

# Longitudinal Comparison of in Clinic and at Home Administration of the Cogstate Brief Battery and Demonstrated Practice Effects in the Mayo Clinic Study of Aging

N.H. Stricker<sup>1</sup>, E.S. Lundt<sup>2</sup>, E.C. Alden<sup>1</sup>, S.M. Albertson<sup>2</sup>, M.M. Machulda<sup>1</sup>, W.K. Kremers<sup>2</sup>, D.S. Knopman<sup>4</sup>, R.C. Petersen<sup>4</sup>, M.M. Mielke<sup>3,4</sup>

1. Division of Neurocognitive Disorders, Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota, USA; 2. Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA; 3. Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA; 4. Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA

Corresponding Author: Nikki H. Stricker, Ph.D., Mayo Clinic, 200 First Street SW, Rochester, MN 55905; 507-284-2649 (phone), 507-284-4158 (fax), Stricker.Nikki@mayo.edu

J Prev Alz Dis 2020;1(7):21-28  
Published online September 16, 2019, <http://dx.doi.org/10.14283/jpad.2019.35>

## Abstract

**BACKGROUND:** The Cogstate Brief Battery (CBB) is a computerized cognitive assessment that can be completed in clinic or at home.

**DESIGN/OBJECTIVE:** This retrospective study investigated whether practice effects / performance trajectories of the CBB differ by location of administration.

**PARTICIPANTS/SETTING:** Participants included 1439 cognitively unimpaired individuals age 50-75 at baseline participating in the Mayo Clinic Study of Aging (MCSA), a population-based study of cognitive aging. Sixty three percent of participants completed the CBB in clinic only and 37% completed CBB both in clinic and at home.

**MEASUREMENTS:** The CBB consists of four subtests: Detection, Identification, One Card Learning, and One Back. Linear mixed effects models were used to evaluate performance trajectories in clinic and at home.

**RESULTS:** Results demonstrated significant practice effects between sessions 1 to 2 for most CBB measures. Practice effects continued over subsequent testing sessions, to a lesser degree. Average practice effects/trajectories were similar for each location (home vs. clinic). One Card Learning and One Back accuracy performances were lower at home than in clinic, and this difference was large in magnitude for One Card Learning accuracy. Participants performed faster at home on Detection reaction time, although this difference was small in magnitude.

**CONCLUSIONS:** Results suggest the location where the CBB is completed has an important impact on performance, particularly for One Card Learning accuracy, and there are practice effects across repeated sessions that are similar regardless of where testing is completed.

*Key words:* Neuropsychology, computerized testing, cognitively unimpaired, memory, reaction time.

## Introduction

There is growing interest in improving access to cognitive assessment tools for research and clinical use by allowing for administration of

cognitive measures in unsupervised settings, including at home. The ability to complete cognitive measures at home has implications for enriching clinical trials, and a large scale demonstration of this goal is already underway through the Brain Health Registry (1). At home assessment can also facilitate collection of clinical outcome data by allowing collection of follow-up data for participants who may live far from clinical centers or for longer periods of time than otherwise feasible when a clinic visit is required. As such, developing computerized testing platforms that can be reliably administered both in supervised clinical settings and unsupervised settings is important. Cogstate is one computerized platform that has numerous tests available, including the Cogstate Brief Battery (CBB), which consists of four subtests that measure attention, working memory, processing speed, and visual learning. In addition, prior data has demonstrated the feasibility of using the CBB in both supervised and unsupervised settings (2). This makes Cogstate an appealing option for use as a screening measure that can be completed in the clinic or home environment.

There is also significant interest in cognitive measures that may help identify individuals at risk of mild cognitive impairment (MCI) or dementia and who would benefit from further clinical work up. Given CBB can be completed at home or during general medical appointments, it may be able to address this clinical need. In addition, existing studies suggest CBB is sensitive to early cognitive decline, and that CBB performance can help identify individuals with cognitive impairment (3, 4). The CBB has FDA approval under the name Cognigram™ for use as a digital cognitive assessment tool in individuals 6-99 years of age. However, it is not known whether there are differences in performance across supervised and unsupervised settings that could differentially impact the sensitivity of the CBB. Currently, the CBB is being used in large scale epidemiological and experimental studies. While some studies did not find evidence for differences in performance across

supervised and unsupervised settings (2), other research has suggested potential variations in performance. For example, the Mayo Clinic Study of Aging (MCSA) has administered Cogstate since 2012, and pilot data ( $n = 194$ ) for participants completing the CBB first in clinic and then at home within 6 months showed small but statistically significant performance differences by location of administration, with participants performing faster at home compared to in the clinic (5). Therefore, additional data with larger sample sizes and more follow-up sessions are needed to fully evaluate the integrity of completing CBB in supervised and unsupervised settings, and to determine whether there are performance differences when individuals complete CBB in clinic and at home.

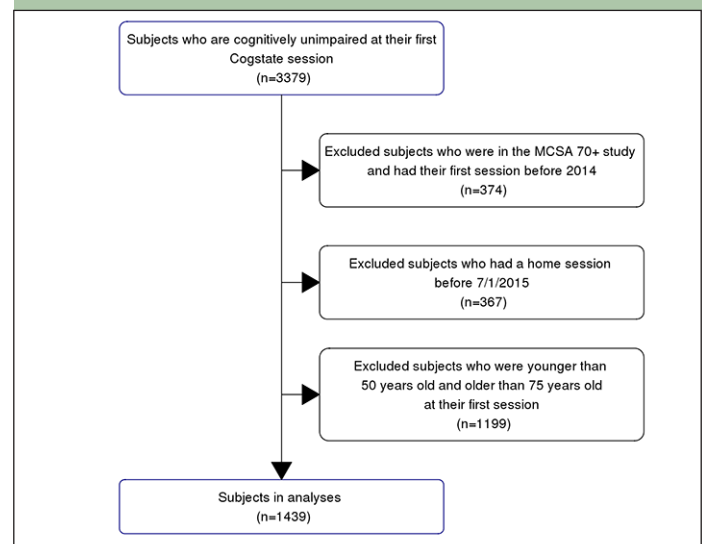
In addition to the ability to complete CBB in both supervised and unsupervised settings, CBB measures were designed to minimize practice effects by having randomly generated alternative forms each time an individual takes the test. This is an important aspect of CBB because one challenge of detecting cognitive decline in older adults is that decline may be relatively subtle but practice effects associated with repeat testing over time can be quite robust. For example, research using traditional neuropsychological assessments suggests practice effects can occur between baseline and follow-up visits on measures of learning and memory, even in individuals with incident MCI (6). Similarly, prior research using CBB over multiple sessions indicates the strongest practice effects occur between the first and second assessment (7). However, while most CBB practice effects stabilized after the third evaluation, sustained practice effects were observed for One Card Learning accuracy (7). Additionally, another study of older adults completing CBB in an unsupervised testing environment demonstrated continued practice effects over multiple testing sessions (8). Given this evidence of practice effects on the CBB, further research is needed to clarify the nature of these practice effects and determine whether practice effects differ across supervised and unsupervised testing environments.

The primary aim of this study was to investigate whether practice effects / performance trajectories of the CBB differ across supervised and unsupervised settings, represented within this study by location of administration (clinic vs. home, respectively). Secondary aims were to 1) further assess differences in test performance across location of administration with a larger independent sample, controlling for the known session 1 to session 2 practice effect that may have confounded our preliminary home vs. clinic analyses (5); and 2) further assess practice effects on the CBB across additional follow-up sessions.

## Method

The MCSA is a population-based study of cognitive aging among Olmsted County, MN, residents. It began in October 2004 and initially enrolled individuals aged 70 to 89 years with follow-up visits every 15 months. The details of the study design and sampling procedures have been previously published; enrollment follows an age- and sex-stratified random sampling design to ensure that men and women are equally represented in each 10-year age strata (9). In 2012, enrollment was extended to cover the ages of 50-90+ following the same sampling methods. Administration of Cogstate began in 2012 for newly enrolled 50-69 year olds and in 2013 for those aged 70 and older during clinic visits. From September 2013 through March 2014 and September 2014 through July 2015 we piloted administration of the CBB at home among MCSA participants in our 50-69 year old cohort. This pilot data ensured acceptability of the at home testing process in our participants and preliminary analyses demonstrated generally comparable performance in the clinic versus at home (5). Although participants did perform faster at home, this difference was viewed as small in magnitude and at-home testing was offered to all MCSA participants starting July 2015. Individuals who completed the pilot testing ( $n = 380$ ) were not included in the current analysis. There was a trend toward fewer follow-up sessions available for individuals over 75 due to a temporary cap on the number of Cogstate sessions in the protocol for older participants. We therefore limited the participants in the current study to individuals who were between the ages of 50-75 at the time of their first Cogstate session.

**Figure 1.** Study flow chart



Study visits included a neurologic evaluation by a physician, an interview by a study coordinator, and neuropsychological testing by a psychometrist (9). The physician examination included a medical history

review, complete neurological examination, and administration of the Short Test of Mental Status (10). The study coordinator interview included demographic information and medical history, and questions about memory to both the participant and informant using the Clinical Dementia Rating (CDR®) Dementia Staging Instrument (11). See Roberts et al. (9) for details about the neuropsychological battery.

For each participant, performance in a cognitive domain was compared with age-adjusted scores of cognitively unimpaired (CU) individuals using Mayo's Older American Normative Studies (12). Participants with scores of  $\geq 1.0$  SD below the age-specific mean in the general population were considered for possible cognitive impairment. A diagnosis of mild cognitive impairment (MCI) or dementia was based on a consensus agreement between the interviewing study coordinator, examining physician, and neuropsychologist, after a review of all participant information (9, 13). Performance on Cogstate was not available for review during consensus conference and thus independent of diagnosis. Individuals who did not meet criteria for MCI or dementia were deemed CU and were eligible for inclusion in the current study. Data for participants who were CU at the time of their first Cogstate session but later were assigned a diagnosis of MCI or dementia at a follow-up visit were included until the visit with the diagnosis to avoid biasing our sample toward individuals with less follow-up data available (also see Supplemental Results for sensitivity analyses).

The study protocols were approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards. All participants provided written informed consent.

### ***Cogstate Brief Battery***

All participants completed their first Cogstate session in clinic. Cogstate was administered on a PC or iPad during MCSA clinic visits (every 15 months), and our prior work describes small platform differences on select outcome variables (14). Participants are permitted to choose whether to complete Cogstate in clinic or at home in between full MCSA study visits. Home testing was completed on a PC through a web browser (i.e., not on an iPad, tablet or phone). Participants electing to complete Cogstate in clinic returned for an in clinic visit at 7.5 month intervals. Participants electing to complete Cogstate at home were sent an email and prompted to complete the testing at 4-month intervals, with a reminder after 2 weeks if not completed.

Each Cogstate administration included a short practice battery followed by a 2-minute rest period and then the complete battery. The practice battery was not used in any analyses. For in clinic visits, the study coordinator was available to help the participants understand the tasks during the practice session. During the test session, the coordinator provided minimal supervision or assistance and typically waited in another room for the

participant to finish. The ability to reliably complete and adhere to the requirements of each task was determined by completion and integrity checks as previously described (14). All data values with a failed completion were excluded and failed integrity values were included and examined as potential outlier values.

Cogstate subtests were administered in the order listed below. Accuracy and reaction time measures were transformed by Cogstate applying a logarithmic base 10 transformation to reaction time data (milliseconds) and arcsine square root transformation to the proportion correct (accuracy) in order to improve normality and reduce skewness.

Detection is a simple reaction time (RT) paradigm that measures psychomotor speed. Participants press "yes" as quickly as possible when a playing card turns face up. RT for correct responses was the primary outcome measure, which is often referred to as speed in other Cogstate manuscripts.

Identification is a choice RT paradigm that measures visual attention. Participants press "yes" or "no" to indicate whether or not a playing card is red as quickly as possible. RT for correct responses was the primary outcome measure. A linear correction was applied to each PC time point (in clinic and at home) to correct for small PC-iPad performance differences in clinic as previously described (14).

One Card Learning is a continuous visual recognition learning task that assesses learning and attention. Participants press "yes" or "no" to indicate whether or not they have seen the card presented previously in the deck. Accuracy was the primary outcome measure.

One Back assesses working memory. Participants press "yes" or "no" to indicate whether or not the card is the same as the last card viewed (one back) as quickly as possible. Accuracy was the primary outcome measure.

### ***Statistical Methods***

Each individual had repeated measures from regular testing where the location was in clinic approximately every 15 months and either in-clinic or at home during the interim. Linear mixed effects (LME) models with random subject-specific intercepts and slopes were used to assess differences between testing locations in each response measure. All models were adjusted for age, sex, and education as additive effects. We captured practice effects using a piecewise linear spline with a bend at session two parameterized with two variables: first practice (sessions 1 to 2) and subsequent practice (session 2 to 3, 3 to 4, and so on). A positive beta for first practice and subsequent practice both imply increasing scores with more practice for accuracy measures, and a negative beta implies improved performance with more practice for RT measures. To assess whether learning effects differed by location we included an interaction between subsequent practice and location. Non-linearity

**Table 1.** Participant demographics at baseline visit

	Total (N=1439)	Clinic Sessions Only (N=900)	Clinic & Home Sessions (N=539)	p value
Baseline Age Mean (SD)	62.54 (6.82)	63.11 (6.75)	62.58 (6.84)	< 0.001
Sex (Percent Male)	50.2%	50.8%	49.4%	0.600
Years Education Mean (SD)	15.01 (2.35)	14.71 (2.30)	15.50 (2.37)	< 0.001
Cogstate Sessions Mean (SD)	6.75 (4.27)	5.18 (3.48)	9.37 (4.18)	< 0.001
Follow-up (years) Mean (SD)	3.21 (2.19)	2.88 ( 2.28)	3.77 (1.90)	0.044

Note. SD = standard deviation. P-values reported above are from linear model ANOVAs (continuous variables) or Pearson's Chi-square test (frequencies).

**Table 2.** Linear mixed effects regression parameter estimates (standard errors) for predicting four cognition measures

Fixed effects	Detection RT		Identification RT		One Card Learning Accuracy		One Back Accuracy	
	$\beta$ (SE)	p	$\beta$ (SE)	p	$\beta$ (SE)	p	$\beta$ (SE)	p
Age at session	0.0031(0.0003)	<0.001	0.0021(0.0002)	<0.001	-0.0025(0.0003)	<0.001	-0.0026(0.0004)	<0.001
Sex (1=male, 0=female)	-0.0214(0.0045)	<0.001	-0.0035(0.003)	0.25	0.0005(0.0041)	0.90	-0.0163(0.0052)	0.002
Education	-0.0036(0.001)	<0.001	-0.0018(0.0007)	0.005	0.0045(0.0009)	<0.001	0.007(0.0011)	<0.001
First practice	0.0012(0.0027)	0.67	-0.0182(0.0019)	<0.001	0.0425(0.0024)	<0.001	0.018(0.005)	<0.001
Subsequent practice (2+)	-0.0048(0.0004)	<0.001	-0.0015(0.0003)	<0.001	0.0076(0.0004)	<0.001	0.0066(0.0007)	<0.001
Location (Clinic vs. Home)	-0.0099(0.0038)	0.010	0.0001(0.0028)	0.97	-0.0609(0.0037)	<0.001	-0.0316(0.0072)	<0.001
Subsequent Practice x Location	0.0011(0.0006)	0.06	0.0008(0.0004)	0.08	-0.0002(0.0006)	0.79	0.0012(0.0011)	0.30
<i>SD of random effects</i>								
Intercept	0.7758		0.0479		0.0627		0.0855	
Slope first practice	0.0489		0.0144		0.0124		0.0599	
Slope subsequent practice	0.0059		0.0019		0.0030		0.0043	
Residual	0.0669		0.0523		0.0688		0.1393	

Note. RT = reaction time. SE = standard error. SD = standard deviation. Age in years; sex is 1 for males and 0 for females; education is total years with all individuals less than 11 coded as 11; First practice is 0 = first session and 1 = second session; Subsequent practice (2+) is 0 = 2nd session, 1 = 3rd, 2 = 4th, and so on; location is 0 = clinic and 1 = home. Taken together first practice and subsequent practice comprise a piecewise linear spline with a bend at session 2. For Detection and Identification, values represent logarithmic base 10 transformation for reaction time data (collected in milliseconds) and negative beta estimates signify better/improved performance. For One Card Learning and One Back, values represent arcsine transformation for accuracy data and positive beta estimates signify better/improved performance.

of the subsequent practice was also tested by including natural log of subsequent practice but did not reach significance at the 0.05 level and was excluded from our final models.

The distribution of the response variable is known to reflect the distribution of the residuals. Many individuals scored near ceiling (perfect score) on One Back accuracy resulting in a skewed, non-normal distribution despite the transformation applied. However, LME models assume residual errors follow normal distributions. In order to assess the appropriateness of LME for One Back accuracy we fit a second Generalized Linear Mixed effects model (GLMM) with a binomial link. For this response variable, the number of correct responses out of X trials is known to follow a binomial distribution; hence we expect the error distribution to be binomial as well. We observed that our LME model may be biased toward the mean especially at early sessions when compared to the GLMM; however the estimated difference between at home and in-clinic was comparable, thus we report the LME model for consistency across response variables. In our case, the large sample size allowed for a robust estimate of group difference despite the slight departure

from normality. Analyses were conducted using statistical software R version 3.4.1.

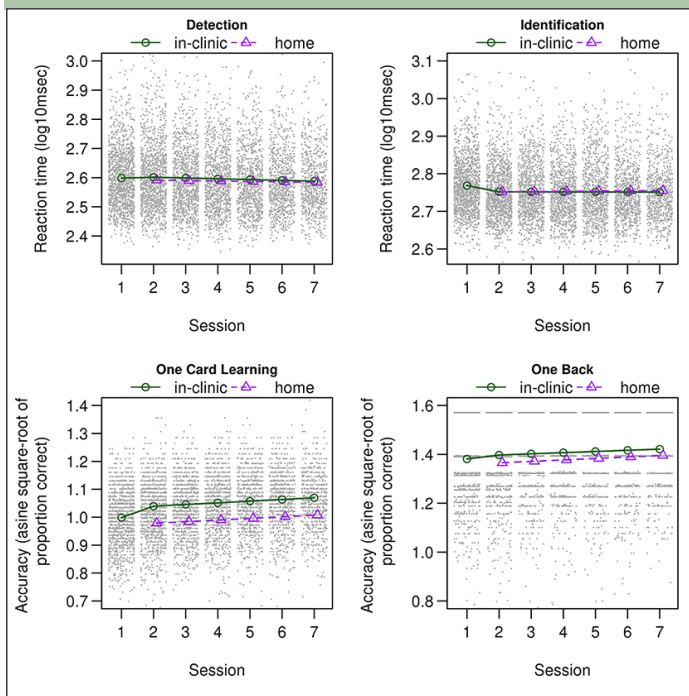
## Results

### Participants

See Table 1 for participant demographics. There were 1439 CU individuals who completed the CBB, with a mean of 3.2 years of follow-up data; over half the sample (50.2%) completed 7 or more Cogstate sessions and 6% completed 14 or more sessions. Sixty three percent of the sample completed Cogstate in clinic only and 37% completed Cogstate both in clinic and at home. Participants completing Cogstate both in clinic and at home were slightly younger, had higher years of education, and had more follow-up sessions available (as expected given the study design), but did not differ by sex. Of all Cogstate sessions completed in clinic, 34.5% were completed on a PC (N = 2559) and the remainder were completed on an iPad (N = 4850). Of all Cogstate sessions completed, 22.8% of sessions (N = 2192) were at home. Completion flag failures were infrequent (< 0.3%;

see Supplemental Table 1), with all completion failures occurring in clinic and none at home. Integrity flag failure rates were comparable across clinic and home sessions for most subtests, although there was a slightly greater rate of integrity failures on One Card Learning at home sessions (2%) relative to in clinic sessions (1%;  $p < .001$ ). Because of this slight difference, we performed sensitivity analyses for One Card Learning and determined that conclusions did not change when excluding data points with a failed integrity flag. We retain this data in our models to improve generalizability of results.

**Figure 2.** Mean of Cogstate scores as a function of session number and location (clinic versus home) for a cognitively unimpaired population 63 years of age with 16 years of education and averaged between males and females, overlaid on scatterplots



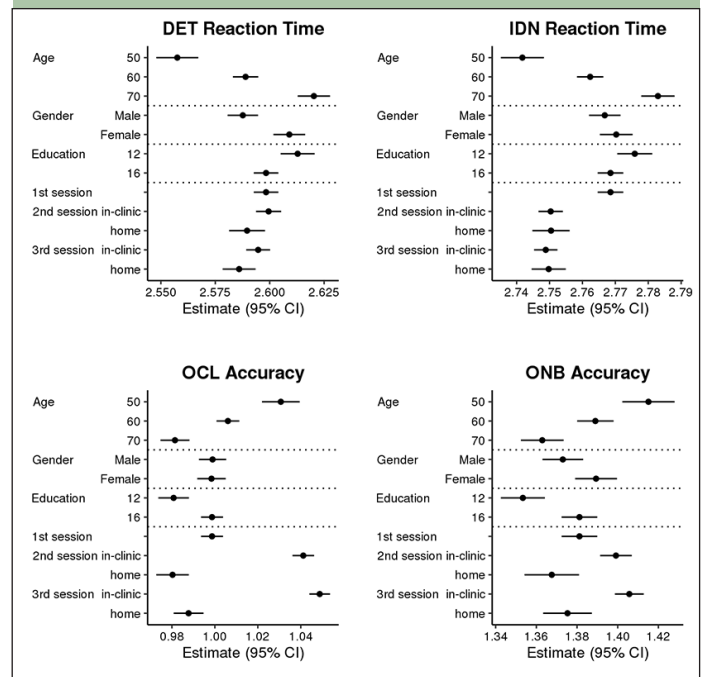
Note. CI = confidence interval. For Detection and Identification, values represent logarithmic base 10 transformation for reaction time data (collected in milliseconds) and lower values signify better performance. For One Card Learning and One Back, values represent arcsine transformation for accuracy data and higher values signify better performance.

### Significant practice effects

Results demonstrate significant practice effects for most Cogstate measures (see Table 2). Consistent with our earlier results that demonstrated the most pronounced practice effect at session 2 (7) and our own exploration of additional types of models in an effort to find the best way to model the observed data, inspection of Figure 2 and Table 2 demonstrates a clear practice effect from session 1 to session 2 on most CBB measures ( $p$ 's  $< .001$ ), except Detection RT. The magnitude of the session 1 to session 2 practice effect is large relative to the magnitude of performance differences associated

with demographic variables, particularly for One Card Learning accuracy (see Figure 3). In addition, practice effects continued across all additional follow up sessions for all Cogstate measures ( $p$ 's  $< .001$ ), although visual inspection of Figure 3 suggests continued practice effects are minimal for RT measures despite coefficients reaching significance.

**Figure 3.** Estimated mean score (95% CI) in cognitive measures from linear mixed-effects models for specified levels of each feature. In the absence of interactions, additive effect age (or sex, or education) shifts estimates up or down by a fixed amount per year of age. Unless indicated in the y-axis description we used age of 63 years, education 16 years, location in-clinic, first session, and averaged between males and females for estimating effects



Note. CI = confidence interval; DET = Detection; IDN = Identification; OCL = One Card Learning; ONB = One Back.

### Practice effects by location were similar

For all Cogstate measures, we did not observe an interaction between location and number of Cogstate sessions ( $p$ 's  $> .05$ ). These results suggest that participants show the same cognitive trajectory / degree of practice effect across Cogstate sessions regardless of completion location.

### Faster at home on Detection

Participants showed faster performance at home on Detection RT ( $p < .01$ ). Visual inspection of Figure 2 suggests that this difference may become less pronounced over repeated sessions, although the interaction failed to reach significance, likely related to significant noise for

this response variable. There was no significant difference across performance in clinic and home for Identification RT.

### *Less accurate at home*

Visual inspection of Figure 1 and results in Table 2 show that One Card Learning and One Back accuracy are lower at home than in clinic ( $p < .001$ ).

### *Magnitude of study findings*

Applying internal Cogstate normative data for change to our model estimated effects provides some insight into the potential impact of our main study findings (15). We used the location (clinic vs. home) estimate from our model (see Table 2) and applied a Reliable Change Index formula (16) that uses within-subject standard deviation provided as part of their normative data to generate a z-score for change across sessions that helps illustrate the size of the effect relative to sampling variability across two sessions. The location of administration difference alone yielded z-scores for change of 0.10 on Detection RT and -.47 on One Card Learning accuracy. Similarly, the estimated difference from session 1 to 2 on One Card Learning accuracy (see Table 2) yields a z-score for change of 0.33. The practice effect is lesser in magnitude after the second session (0.06 z for each interval), but sessions 2 to 7 had an aggregate z-score for change of .30 (approximately the magnitude of the first practice effect). Together, the change from baseline to session 7 would yield a z-score for change of 0.63. That is, the average participant in our sample shows nearly a 2/3 standard deviation improvement in One Card Learning accuracy performance at session 7 relative to baseline. To assist with interpretation of the size of these effects, we also re-ran models on raw, untransformed Cogstate variables and results are presented in Supplementary Table 2 (S2) and Supplemental Results.

## **Discussion**

The main finding of this study is that practice effects on the CBB do not differ by location of administration. In accordance with our secondary aims, we also demonstrate that: 1) there are important differences across CBB sessions completed in clinic and at home, and 2) there are practice effects on the CBB from session 1 to session 2; moreover practice effects continue across numerous follow-up sessions.

This study adds to growing evidence suggesting that practice effects must be considered when using the CBB, despite careful efforts by Cogstate to mitigate practice effects by implementing a short practice trial before CBB administration and the use of randomly generated alternate forms. This is consistent with prior

work showing that alternate forms reduce, but do not eliminate, practice effects (17). This is because a number of other factors besides memory for specific test items can lead to practice effects, including familiarity and increased comfort with test procedures and learned strategies for successfully navigating task demands (18). Our results show a clear practice effect from session 1 to session 2 for most CBB measures (3 out of 4). For example, the difference between session 1 and session 2 on One Card Learning accuracy is similar in magnitude to the average difference between a 50 and 70 year old, and is an approximately 3.7% improvement on the raw score scale (see Figure 3 and S2). The magnitude of the practice effect from session 1 to 2 is smaller for IDN and ONB accuracy, and slightly less than a 10 year age difference. There are also continued practice effects after session 2 for all variables. These continued practice effects are more notable for accuracy measures, particularly One Card Learning. Continued practice effects are considered negligible for RT measures. Consistent with our results, Valdes et al. (8) similarly showed significant practice effects on 3 out of 4 CBB measures when administered monthly at home in a sample of older adults. They also found that prior computer use impacted the practice effect observed on One Card Learning and One Back accuracy, with greater improvements over sessions seen in individuals with less frequent computer use. Overall, our results suggest there are practice effects on CBB, particularly across sessions 1 and 2. However, practice effects, or performance trajectories, of CBB performance are similar regardless of where an individual completes the CBB.

Although trajectories of CBB performance are similar across location of administration, our results suggest there are some potentially important performance differences by location. Most notably, individuals perform less accurately at home than in clinic (approximately 5.3% lower at home on raw percent accuracy; see S2). Figure 3 demonstrates the magnitude of this effect relative to the impact of other variables in the model. For One Card Learning accuracy, the difference between clinic and home performance at session 2 is slightly greater than the effect of 20 years difference in age; it is also greater than the initial practice effect from session 1 to session 2. Our previously published pilot data that were not included in the current study did not show a significant One Card Learning accuracy difference, and did not include One Back accuracy as an outcome variable (5). The session 1 to session 2 practice effect on One Card Learning accuracy may have obscured finding a home vs. clinic difference in our prior manuscript, as within that study all participants completed CBB in the clinic first, then at home. Participants also completed Cogstate in clinic first within the current study, but because participants who completed Cogstate at home also had subsequent Cogstate sessions in clinic, and our model takes into account the number of Cogstate

sessions, this allows us to better estimate the home effect. Findings by Cromer et al. (2) did not show differences in any CBB measure compared across supervised and unsupervised sessions, including One Card Learning accuracy. Importantly, that study counterbalanced test order, although the sample size was small ( $n = 57$ ) limiting the power to detect subtle differences. Similar to our earlier findings, we again found evidence that participants perform faster at home than in clinic on Detection RT, but we interpret the magnitude of this effect as small and not clinically meaningful. Our prior results also showed faster performance on One Back RT and One Card Learning RT, which were not included as primary response variables for the current study. Consistent with our prior results, there was no difference in Identification RT across location.

These results have important implications for future study designs. For studies focusing on detecting change in performance over time, investigators should consider study designs that would help minimize change due to practice effects. Because the largest practice effects were observed between session 1 and 2 in our data, administering two CBB sessions during the baseline visit, and excluding session 1 from further analysis could be considered, as using CBB session 1 as a benchmark for future change may obscure true decline given typical practice effects. Valdes et al. (8) also recommended this approach based on their data. Similarly, we recommend investigators choose to either have participants complete the CBB in the clinic for all sessions, or have participants complete all sessions at home. If home administration is the only feasible option for longitudinal follow-up, investigators could choose to administer the first CBB in clinic to familiarize participants with the procedure as described above, but use a first CBB session administered at home shortly after as the longitudinal baseline. For populations that are less familiar with using computers at home, three baseline sessions could even be considered; one in clinic, a second at home to familiarize participants with procedures for completing the test at home, and a third session at home to serve as the study baseline.

These results also have important implications for the application of the CBB for clinical use or to inform diagnostic status in research studies. A working memory/learning composite score based on supervised One Card Learning and One Back accuracy performance has previously demonstrated good sensitivity and specificity for differentiating individuals with MCI and AD dementia from cognitively unimpaired individuals (4). Given our findings of significantly lower accuracy performance at home, it will be important to validate the diagnostic accuracy of the CBB in unsupervised settings.

Cogstate was designed to be sensitive to detecting change over time. Although the CBB has demonstrated sensitivity to memory decline in CU and MCI participants with high amyloid based on Pittsburgh compound B (PiB)-positron emission tomography neuroimaging

(20), studies have not yet been done to demonstrate whether the application of available internal Cogstate norms for change is sensitive to subtle cognitive decline at the level of the individual, particularly for detection of MCI and AD dementia. In individuals with concussion, change on the CBB (aka CogSport/Axon) using study specific normative data (21) was less sensitive but more specific for concussion relative to a single assessment. Our finding of practice effects that continue after session 2 raises questions about the best method for determining whether a significant change has occurred beyond a single follow-up assessment. Internal Cogstate normative data provide within-subject standard deviation (WSD) values for most CBB primary outcome variables using a test-retest interval of approximately 1 month (15). This can be used in a reliable change index formula to calculate a z-score for change (16, 22). This helps take into account the session 1 to session 2 practice effect when determining if a change is significant, but it is not clear if application of this single interval WSD is appropriate for subsequent follow-up sessions, particularly for One Card Learning accuracy that demonstrated the most robust continued practice effect over time in our results. For these reasons, the use of a control group is critical for clinical trials using Cogstate as an outcome measure. Because the internal normative data provided by Cogstate are not in the public domain, this limits the ease of reproducing our illustration of the magnitude of these results.

Strengths of our study include the population-based design and large sample size. There are also several limitations. First, all participants completed Cogstate in the clinic first, which is a significant confound when comparing CBB performance in clinic and at home. A counter-balanced design would provide a better test of home-clinic differences. Second, the differing follow-up intervals across participants electing to complete Cogstate only in clinic versus also at home complicated interpretation of these results, but sensitivity analyses suggest this did not significantly impact findings (see Supplemental Results). Future studies would benefit from using the same follow-up interval regardless of location of administration. Although the fact that participants complete Cogstate in clinic and at home complicates comparison of trajectories, it also helps us to better estimate home vs. clinic effects. Future studies would benefit from including a measure of the frequency of computer use and confidence with computers to help determine whether that impacts results. Future studies would also benefit from examining whether practice effects are also observed in clinical populations, such as individuals with MCI, as this has been reported on some traditional neuropsychological tests (6).

In summary, results suggest the location where the CBB is completed has an important impact on performance, particularly for One Card Learning accuracy. CBB performance over time is influenced

by practice effects on most measures, which are most prominent from session 1 to 2, but also continue over time. Practice effects and trajectories of performance over time are similar regardless of where testing is completed.

*Funding:* This work was supported by the Rochester Epidemiology Project (R01 AG034676), the National Institutes of Health (grant numbers P50 AG016574, U01 AG006786, and R01 AG041851), a grant from the Alzheimer's Association (AARG-17-531322), the Robert Wood Johnson Foundation, The Elsie and Marvin Dekelboum Family Foundation, Alzheimer's Association, and the Mayo Foundation for Education and Research. NHS and MMMi serve as consultants to Biogen and Lundbeck. DSK serves on a Data Safety Monitoring Board for the DIAN-TU study and is an investigator in clinical trials sponsored by Lilly Pharmaceuticals, Biogen, and the University of Southern California. RCP has served as a consultant for Hoffman-La Roche Inc., Merck Inc., Genentech Inc., Biogen Inc., Eisai, Inc. and GE Healthcare. The sponsors had no role in the design and conduct of the study; in the collection, analysis, and interpretation of data; in the preparation of the manuscript; or in the review or approval of the manuscript. The authors report no conflicts of interest.

*Acknowledgements:* The authors wish to thank the participants and staff at the Mayo Clinic Study of Aging.

## References

- Weiner MW, Nosheny R, Camacho M, Truran-Sacrey D, Mackin RS, Flenniken D, et al. The Brain Health Registry: An internet-based platform for recruitment, assessment, and longitudinal monitoring of participants for neuroscience studies. *Alzheimers Dement*. 2018;14:1063-76.
- Cromer JA, Harel BT, Yu K, Valadka JS, Brunwin JW, Crawford CD, et al. Comparison of Cognitive Performance on the Cogstate Brief Battery When Taken In-Clinic, In-Group, and Unsupervised. *Clin Neuropsychol*. 2015;29:542-58.
- Darby DG, Pietrzak RH, Fredrickson J, Woodward M, Moore L, Fredrickson A, et al. Intraindividual cognitive decline using a brief computerized cognitive screening test. *Alzheimers Dement*. 2012;8:95-104.
- Maruff P, Lim YY, Darby D, Ellis KA, Pietrzak RH, Snyder PJ, et al. Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychology*. 2013;1:30.
- Mielke MM, Machulda MM, Hagen CE, Edwards KK, Roberts RO, Pankratz VS, et al. Performance of the CogState computerized battery in the Mayo Clinic Study on Aging. *Alzheimer's & Dementia*. 2015;11:1367-76.
- Machulda MM, Pankratz VS, Christianson TJ, Ivnik RC, Mielke MM, Roberts RO, et al. Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the Mayo Clinic Study of Aging. *The Clinical Neuropsychologist*. 2013;27:1247-64.
- Mielke MM, Machulda MM, Hagen CE, Christianson TJ, Roberts RO, Knopman DS, et al. Influence of amyloid and APOE on cognitive performance in a late middle-aged cohort. *Alzheimer's & Dementia*. 2016;12:281-91.
- Valdes EG, Sadeq NA, Harrison Bush AL, Morgan D, Andel R. Regular cognitive self-monitoring in community-dwelling older adults using an internet-based tool. *J Clin Exp Neuropsychol*. 2016;38:1026-37.
- Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, et al. The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*. 2008;30:58-69.
- Kokmen E, Smith GE, Petersen RC, Tangalos E, Ivnik RC. The short test of mental status: Correlations with standardized psychometric testing. *Arch Neurol*. 1991;48:725-8.
- Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*. 1993;43:2412-4.
- Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, et al. Mayo's Older Americans Normative Studies: WAIS-R, WMS-R and AVLT norms for ages 56 through 97. *The Clinical Neuropsychologist*. 1992;6:1-104.
- Petersen RC, Roberts RO, Knopman DS, Geda YE, Cha RH, Pankratz VS, et al. Prevalence of mild cognitive impairment is higher in men: The Mayo Clinic Study of Aging. *Neurology*. 2010;75:889-97.
- Stricker NH, Lundt ES, Edwards KK, Machulda MM, Kremers WK, Roberts RO, et al. Comparison of PC and iPad administrations of the Cogstate Brief Battery in the Mayo Clinic Study of Aging: assessing cross-modality equivalence of computerized neuropsychological tests. *Clin Neuropsychol*. 2018;1-25.
- Cogstate. Cogstate Pediatric and Adult Normative Data. New Haven, CT: Cogstate, Inc.; 2018.
- Lewis MS, Maruff P, Silbert BS, Evered LA, Scott DA. The influence of different error estimates in the detection of postoperative cognitive dysfunction using reliable change indices with correction for practice effects. *Arch Clin Neuropsychol*. 2007;22:249-57.
- Calamia M, Markon K, Tranel D. Scoring higher the second time around: Meta-analysis of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*. 2012;26:543-70.
- Heilbronner RL, Sweet JJ, Attix DK, Krull KR, Henry GK, Hart RP. Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test administrations in clinical and forensic contexts. *Clin Neuropsychol*. 2010;24:1267-78.
- Mackin RS, Insel PS, Truran D, Finley S, Flenniken D, Nosheny R, et al. Unsupervised online neuropsychological test performance for individuals with mild cognitive impairment and dementia: Results from the Brain Health Registry. *Alzheimers Dement (Amst)*. 2018;10:573-82.
- Lim YY, Maruff P, Pietrzak RH, Ellis KA, Darby D, Ames D, et al. Abeta and cognitive change: examining the preclinical and prodromal stages of Alzheimer's disease. *Alzheimers Dement*. 2014;10:743-51 e1.
- Louey AG, Cromer JA, Schembri AJ, Darby DG, Maruff P, Makdissi M, et al. Detecting cognitive impairment after concussion: Sensitivity of change from baseline and normative data methods using the CogSport/Axon cognitive test battery. *Arch Clin Neuropsychol*. 2014;29:432-41.
- Hinton-Bayre AD. Deriving reliable change statistics from test-retest normative data: comparison of models and mathematical expressions. *Arch Clin Neuropsychol*. 2010;25:244-56.