also underwent a [^{11}C]PiB positron emission tomography (PET) scan on a Siemens ECAT HR^{+} camera operated in 3-dimensional (3-D) mode (septa retracted); 14 of 20 MCI, 45 of 52 AD, 9 of 9 DLB, and 9 of 10 FTD participants received [^{11}C]PiB PET scans, however HC participants were not scanned. [^{11}C]PiB PET images were acquired as a dynamic series of 17 scan frames over a total of 80 minutes.[24] Parametric [^{11}C]PiB distribution-volume ratio images (DVR) were computed by averaging the last 4 scan frames (40-80 minutes) normalized to the mean value of the cerebellar hemisphere gray matter. Standardized participant [^{11}C]PiB PET transaxial image data sets stripped of identifiers were evaluated in a blinded manner by 1 expert interpreter (KF). PiB deposition was judged abnormal if the cortical PiB deposition exceeded subjacent white matter deposition.[24] Visual assessment of cortical PiB deposition has been found to exhibit accuracy comparable to quantitative analyses of PiB binding.[25]

CogState Tasks

Computerized testing was performed in the current study using tasks from the CogState battery. For each task, instructions were provided on the computer screen, followed by a playing card presented facedown in the center of the screen on a green background. After a short interval (around 2.5 seconds), the card turned faceup and the participants were required to respond "yes" or "no" based on questions that varied for each task, as follows:

Detection task (DET): A simple reaction time task that requires the participant to respond as quickly as possible when the central card is turned faceup.

Identification task (IDN): A choice reaction time task that measures visual attention and requires the participant to respond differently if the faceup card is red or black.

One-back task (OBK): An OBK task that assesses working memory and attention and requires the participant to determine whether the faceup card was the same as the preceding card.

One card learning (OCL): A continuous visual recognition learning task that assesses visual recognition memory and attention and requires the participant to determine whether the faceup card had appeared in the current task previously.

Table 1. Demographic and Traditional Neuropsychological Variables for Each Group[a]

Measure	HC	MCI	AD	DLB	FTD	Effect Size
n	23	20	52	9	10	
Age	68.4 (9.5)	73.5 (5.9)	70.8 (8.7)	70.4 (8.5)	64.2 (8.1)	0.08
Gender	47.8%	52.4%	58.8%	77.8%	90%	0.25
Education	16.2 (2.9)	15.7 (3.2)	15.4 (3.7)	14.3 (2.3)	13.7 (3.1)	0.05
% Taking cholinesterase inhibitors	4.3	23.8[b]	80.4[c]	88.9[c]	60.0[c]	
PiB +	–	50.0%	91.1%	44.4%	22.2%	0.45
MMSE	29.1 (.9)	27.0 (2.3)	21.8 (4.8)[c]	22.1 (3.8)[c]	21.4 (3.7)[c]	0.42
GDS	1.5 (2.0)	1.7 (2.1)	1.8 (2.4)[b]	6.0 (4.2)[c,d]	3.8 (2.2)[c,d]	0.20
VR DR	46.5 (26.5)	22.1 (21.2)[b]	5.3 (7.0)[c]	8.6 (12.2)[b]	11.4 (17.1)[b]	0.49
LM DR	27.9 (8.4)	13.2 (9.8)[b]	5.6 (6.7)[c]	9.2 (9.6)[b]	8.3 (8.1)[b]	0.56
WCST	46.1 (6.8)	40.6 (10.1)	31.9 (11.5)[c]	26.8 (11.3)[b]	34.2 (9.2)[c]	0.33
TMT B	71.2 (24.8)	103.7 (46.2)	227.2 (100)[c]	253.3 (113)[c]	209.6 (103)[c]	0.50
COWA	41.9 (9.8)	41.9 (14.2)	29.5 (14.3)[c]	21.4 (11.5)[c]	17.0 (12.3)[c]	0.30
Digit span	8.8 (1.7)	7.8 (1.9)	6.8 (2.5)[c]	6.0 (2.1)[c]	6.8 (2.0)[c]	0.15

Abbreviations: HCs, healthy controls; MCI, mild cognitive impairment; AD, Alzheimer's disease; FTD, frontotemporal dementia; DLB, dementia with Lewy bodies; Gender, percentage of participants, male; PET, positron emission tomography; PiB+, percentage of participants with a positive [^{11}C]PiB PET scan status; MMSE, Mini-Mental State Examination GDS, Geriatric Depression Scale (short-form, cut-off 5/15), VR DR, Visual Reproduction Delayed Recall; LM DR, Logical Memory Delayed Recall; WCST, Wisconsin Card Sorting Test; TMT B, Trail Making Test Part B; COWA, Controlled Word Association Test.

[a] Values represented as means and standard deviations (in parentheses). Effect sizes were measured using partial eta square values (η^2).

[b] Different from HC group only, $P < .001$.

[c] Different from both HC and MCI groups, $P < .001$.

[d] Different from AD group, $P < .001$.

Divided attention (IDM): Assesses the participant's ability to perform parallel cognitive activities, requiring monitoring of the movement of a line of 5 cards horizontally on the computer screen while at the same time performing the OBK task.

Associative learning (ASSL): Requires the participant to determine whether pairs of cards match a legend of 5 card pairs presented at the top of the screen. Upon correctly matching, the participant has to judge subsequent trials based on memory alone.

For each computerized task, speed and or accuracy measures were computed and evaluated. The selection of outcome measures (speed, accuracy, or both) were chosen based on their ability to optimally measure change and reduce ceiling or floor effects. For example, on relatively basic tasks like simple and choice reaction time, accuracy measures were not selected because notable ceiling effects exist for those tasks (most participants were 100% accurate). The speed measure was computed as the mean of the distribution of base 10 logarithmic transformed reaction times in milliseconds, and the accuracy as the arcsine transformed proportion of correct responses (correct responses divided by total responses).[7,10] These transformations aimed to normalize the data distributions for parametric analyses. In addition, test scores that failed to meet test completion criteria ($\geq$75% trials completed) were excluded; 3 participants from the probable AD group were excluded from the study due to poor completion rate during the 2 CogState trials.

Data Analysis

For the primary computerized task analyses, a series of repeated measures analyses of variance (ANOVAs) were performed on each of the computerized task reaction time and accuracy variables in each diagnostic group, but only tasks selected to be optimal in detecting change based on previous pilot data were included in the analyses. The main effects for these analyses were diagnostic group and trial (1-2), and the interaction effect was a group $\times$ trial interaction; for any significant interaction effects, post hoc comparisons were performed using Least Squared Means to determine specific group differences. For the majority of the remaining neuropsychological and demographic test variables, ANOVAs were performed using diagnostic group as the independent variable, and post hoc comparisons were run to determine specific group differences. For categorical-dependent variables (gender, handedness, and PiB status), chi-square analyses were performed with diagnostic group as the independent variable, and secondary post hoc comparisons were also performed. Measures of effect size were expressed as partial eta square (η^2) values for omnibus test analyses, and Cohen *d* values for post hoc analyses.

Stability analyses were performed for each computerized task using test-retest reliabilities between the first and second task trials, utilizing Pearson product-moment correlations. Stability analyses were performed on the HC, MCI, and AD groups only due to the small sample size for the FTD and DLB groups.

Results

Of the 114 overall participants (mean age = 70.2 $\pm$ 8.6 years, mean education = 15.4 $\pm$ 3.4 years, 60% male), 23 were healthy volunteers (HCs), and the rest were classified as follows: 20 participants with MCI, 52 participants with probable AD, 10 participants with FTD, and 9 with DLB. Table 1 shows

Table 2. Baseline (Trial 1) Performance for Each CogState Task Across Diagnostic Groups[a]

Measure	HC	MCI	AD	DLB	FTD	Effect Size
DETs	319.8 (1.3)	333.7 (1.2)	480.5 (1.5)[b]	724.1 (2.3)[b]	553.6 (1.3)[b]	0.26
IDNs	541.5 (1.2)	578.2 (1.2)	726.0 (1.4)[b]	902.4 (1.4)[b]	750.0 (1.2)[b]	0.24
OBKs	816.6 (1.2)	1018.0 (1.3)	1110.7 (1.4)[c]	1600.8 (1.5)[d]	1067.6 (1.3)[c]	0.25
IDMs	503.1 (1.3)	582.5 (1.3)	668.9 (1.4)[c]	885.9 (1.5)[e]	592.0 (1.5)	0.17
OBKa	91.1 (17.6)	73.5 (34.7)[b]	50.1 (23.4)[b]	28.2 (15.4)[d]	63.5 (36.8)[b]	0.51
OCLa	70.2 (12.6)	64.9 (12.2)	54.9 (14.4)[b]	46.3 (10.0)[b]	48.8 (19.4)[b]	0.32
ASSLa	70.9 (17.0)	67.4 (16.8)	50.8 (17.4)[b]	36.8 (18.1)[b]	56.0 (13.0)[c]	0.39

Abbreviations: HCs, healthy controls; MCI, mild cognitive impairment; AD, Alzheimer's disease; FTD, frontotemporal dementia; DLB, dementia with Lewy bodies; DETs, detection task speed; IDNs, identification task speed; OBKs, one-back task speed; IDMs, divided attention speed; OBKa, one-back task accuracy; OCLa, one card learning accuracy; ASSLa, associate learning accuracy.

[a] Values represented as means and standard deviations (in parentheses). Back-transformed units for speed tasks are milliseconds, and accuracy tasks are percentage of correct responses. Effect sizes were measured using partial eta square values (η^2).

[b] Different from both HC and MCI groups, $P < .05$.

[c] Different from HC group only, $P < .001$.

[d] Different from all other groups, $P < .001$.

[e] Different from HC group only, $P < .05$.

demographic, UDS neuropsychological, and [^{11}C]PiB results. The groups were similar in terms of age, education, and gender distribution, therefore these variables were not used as covariates in further analyses. Most participants were of European ancestry (96%). A larger proportion of participants from the dementia groups were prescribed cholinesterase inhibitors at the time of testing than either the HC or MCI groups, and dementia groups had significantly lower MMSE scores than either HC or MCI groups, but no differences existed on the MMSE between the dementia groups (or between HC and MCI groups). The FTD and DLB groups had significantly higher GDS depression scores than the other groups, though all means remained within the normal range indicating that these differences were not clinically significant.

Traditional neuropsychological test data and PiB status were included as a gauge of severity of clinical impairment in our current sample and presence of putative AD pathology in each clinically determined disease group, respectively. No differences existed for the executive functioning (Trails B and WCST), verbal fluency (COWAT), and working memory (DS) measures between HC and MCI groups, or between the particular dementia groups. Most dementia groups, however, had significantly worse performance than both HC and MCI groups. For the auditory and verbal memory measures (VR and LM), both AD and MCI groups were significantly worse than the HC group. In addition, a larger proportion of participants from the AD group was PiB+ (91%) at the time of testing than the MCI group (50%), and the AD group also had a significantly higher proportion of PiB+ individuals than the other dementia groups (44% and 22% for DLB and FTD groups, respectively) as compared to AD groups.

CogState Analyses

In order to measure stability over time across clinical groups, repeated measures ANOVAs were performed for the subtests from the computerized battery. Table 2 displays CogState performance values at trial 1 for each diagnostic group. Figure 1 presents a graphical representation of change in performance over time in the current analyses, utilizing back-transformed change scores from trial 2 - trial 1; negative values in Figure 1 reflect improved reaction time at trial 2, whereas positive values reflect slower reaction time at trial 2. At baseline (trial 1; Table 2), the CogState tasks consistently discriminated dementia groups (AD, DLB, and FTD) from HC or MCI groups for both reaction time and accuracy variables. Except for the OBK, the CogState variables did not differentiate between HC and MCI groups or within dementia groups.

For the repeated administration of the CogState battery, the results are as follows for the reaction time variables: No significant interaction existed between group diagnosis and trial for either the detection (simple reaction time; $P = .60$) or the identification (choice reaction time; $P = .77$) variables, nor did there exist a main effect for trial ($P = .09$ and $P = .41$, respectively). Alternatively, a main effect existed for group diagnosis for both detection speed, $F(4, 88) = 9.10$, $P < .001$, $\eta^2 = .29$, and identification speed, $F(4, 86) = 8.61$, $P < .001$, $\eta^2 = .29$; post hoc tests revealed that there were significant differences between the HC and all dementia groups ($P < .05$ for AD, FTD, and DLB) for both tasks, but no significant differences between HC and the MCI groups on either measure. For the working memory (OBK) and IDM reaction time variables, interaction effects were present (Wilk Lambda = .79, $F(4, 83) = 5.38$, $P < .01$, $\eta^2 = .21$, and Wilk Lambda = .80, $F(4, 81) = 5.18$, $P < .01$, $\eta^2 = .20$, respectively). As seen in Figure 1, post hoc comparisons indicated that the DLB group improved on both the OBK speed and IDM speed tasks (both $P < .05$, $d = 1.06$ and $d = 0.80$, respectively), and to a lesser extent the MCI group became faster on the OBK speed variable ($P < .05$, $d = 0.56$).

No interaction effects existed for the working memory (OBK) or incidental learning (OCL) accuracy subtests, $P = .77$ and $P = .68$, respectively. However, there were main effects for both trial, Wilk Lambda = .83, $F(1, 83) = 16.97$, $P < .001$, $\eta^2 = .17$, and group diagnosis, $F(4, 83) = 26.41$, $P < .001$, $\eta^2 = .56$, for the OBK subtest. As Figure 1 suggests,

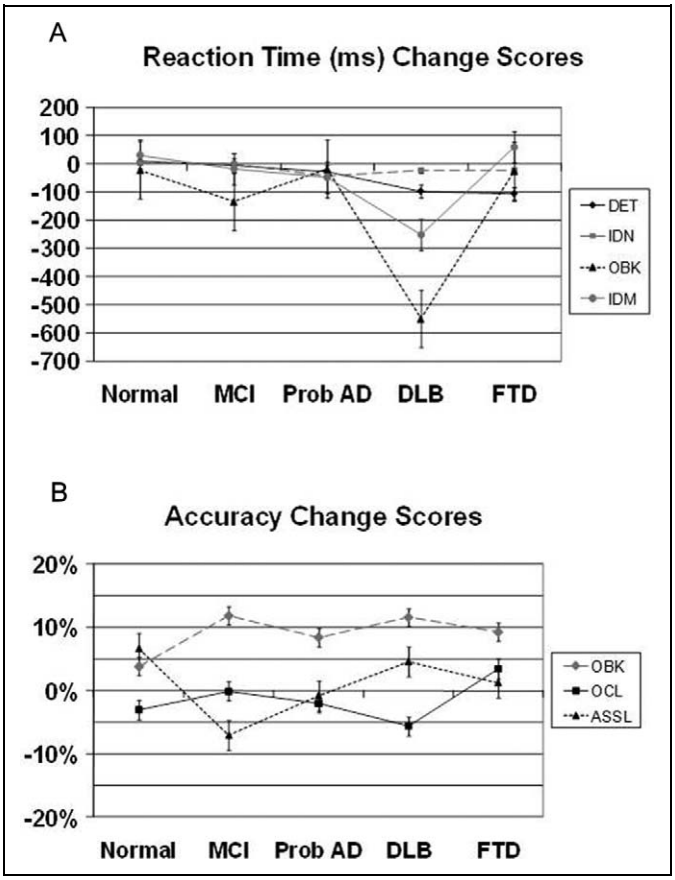

Figure 1. Trial 2 - Trial 1 back-transformed change scores for CogState reaction time (A) and accuracy (B) variables for each group.

there was better performance at trial 2 across groups; also, post hoc tests revealed that the HCs performed better than all other groups ($P < .05$). No main effect for trial existed for the OCL accuracy subtest, $P = .38$. Similar to the other analyses, however, a main effect for diagnostic group existed, $F(4, 82) = 17.47$, $P < .001$, $\eta^2 = .46$, with post hoc tests again revealing that the HC group performed significantly better than all dementia groups ($P < .05$) but not better than the MCI group. Lastly, for the Associated Learning accuracy (ASSL) measure, a significant group × trial interaction effect existed, Wilk Lambda = .87, $F(4, 80) = 3.09$, $P < .05$, $\eta^2 = .13$; as displayed in Figure 1, the HC group became more accurate at trial 2 ($P < .01$), with no significant changes over time in the other groups (despite its appearance, $P = .25$ for the DLB group).

In addition, the coefficients of stability (test-retest reliabilities) between CogState trials 1 and 2 were established using the HC, MCI, and AD groups. See Table 3 for results.

Discussion

The results of this study support prior research[5,9] demonstrating the reliability of a brief computerized battery upon repeated administration, showing this for the first time in various dementing conditions. In the current study, healthy controls and participants with consensus-confirmed subclinical (MCI) or neurological (AD, FTD, and DLB) diagnoses were administered a standard neuropsychological battery in order to determine the severity of cognitive impairments and were administered [^{11}C]PiB PET scans to evaluate putative AD pathology. Similar to the established literature,[26] participants in the dementia groups were impaired on traditional neuropsychological measures relative to the HC and MCI groups, but significant differences were relatively rare between dementia groups. This partly reflects the study of a group of mildly demented individuals. In addition, the AD group had a higher prevalence of individuals with PiB+ status (91%) than the MCI group. This high rate of PiB binding in our AD sample was consistent with the established literature,[27,28] as was our moderate rate of PiB binding for DLB and MCI groups.[28] Our study's 22% PiB+ status for the FTD group is higher than one report in which no PiB binding was present in a clinically diagnosed FTD population.[28] Other clinical samples have shown PiB+ rates of 20% to 25%[27,29] in clinically diagnosed FTD populations. These higher rates suggest these patients may have had mixed FTD/AD dementia or frontal variant AD.[27] Taken together, the neuropsychological and imaging data suggest that our current set of participants reflects a reasonably representative sample for each clinical or subclinical diagnosis.

Table 3. Test-Retest Reliability Values for Control, MCI, and AD Groups

r (Test-Retest)[a]	HC	MCI	AD
DETs	.78[b]	.33	.71[b]
IDNs	.79[b]	.64[c]	.80[b]
OBKs	.78[b]	.73[b]	.59[b]
IDMs	.78[b]	.66[c]	.79[b]
OBKa	.23	.75[b]	.63[b]
OCLa	.54[b]	−.19	.59[b]
ASSLa	.53[c]	.28	.64[b]

Abbreviations: HCs, healthy controls; MCI, mild cognitive impairment; AD, Alzheimer's disease; DETs, detection task speed; IDNs, identification task speed; OBKs, one-back task speed; IDMs, divided attention speed; OBKa, one-back task accuracy; OCLa, one card learning accuracy; ASSLa, associate learning accuracy.
[a] Coefficients of stability were not provided for FTD and DLB groups due to their small sample size.
[b] Coefficient of stability, $P < .01$.
[c] Coefficient of stability, $P < .05$.

For the CogState analyses, similar to previous research,[11] the baseline and repeated administration of the computerized battery displayed utility in discriminating participants with various forms of dementia (AD, DLB, and FTD) from those without (HC and MCI). Although the particular computerized CogState task measures chosen for this study could not differentiate between the dementia groups in the current study, they would likely have clinical utility in discriminating impaired patients from within the general population.

The current study primarily evaluated the stability of a brief computerized battery across diagnostic groups by examining performance change among groups over time, as well as by determining the test-retest reliability for the control, MCI, and AD groups. Overall, the CogState battery appears to

be relatively stable across serial administrations when examining normal and clinical groups over a short period of time. As observed in Figure 1, of the 6 CogState tasks evaluated in the current study, only the working memory (reaction time and accuracy) and IDM (speed) tasks displayed practice effects across groups. For those particular tasks, all groups appeared to improve in accuracy during the working memory task, whereas only a few groups appeared to show improved reaction times for working memory (MCI and DLB groups) or IDM (DLB group); of note, this modest reaction time improvement for the DLB group can likely be explained by a regression toward the mean, for DLB performance was weakest relative to all other groups at trial 1 (see Table 2).

Also of note in our results is the general lack of practice effects from our control group from trial 1 to trial 2, outside of the ASSL task. These results conflict with multiple studies,[5,7,9] which found practice effects between the first 2 trials of a 4-trial administration within relatively short periods of time for control groups. Possible explanations for this difference are related to alterations within task instructions, for Darby et al provided only nonverbal interactive instruction,[9] whereas the current study utilized updated instructions that provided explicit textual and verbal information with the intention of reducing practice effects. Additionally, the CogState tasks selected for the study were redesigned by the manufacturer to minimize practice effects after the first administration for optimal measures of change (personal communication).

A review of the test-retest reliability for the HCs suggests that except for the working memory accuracy variable, the computerized battery also displays reasonable coefficients of stability (Table 3). This is consistent with a prior study in older community volunteers[30] where the only task showing any statistically significant improvement over 5 administrations performed over 12 months was the OBK speed measure. In addition, there did not appear to be major differences in the stability coefficients when comparing AD and control groups, outside of mild differences on the working memory variable. Interestingly, the stability coefficients for the MCI group were smaller than the other groups for most tasks, which may reflect the heterogeneous make-up of the MCI group, consistent with the historically inconsistent performance of MCI groups in published studies.[31] Compared to the previously reported larger Australian sample of controls ($n = 103$),[5] the current results for our control group showed higher test-retest reliabilities for all reaction time variables and for incidental learning and ASSL variables. Despite our small control group sample size, the reaction time variables all displayed correlations greater than .75, and the accuracy of OCL and ASSL coefficients were greater than 0.50. Although $r = .50$ is still below the generally agreed upon level of acceptable test-retest reliability ($r = .70$),[32] all values above .50 were statistically significant ($P < .05$ or greater) at that level. Only the working memory accuracy variable had a coefficient below 0.50 ($r = .23$) for the control group, which is consistent with it being the only variable to display an overall time effect across diagnoses on repeated measures ANOVA. Taken together, the results of this study suggest that this computerized battery provides several reasonable measures of cognition with only minimal learning across groups when repeated after a very short delay.

The current study is not without limitations. Of note were the relatively small sample sizes in each of our diagnostic groups. When compared to a prior test-retest reliability study that included 103 controls,[5] our control sample was smaller ($n = 22$). While it is known that limited sample size may contribute to spurious findings, our American sample displayed results consistent with the findings of the Australian sample. Also, although small sample sizes reduce the statistical power (1-β) available to identify "true" differences in both correlational and group analyses,[33] our sample was large enough to result in statistical significance among the correlational analyses for coefficients of stability at .50 or above. For the repeated measures analyses, reduced sample sizes likely resulted in decreased statistical power to discriminate traditional neuropsychological and CogState performance between dementia groups; given the challenge of recruiting and maintaining DLB and FTD populations, however, it was felt that their positive impact on clinical relevance outweighed their negative impact on statistical power and they were consequently included in the analyses. In the future, larger samples of not only healthy American controls, but FTD and DLB participants, should be evaluated to confirm these results. In addition, the MMSE values for the FTD and DLB are not statistically different to the AD group, suggesting relatively advanced stages of dementia for these 2 groups relative to the AD group.[34,35] Future research should include FTD and DLB populations at earlier stages in the disease course to better evaluate CogState as a screening tool.

Further, the current study only evaluated practice effects across 2 time points, which likely limits conclusions drawn about practice effects for the selected CogState tasks. Though using different training (nonverbal interactional instruction only) than the current study (explicit textual and verbal information), Darby and colleagues[9] continued to show practice effects on CogState tasks for the control group across 4 time points; and although Pietrzak displayed no practice effects from time 1 to time 4, their participants underwent a "practice" trial before time 1 specifically to eliminate practice effects.[8] Given this fact, along with the current study's finding of mild practice effects for the working memory and ASSL tasks at trial 2, replication and extension of the current study over 3 or 4 time points would be important to further support the conclusion of limited practice effects.

In conclusion, the current study supported the reliability and utility of the CogState brief battery as a screening tool when evaluating subclinical (HCs and MCI) and dementia (AD, DLB, and FTD) groups. Repeated administration of the CogState tasks provides similar measures of cognitive ability even at short test-retest intervals and may be useful for the assessment of therapeutic interventions in AD and other dementias.

Declaration of Conflicting Interests

No authors associated with the University of Michigan have reported conflicts of interest.

Dr. Darby is employed by CogState Ltd.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: grant NIH-NIA P50 AG08671 and the Michigan Alzheimer's Disease Research Center. The University of Michigan received no financial benefits for completion of this study from any source other than those listed above.

References

1. Johnson DK, Storandt M, Morris JC, Langford ZD, Galvin JE. Cognitive profiles in dementia: Alzheimer disease vs healthy brain aging. *Neurology*. 2008;71(22):1783-1789.
2. Anastasi A, Urbina S. *Psychological Testing*. Upper Saddle River, NJ: Prentice Hall; 1997.
3. McCaffrey R, Duff K, Westervelt H. *Practitioner's guide to evaluating change with neuropsycholgical assessment instruments*. New York, NY: Kluwer Academic/ Plenum Publishers; 2000.
4. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci*. 2010; 11:118.
5. Collie A, Maruff P, Darby DG. The effects of practice on cognitive test performance of neurologically normal individuals at brief test-retest intervals. *J Int Neuropsychol Soc*. 2003;9(3): 419-428.
6. Dingwall KM, Lewis MS, Maruff P, Cairney S. Reliability of repeated cognitive testing in healthy Indigenous Australian adolescents. *Australian Psychol*. 2009;44(2):224-234.
7. Falleti MG, Maruff P, Collie A, Darby DG. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J Clin Exp Neuropsychol*. 2006; 28(7):1095-1112.
8. Pietrzak RH, Snyder P, Jackson CE, et al. Stability of cognitive impairment in chronic schizophrenia over brief and intermediate re-test intervals. *Hum Psychopharmacol Clin Exper*. 2009;24(2): 113-121.
9. Darby D, Maruff P, Collie A, McStephen M. Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology*. 2002;59(7):1042-1046.
10. Maruff P, Thomas E, Cysique L, et al. 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(2):165-178.
11. Ellis KA, Bush AI, Darby D, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr*. 2009; 21(4):672-687.
12. Morris JC, Weintraub S, Chui HC, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Dis Assoc Disord*. 2006; 20(4):210-216.
13. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256(3):183-194.
14. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services task force on Alzheimer's disease. *Neurology*. 1984;34(7):939-944.
15. Neary D, Snowden JS, Gustafson L, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology*. 1998;51(6):1546-1554.
16. McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology*. 2005;65(12):1863-1872.
17. Wechsler D. *WAIS-III/WMS-III Technical Manual*. San Antonio, TX: The Psychological Corporation; 1997.
18. Wechsler D. *Wechsler Memory Scale-Revised Manual*. San Antonio, TX: The Psychological Corporation; 1987.
19. Armitage SG. An analysis of certain psychological tests used in the evaluation of brain injury. *Psychol Monogr*. 1946; 60(5):1-48.
20. Heaton RK. *Wisconsin Card Sorting Test (WCST)*. Odessa, FL: Psychological Assessment Resources; 1981.
21. Benton A, Hamsher Kd, Sivan A. *Multilingual Aphasia Examination*. 3rd ed. S an Antonio, TX: Psychological Corp.; 1994.
22. Folstein MF, Folstein SE, McHugh PR. Mini-mental State. *J Psychiatric Res*. 1975;12:189-198.
23. Arthur A, Jagger C, Lindesay J, Graham C, Clarke M. Using an annual over-75 health check to screen for depression: validation of the short Geriatric Depression Scale (GDS15) within general practice. *Int J Geriatr Psychiatry*. 1999;14(3): 431-439.
24. Koeppe RA, Frey KA. Equilibrium analysis of [11C]PIB studies. *Neuroimage*. 2008;41(3):T30-T30.
25. Suotunen T, Hirvonen J, Immonen-Räihä P, et al. Visual assessment of [(11)C]PIB PET in patients with cognitive impairment. *Eur J Nucl Med Mol Imaging*. 2010;37(6): 1141-1147.
26. Salmon DP, Bondi MW. Neuropsychological assessment of dementia. *Annu Rev Psychol*. 2009;60(5):257-282.
27. Rabinovici GD, Furst AJ, O'Neil JP, et al. 11C-PIB PET imaging in Alzheimer disease and frontotemporal lobar degeneration. *Neurology*. 2007;68(15):1205-1212.
28. Rowe CC, Ng S, Ackermann U, et al. Imaging beta-amyloid burden in aging and dementia. *Neurology*. 2007;68(20): 1718-1725.
29. Engler H, Santillo AF, Wang SX, et al. In vivo amyloid imaging with PET in frontotemporal dementia. *Eur J Nucl Med Mol Imaging*. 2008;35(2):100-106.
30. Fredrickson J, Maruff P, Woodward M, et al. Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology*. 2010; 34(2):65-75.

31. Collie A, Maruff P, Currie J. Behavioural characterization of mild cognitive impairment. *J Clin Exp Neuropsychol*. 2002;24:720-733.
32. Nunnally JO. *Psychometric Theory*. New York, NY: McGraw-Hill; 1978.
33. Sharp DS, Gahlinger PM. Regression analysis in biological research: sample size and statistical power. *Med Sci Sports Exerc*. 1988;20(6):605-610.
34. Chow TW, Hynan LS, Lipton AM. MMSE scores decline at a greater rate in frontotemporal degeneration than in AD. *Dement Geriatr Cogn Disord*. 2006;22(3):194-199.
35. Nelson PT, Kryscio RJ, Jicha GA, et al. Relative preservation of MMSE scores in autopsy-proven dementia with Lewy bodies. *Neurology*. 2009;73(14):1127-1133.