HHS Public Access

Author manuscript

Alzheimers Dement. Author manuscript; available in PMC 2022 April 01.

Published in final edited form as:

Alzheimers Dement. 2021 April ; 17(4): 584–594. doi:10.1002/alz.12219.

Diagnostic accuracy of the Cogstate Brief Battery for prevalent MCI and prodromal AD (MCI A+T+) in a population-based sample

Eva C. Alden[1], **Shehroo B. Pudumjee**[1], **Emily S. Lundt**[2], **Sabrina M. Albertson**[2], **Mary M. Machulda**[1], **Walter K. Kremers**[2], **Clifford R. Jack Jr**[3], **David S. Knopman**[4,5], **Ronald C. Petersen**[5], **Michelle M. Mielke**[4,5], **Nikki H. Stricker**[1]

[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]Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA

[4]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

[5]Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA

Abstract

INTRODUCTION: This study evaluated the diagnostic accuracy of the Cogstate Brief Battery (CBB) for mild cognitive impairment (MCI) and prodromal Alzheimer's disease (AD) in a population-based sample.

METHODS: Participants included adults ages 50+ classified as cognitively unimpaired (CU, n=2,866) or MCI (n=226), and a subset with amyloid (A) and tau (T) PET who were AD biomarker negative (A−T−) or had prodromal AD (A+T+).

RESULTS: Diagnostic accuracy of the Learning/Working Memory Composite (Lrn/WM) for discriminating all CU and MCI was moderate (AUC=0.75), but improved when discriminating CU A−T− and MCI A+T+ (AUC=0.93) and when differentiating MCI participants without AD biomarkers from those with prodromal AD (AUC=0.86). Conventional cut-offs yielded lower than expected sensitivity for both MCI (38%) and prodromal AD (73%).

DISCUSSION: Clinical utility of the CBB for detecting MCI in a population-based sample is lower than expected. Caution is needed when using currently available CBB normative data for clinical interpretation.

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.

Conflict of Interest/Disclosure Statement

The authors have no conflicts of interest to report.

Keywords

Sensitivity and Specificity; Neuropsychology; Memory; Mild Cognitive Impairment; Alzheimer's disease; Amyloid; Tau; Cognigram; One Card Learning; One Back

BACKGROUND

There is a growing need for sensitive and reliable cognitive screening instruments for use in primary care settings [1, 2]. Computerized cognitive assessments that can be administered in the clinic or remotely have the potential to increase access to evidence-based screening tools in otherwise underserved populations, facilitate longitudinal monitoring, and allow for early identification of mild cognitive impairment (MCI) [3].

The Cogstate Brief Battery (CBB) is a computerized measure marketed under the name Cognigram™ as a medical device for cognitive screening and monitoring in individuals ages 6-99. It provides objective measures of psychomotor function, attention, working memory, and visual memory [4], is available for prescription use for single measurement or change over time, and can be completed in a clinical setting or at home. One goal of the CBB / Cognigram is to aid in distinguishing MCI or dementia from normal aging. While use of the CBB has gained significant traction in Alzheimer's disease (AD) research and clinical trials [5, 6], more translational studies are needed. One study reported the Learning/Working Memory (Lrn/WM) Composite from the CBB has good diagnostic accuracy for AD dementia and amnestic MCI (aMCI) in settings that resemble a memory disorders clinic [7]. Sabbagh and colleagues [1, 2] recently emphased the importance of validating cognitive screening measures in populations representative of primary care setting. Thus, evaluating the diagnostic accuracy of the CBB for MCI in a population-based sample is important to improve our understanding of its utility for use in primary care settings.

One challenge in investigating the diagnostic accuracy of cognitive measures for MCI is that MCI is a syndrome and not a specific disease entity, with inherent variability in clinical presentation and prognosis [8]. For example, while many individuals with MCI develop dementia, others may remain stable or even revert to cognitively normal [9]. In addition, there is heterogeneous neuropathology underlying MCI [10, 11], and a sizeable proportion of individuals with MCI do not have underlying AD pathology [12, 13]. For instance, Landau, Horng [14] found that 34% of amnestic MCI (aMCI) participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI) were amyloid negative based on florbetapir PET. Further, clinical phenotypes do not necessarily inform the biological diagnosis, as amnestic and other MCI subtypes show proportionally similar underlying AD, vascular, and Lewy body neuropathology [11]. Current research standards propose using the amyloid, tau, and neurodegeneration (or ATN) biomarker framework to support a biological diagnosis of AD [15, 16], but few studies of diagnostic accuracy have included biomarker status among MCI participants [17]. Therefore, including AD biomarkers in studies of the diagnostic accuracy of cognitive measures is an important way to improve the reference standard for diagnosis and provide information about the measures' utility for informing underlying etiology [18].

Our primary study aim was to evaluate the diagnostic accuracy of a single administration of the CBB for differentiating cognitively unimpaired (CU) participants from MCI participants in a population-based sample to help inform whether the CBB may have utility as a screening measure in primary care clinics. Given the known heterogeneity of MCI as reviewed above, secondary aims included assessing a subsample of individuals who also had amyloid and tau PET biomarkers to investigate the diagnostic accuracy of the CBB for differentiating (1) CU participants who were brain amyloid and tau negative (CU A−T−) from those with prodromal AD defined as having a diagnosis of MCI and positive amyloid and tau PET biomarkers (MCI A+T+) [16, 19], and (2) MCI A+T+ from MCI A−T− participants. We hypothesized that the CBB Lrn/WM Composite would demonstrate good diagnostic accuracy for differentiating MCI from CU, and that diagnostic accuracy would increase when limiting the sample to biomarker-refined subgroups of CU A−T− and MCI A+T+. We also hypothesized the CBB Lrn/WM Composite would differentiate MCI A−T− and MCI A+T+, as the latter group would be more likely to show memory impairment. Although study hypotheses focus on the Lrn/WM Composite, we include data for the Attention/Psychomotor Composite given our inclusion of all types of MCI participants, and to better understand the diagnostic accuracy of the full CBB in a population-based sample. Although the Attention/Psychomotor Composite has previously demonstrated minimal utility in MCI, correlations with processing speed measures have been reported [20] and measures comprising this composite have shown group differences in individuals with Lewy Body Dementia relative to healthy older adults [7, 21].

METHODS

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

Participants.

Participants were from the Mayo Clinic Study of Aging (MCSA), a prospective, population-based study of aging. The MCSA recruits residents of Olmsted County, MN, using an age and sex stratified sampling design [22]. Participants without medical contraindication were invited to undergo imaging studies. Each MCSA study visit included the following components: (1) a neurological examination, review of medical history, and administration of the Short Test of Mental Status [23]; (2) interview with a study coordinator to obtain demographic information, medical history, and participant and informant ratings of memory using the Clinical Dementia Rating® scale [24]; and (3) neuropsychological testing. A more comprehensive account of these components, including information about the neuropsychological battery, is available [22]. Using previously published criteria [8, 25] to guide consensus agreement between the physician, study coordinator, and neuropsychologist, participants are assigned a diagnosis of CU, MCI, or dementia. Participants were classified as having amnestic MCI if there was evidence of memory impairment regardless of whether there was impairment in other areas (i.e. single and multidomain). All other participants diagnosed with MCI who did not have memory impairment were termed non-amnestic MCI. Diagnostic decisions at each study visit were

made blind to prior clinical information, prior diagnoses, and biomarker information. Cogstate performance was not considered for diagnosis.

Inclusion Criteria.

The present study included individuals 50 or older who were classified as either CU or MCI at the time of their baseline Cogstate session. A subset of individuals had amyloid PET (Pittsburgh Compound B) and tau PET (AV1451) scans within two years of their baseline Cogstate evaluation and were included in biomarker subgroups.

Cogstate Brief Battery.

The CBB includes 4 individual card tasks that measure psychomotor function (Detection), attention (Identification), working memory (One Back), and visual recognition memory (One Card Learning). To normalize the variables, accuracy data from One Card Learning (OCL) and One Back were transformed using arcsine transformation; reaction time data from Detection and Identification were transformed using logarithmic base 10 transformation [26–29]. The Lrn/WM Composite is an average of OCL accuracy and One Back accuracy subtests. The Attention/Psychomotor Composite is an average of the Detection and Identification subtests. We applied age-corrected normative data provided by Cogstate by calculating z-scores based on the means and SDs by available 10-year age bands, and averaged those age-corrected z-scores for the composites [30]. We also present data by subtest to aid in understanding how subtest performance influences composite score results.

The current study only included baseline CBB performance, which was typically completed either on a PC or an iPad in clinic; a few participants completed their baseline CBB on a PC at home (see Table 4). Data values with a failed completion flag were identified as those completing fewer than 75% of trials within each task and were not considered for further analyses. Cogstate version 7 was used. See our prior work for a more detailed description of the use of Cogstate in the MCSA [31, 32]. Note that the Lrn/WM data presented for the CU A−T− subgroup was also reported in a prior publication focused on comparisons with CU A+T− and CU A+T+ subgroups [33].

Biomarker Acquisition.

PiB-PET and tau-PET are obtained using PET/CT. Using thresholds of SUVR≥1.48 (centiloid 22; [34]) and SUVR≥1.25, individuals were considered positive for amyloid and tau, respectively [35–37]. Biomarker subgroups were based on the recently proposed research framework for a biological diagnosis of Alzheimer's disease [16] and included CU participants with negative amyloid and tau PET biomarkers (CU A−T−), participants with MCI and positive amyloid and tau PET biomarkers (MCI A+T+), and participants with MCI who had negative amyloid and tau PET biomarkers (MCI A−T−). Although the proposed research framework includes neurodegeneration (N*) as a biomarker, it is not specific to AD and therefore is used only for staging, not for diagnostic purposes.

Analysis.

Linear model ANOVA test for means and Chi-squared test for frequencies were used to illustrate demographic and clinical differences across CU, MCI and biomarker subgroups. Effect sizes were computed using weighted and pooled standard deviations (Hedge's *g*). AUROC analyses were conducted to assess the ability of each CBB measure to discriminate between CU and MCI groups, as well as between groups based on amyloid and tau positivity. Conventional cut-offs are based on existing, generalizable clinical interpretive standards and facilitate clinical translation of study results. Therefore, results focus on application of a conventional ≤ −1 standard deviation (SD) below the mean cut-off based upon age-corrected normative scores (equivalent to age-corrected z-score ≤ −1), consistent with Maruff, Lim [7] and cut-offs used in Cognigram™. Study-specific, data-driven "optimal" cut-offs generated from the AUROC analysis are also reported and provide the best balance between sensitivity and specificity for the specific samples used in these analyses. Youden's J statistic (J = max{sensitivities + specificities}) is employed and the optimal cut-off is the threshold that achieves the maximum distance to the identity (diagonal) line [38]. Optimal cut-offs are susceptible to bias, particularly in small sample sizes (n < 40), and therefore may be limited in terms of generalizability to other samples [39]. The current study is a retrospective analysis and followed reporting standards for studies of diagnostic test accuracy in dementia [18]. We also evaluated frequency of low test performance. For CU individuals, approximately 16% are expected to show low test performance based on typical normative standards that assume a normal distribution of performance.

RESULTS

CU vs. all MCI.

Group Comparisons.—Performance on both the Lrn/WM Composite (Hedge's $g = 0.97$) and the Attention/Psychomotor Composite (Hedge's $g = 0.88$) was lower in the MCI group relative to the CU group (see Table 1). In addition, the MCI group showed a higher frequency of low test performance (≤ −1SD) for both Composites.

Diagnostic accuracy.—Overall diagnostic accuracy of the Lrn/WM Composite for differentiating CU and MCI participants was moderate (see Table 2). The conventional cut-off of z ≤ −1 resulted in poor sensitivity with only 38% of MCI subjects performing below this cut-off, though specificity was excellent (91%). The derived optimal cut-off of z = −0.21 yielded moderate sensitivity and specificity (both 70%). Overall diagnostic accuracy of the Attention/Psychomotor Composite was moderate for differentiating CU and MCI (see Table 3). The following sensitivity analyses show that these diagnostic accuracy results are not significantly impacted by controlling for covariates or when examining MCI subtypes separately.

Impact of Covariates.—Given our focus on understanding how currently available age-adjusted normative data perform in this validation study, we do not adjust for covariates for our primary analyses. However, participants with MCI were older and had 1 less year of education on average relative to the CU groups. To ensure these demographic differences

were not driving results for primary analyses, we computed a covariate-adjusted AUROC analyses controlling for age, sex, education, and device type (PC/iPad). The pattern of results remained the same (see Supplementary Table 1).

Exploring results by MCI subtype.—Supplementary analyses investigated whether results varied by MCI subtypes. Mean comparisons across aMCI (n=174) and non-amnestic MCI (naMCI; n=52) subtypes support collapsing these subgroups into all MCI participants for our primary analyses (see Supplemental Table 2). Performances were comparable across subtypes for OCL accuracy, Detection RT, and Identification RT (all p's > 0.05; see Supplemental Table 3). The naMCI groups showed lower performance on One Back accuracy ($p < .05$) than the aMCI group. Similar to the all MCI results, the aMCI group and the naMCI group showed lower mean performances across Lrn/WM Composite, Attention/Psychomotor Composite, and all CBB subtests relative to the CU group (all p's < .001). Diagnostic accuracy results for differentiating CU and aMCI participants were very similar to results when differentiating CU from all MCI (see Supplementary Table 4). Use of a conventional cut-off of z ≤ −1 resulted in subtly lower sensitivity for aMCI relative to all MCI (35% vs. 38%) for the Lrn/WM Composite. Total AUC was slightly higher for differentiating CU and naMCI participants relative to when differentiating CU and all MCI participants for both Lrn/WM (total AUC = 0.81 vs. 0.75) and Attention/Psychomotor (total AUC = 0.73 vs. 0.70) Composites.

CU A−T− vs. MCI A+T+.

Group Comparisons.—The MCI A+T+ group had significantly lower performance on the Lrn/WM Composite (Hedge's $g = 2.12$), and a higher frequency of low performance relative to the CU A−T− group (see Table 3). The MCI A+T+ group showed a trend toward lower performance relative to the CU A−T− group on the Attention/Psychomotor Composite (Hedge's $g = 0.48$); the frequency of low performance was higher than expected in the CU A−T− group (37.0%) based on typical normative expectations and did not differ from the MCI A+T+ group.

Diagnostic accuracy.—Overall diagnostic accuracy for differentiating CU A−T− from MCI A+T+ was excellent. A conventional cut-off of z ≤ −1 yields moderate sensitivity (73%) and excellent specificity (95%) for the Lrn/WM Composite. The derived optimal cut-off for the Lrn/WM Composite is well within normal limits at z= −0.32, which would be challenging to apply clinically. The Attention/Psychomotor Composite did not differentiate groups better than chance.

MCI A−T− vs. MCI A+T+.

Group Comparisons.—The MCI A+T+ group had significantly lower performance relative to the MCI A−T− group on the Lrn/WM Composite (Hedge's g= 1.43), and higher frequency of low performance. The MCI A+T+ and MCI A−T− groups showed comparable performance on the Attention/Psychomotor Composite (Hedge's $g = 0.22$).

Diagnostic accuracy.—Overall diagnostic accuracy for differentiating MCI A+T+ from MCI A−T− participants was good for the Lrn/WM Composite (AUC = 0.86). A

conventional cut-off yields adequate sensitivity (73%) and good specificity (86%) for differentiating MCI A−T− and MCI A+T+; an optimal cut-off of −.79 improves sensitivity slightly to 80% while maintaining equivalent specificity. The diagnostic accuracy of the Attention/Psychomotor Composite was not better than chance for differentiating MCI A−T− and MCI A+T+.

Subtest Level Results.—Tables 4 and 5 display subtest level data for all four CBB subtests. CU participants, including CU A−T− participants, demonstrated a high frequency of below cut-off performance on both subtests comprising the Attention/Psychomotor Composite. For subtests comprising the Learning/Working Memory composite, frequency of low performance in CU participants was in line with typical normative expectations for the One Back subtest, and lower than expected for the One Card Learning subtest. The MCI A+T+ group had lower mean performance on OCL and One Back subtests compared to the MCI A−T− group. The MCI A+T+ and MCI A−T− groups did not differ on the Detection or Identification subtests.

DISCUSSION

The present study evaluated the diagnostic accuracy of the CBB for detecting MCI in a population-based sample, which may approximate expected test performance for patients seen in primary care clinics. Findings suggest the diagnostic accuracy of the Lrn/WM Composite for differentiating CU from MCI participants was moderate overall, but showed unexpectedly low sensitivity (38%) to all MCI with application of a conventional cut-off of ≤ −1 SD. With regard to diagnostic accuracy among biomarker-refined subgroups, the Lrn/WM Composite was better at differentiating biomarker negative CU (CU A−T−) from biomarker positive MCI (MCI A+T+) participants relative to all CU and MCI comparisons. In addition, the Lrn/WM Composite shows some promise for differentiating MCI A+T+ from MCI A−T−, with 86% total AUC. In contrast, the Attention/Psychomotor speed Composite did not differentiate among biomarker refined subgroups, and may not be a useful indicator for differential diagnosis.

There are no clear criteria regarding minimum standards of sensitivity and specificity for the diagnostic accuracy of MCI for cognitive measures [17]. Sabbagh and colleagues describe that an ideal detection tool for MCI in primary care should have high sensitivity to ensure that individuals in need of follow-up care will not be missed [1]. In contrast, moderate specificity may be acceptable for a screening measure. When applying a conventional cut-off, we see the opposite pattern as that desired for primary care for the CBB Lrn/WM Composite, with unexpectedly low sensitivity levels and high specificity. The relatively limited clinical utility of the Lrn/WM Composite for differentiating all CU and MCI participants was inconsistent with study hypotheses, given a previous study reported the Lrn/WM Composite accurately discriminated aMCI from normal aging with 80.4% sensitivity and 84.7% specificity [7]. Supplemental analyses show that results did not improve when our sample was restricted to those with aMCI. One reason for the markedly lower sensitivity observed in our study may be differences in sample characteristics, as most computerized assessment measures have not been validated in population-based cohorts [1, 40]. While the previous study was conducted in a setting similar to a memory clinic,

our sample was derived from a population-based study with broad inclusion criteria. Other studies of computerized testing in primary care settings have reported similar findings, namely that tests may be less sensitive and less reliable for detecting dementia in a primary care setting [41].

The discrepancy between the relatively limited clinical utility of the CBB for detecting all MCI versus the high diagnostic accuracy observed in the biomarker refined samples illustrates the importance of understanding test performance in population-based samples [1, 40]. Results that refine subjects to CU A−T− and MCI A+T+ groups are more similar to prior findings in study samples with careful exclusionary criteria and a high likelihood of MCI due to Alzheimer's disease [7]. However, the 73% sensitivity of the Lrn/WM Composite for prodromal AD (MCI A+T+) remains lower than anticipated given these individuals are on the cusp of transitioning to AD dementia, when memory measures typically show very high sensitivity [17]. Consistent with prior findings [21], our results suggest the measures comprising the Lrn/WM Composite are not sufficiently sensitive to accurately identify early memory impairment due to AD pathology until individuals meet criteria for dementia. For example, a limited proportion of MCI A+T+ participants showed low performance on the OCL subtest (40%).

A secondary but important finding is that these results raise questions about the CBB's normative data and underlying psychometric properties. Although the CBB does show some ability to differentiate CU and MCI groups based on total AUC values, applying internal Cogstate norms yields unexpected results based on typical expectations for normative data characteristics and may significantly limit the CBB's clinical utility. Cogstate norms were derived from individuals from numerous countries enrolled in clinical trials, research, and academic studies [30]. Cogstate normative data do not take into account device type or location where the test was completed, and the importance of considering device type is increasingly recognized [42]. Consistent with our prior results [31, 32], an illustration of device type impact is seen on the Attention/Psychomotor Composite, where we observed a *higher* frequency of low test performance (37.0%) relative to typical normative expectations among CU A−T− participants. Use of CBB in primary care with the current norms has a high risk of interpretation errors, particularly when done by providers without expertise in psychometrics and interpretation of normative data.

The present study has some limitations. First, sample sizes for biomarker-refined MCI subgroups were relatively small, and replication in a larger sample is needed. In addition, the present study only evaluated CBB diagnostic accuracy for a single time point, and evaluation of longitudinal change across repeat administrations is needed. Finally, although the current results have direct relevance for the use of Cognigram™, there are subtle differences in test instructions and aspects of the normative data across the Cogstate Brief Battery version used in this research study and Cognigram™.

In summary, findings from this study suggest a single baseline administration of the CBB has modest clinical utility for identifying MCI in a population-based sample. Given the low sensitivity of the CBB to all MCI, a high rate of false negative results is expected in primary care clinics, which will delay referrals for further work-up and opportunities for

early intervention. However, our results may support targeted use of the CBB in memory clinic settings given demonstrated utility for differentiating biomarker status in those already diagnosed with MCI. The differing results depending on how samples are defined highlight the importance of varying sample characteristics and the reference standard used in studies of diagnostic accuracy [18]. Reducing syndromal heterogenerity by refining samples by biomarker characteristics may offer new insights for test validation studies. Finally, overall clinical utility of the CBB could be improved if updated normative data become available or with further refinement of the test battery, including addition of more sensitive measures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors wish to thank the participants and staff at the Mayo Clinic Study of Aging. This work was supported by the Rochester Epidemiology Project (R01 AG034676), the National Institutes of Health (grant numbers P50 AG016574, P30 AG062677, U01 AG006786, R37 AG011378, R01 AG041851, RF1 AG55151), a grant from the Alzheimer's Association (AARG-17-531322), Zenith Award from the Alzheimer's Association, the Robert Wood Johnson Foundation, The Elsie and Marvin Dekelboum Family Foundation, Alexander Family Alzheimer's Disease Research Professorship of the Mayo Clinic, Liston Award, Schuler Foundation, GHR Foundation, AVID Radiopharmaceuticals, and the Mayo Foundation for Education and Research. We would like to greatly thank AVID Radiopharmaceuticals, Inc., for their support in supplying AV-1451 precursor, chemistry production advice and oversight, and FDA regulatory cross-filing permission and documentation needed for this work. 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.

References

1. Sabbagh M, et al., Early Detection of Mild Cognitive Impairment MCI in Primary Care. J Prev Alzheimers Dis, 2020.
2. Petersen RC, and Yaffe K, Issues and Questions Surrounding Screening for Cognitive Impairment in Older Patients. JAMA, 2020. 323.
3. Sabbagh M, et al., Early Detection of Mild Cognitive Impairment MCI in an At Home Setting. J Prev Alzheimers Dis, 2020.
4. Maruff P, 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): p. 165–78. [PubMed: 19395350]
5. Weiner MW, et al., The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimer's & Dementia, 2017. 13(5): p. 561–571.
6. Mackin RS, 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: p. 573–582. [PubMed: 30406176]
7. Maruff P, 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(1): p. 30–30. [PubMed: 25566378]
8. Petersen RC, Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine, 2004. 256(3): p. 183–194. [PubMed: 15324362]
9. Roberts RO, et al., Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. Neurology, 2014. 82(4): p. 317–325. [PubMed: 24353333]
10. Kapasi A, DeCarli C, and Schneider JA, Impact of multiple pathologies on the threshold for clinically overt dementia. Acta Neuropathologica, 2017. 134(2): p. 171–186. [PubMed: 28488154]

11. Schneider JA, et al., The neuropathology of probable Alzheimer disease and mild cognitive impairment. Ann Neurol, 2009. 66(2): p. 200–8. [PubMed: 19743450]
12. Jack CR Jr., et al., Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework. JAMA Neurol, 2019.
13. Petersen RC, et al., Mild cognitive impairment due to Alzheimer disease in the community. Ann Neurol, 2013. 74(2): p. 199–208. [PubMed: 23686697]
14. Landau S, et al., Amyloid negativity in patients with clinically diagnosed AD and MCI. Neurology, 2016. 86.
15. Jack CR Jr., et al., A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology, 2016. 87(5): p. 539–47. [PubMed: 27371494]
16. Jack CR Jr., et al., NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement, 2018. 14(4): p. 535–562. [PubMed: 29653606]
17. Weissberger GH, et al., Diagnostic Accuracy of Memory Measures in Alzheimer's Dementia and Mild Cognitive Impairment: a Systematic Review and Meta-Analysis. Neuropsychol Rev, 2017. 27(4): p. 354–388. [PubMed: 28940127]
18. Noel-Storr AH, et al., Reporting standards for studies of diagnostic test accuracy in dementia: The STARDdem Initiative. Neurology, 2014. 83(4): p. 364–73. [PubMed: 24944261]
19. Albert MS, et al., The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, 2011. 7(3): p. 270–9. [PubMed: 21514249]
20. Collie A, et al., CogSport: Reliability and Correlation with Conventional Cognitive Tests Used in Postconcussion Medical Evaluations. Clinical Journal of Sport Medicine, 2003. 13: p. 28–32. [PubMed: 12544161]
21. Hammers D, et al., Validity of a brief computerized cognitive screening test in dementia. J Geriatr Psychiatry Neurol, 2012. 25(2): p. 89–99. [PubMed: 22689701]
22. Roberts RO, et al., The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology, 2008. 30(1): p. 58–69. [PubMed: 18259084]
23. Kokmen E, et al., The short test of mental status: Correlations with standardized psychometric testing. Archives of Neurology, 1991. 48(7): p. 725–8. [PubMed: 1859300]
24. Morris JC, The Clinical Dementia Rating (CDR): Current version and scoring rules. Neurology, 1993. 43(11): p. 2412–2414.
25. American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders. 4th ed.2000, Washington, D.C.
26. Collie A, et al., The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. Journal of the International Neuropsychological Society, 2003. 9(3): p. 419–428. [PubMed: 12666766]
27. Fredrickson J, et al., Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. Neuroepidemiology, 2010. 34(2): p. 65–75. [PubMed: 20016215]
28. Lim Y, et al., Use of the CogState Brief Battery in the assessment of Alzheimer's disease related cognitive impairment in the Australian Imaging, Biomarkers and Lifestyle (AIBL) study. Journal of Clinical and Experimental Neuropsychology, 2012. 34(4): p. 345–358. [PubMed: 22248010]
29. Pietrzak RH, et al., An examination of the construct validity and factor structure of the Groton Maze Learning Test, a new measure of spatial working memory, learning efficiency, and error monitoring. Archives of Clinical Neuropsychology, 2008. 23(4): p. 433–445. [PubMed: 18448309]
30. Cogstate, Cogstate Pediatric and Adult Normative Data. 2018, New Haven, CT: Cogstate, Inc.
31. Stricker NH, 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. The Clinical neuropsychologist, 2019. 33(6): p. 1102–1126. [PubMed: 30417735]

32. Stricker NH, et al., 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. The journal of prevention of Alzheimer's disease, 2020. 7(1): p. 21–28.

33. Stricker NH, et al., Diagnostic and Prognostic Accuracy of the Cogstate Brief Battery and Auditory Verbal Learning Test in Preclinical Alzheimer's Disease and Incident Mild Cognitive Impairment: Implications for Defining Subtle Objective Cognitive Impairment Journal of Alzheimer's Disease, JAD, in press.

34. Klunk WE, et al., The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement, 2015. 11(1): p. 1–15e1–4. [PubMed: 25443857]

35. Jack CR Jr., et al., 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain, 2008. 131(Pt 3): p. 665–80. [PubMed: 18263627]

36. Jack CR Jr., et al., Defining imaging biomarker cut points for brain aging and Alzheimer's disease. Alzheimers Dement, 2017. 13(3): p. 205–216. [PubMed: 27697430]

37. Vemuri P, et al., Tau-PET uptake: Regional variation in average SUVR and impact of amyloid deposition. Alzheimers Dement (Amst), 2017. 6: p. 21–30. [PubMed: 28138510]

38. Perkins NJ, and Schisterman EF, The Youden Index and the optimal cut-point corrected for measurement error. Biom J, 2005. 47(4): p. 428–41. [PubMed: 16161802]

39. Leeflang MM, et al., Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem, 2008. 54(4): p. 729–37. [PubMed: 18258670]

40. De Roeck EE, et al., Brief cognitive screening instruments for early detection of Alzheimer's disease: a systematic review. Alzheimers Res Ther, 2019. 11(1): p. 21. [PubMed: 30819244]

41. Seitz DP, et al., Mini-Cog for the diagnosis of Alzheimer's disease dementia and other dementias within a primary care setting. Cochrane Database Syst Rev, 2018. 2: p. CD011415. [PubMed: 29470861]

42. Germine L, Reinecke K, and Chaytor NS, Digital neuropsychology: Challenges and opportunities at the intersection of science and software. The Clinical Neuropsychologist, 2019. 33(2): p. 271–286. [PubMed: 30614374]

Table 1.

Demographic characteristics, mean performance, and frequency of low test performance across CU and MCI groups.

	CU (N = 2,866)	MCI (N = 226)	*p*-value
Age Mean (SD)	69.82 (11.31)	75.96 (10.28)	< 0.001[a]
Education Mean (SD)	14.93 (2.50)	13.62 (2.70)	< 0.001[a]
Sex (% Male)	1433 (50.0%)	119 (52.7%)	0.442[b]
Short Test of Mental Status [c]	35.70 (2.05)	31.01 (2.76)	< 0.001[a]
CBB Lrn/WM Composite Mean (z)	0.12 (0.89)	−0.77 (1.09)	< 0.001[a]
CBB Lrn/WM Composite			< 0.001[b]
Normal z > −1	2599 (91.0%)	138 (62.2%)	
z ≤ −1	257 (9.0%)	84 (37.8%)	
CBB Attn/Psychomotor Composite Mean (z)	−0.53 (0.90)	−1.34 (1.20)	< 0.001[a]
CBB Attn/Psychomotor Composite			< 0.001[b]
Normal z > −1	2089 (73.1%)	105 (46.9%)	
z ≤ −1	768 (26.9%)	119 (53.1%)	

[a] p-values represent linear model ANOVAs for mean comparisons.

[b] p-values represent Pearson's Chi-squared test for frequency comparisons.

[c] Performance is considered as part of consensus diagnosis.

Note. CU = cognitively unimpaired; MCI = mild cognitive impairment (n=174 amnestic MCI; n=52 non-amnestic MCI); CBB = Cogstate Brief Battery; Lrn/WM = Learning / Working Memory; Attn = Attention.

Table 2.

Diagnostic accuracy.

Description	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
CBB Lrn/WM Composite	Optimal (*z*)			
CU vs MCI all subjects	≤ −0.21	0.70 (0.64,0.76)	0.70 (0.68,0.72)	0.75 (0.71,0.78)
CU A−T− vs MCI A+T+	≤ −0.32	0.93 (0.80,1.00)	0.79 (0.72,0.86)	0.93 (0.87,0.99)
MCI A−T− vs MCI A+T+	≤ −0.79	0.80 (0.60,1.00)	0.86 (0.64,1.00)	0.86 (0.73,1.00)
CBB Lrn/WM Composite	Conventional (*z*)[1]			
CU vs MCI all subjects	≤ −1	0.38 (0.32,0.45)	0.91 (0.90,0.92)	
CU A−T− vs MCI A+T+	≤ −1	0.73 (0.47,0.93)	0.95 (0.91,0.99)	
MCI A−T− vs MCI A+T+	≤ −1	0.73 (0.47,0.93)	0.86 (0.64,1.00)	
CBB Attention/Psychomotor Composite	Optimal (*z*)			
CU vs MCI all subjects	≤ −0.73	0.69 (0.63, 0.75)	0.64 (0.62, 0.65)	0.70 (0.66, 0.74)
CU A−T− vs MCI A+T+	≤ −1.32	0.60 (0.33, 0.80)	0.73 (0.66, 0.80)	0.64 (0.49, 0.80)
MCI A−T− vs MCI A+T+	≤ −1.10	0.60 (0.33, 0.87)	0.50 (0.21, 0.79)	0.43 (0.21, 0.65)
CBB Attention/Psychomotor Composite	Conventional (*z*)[1]			
CU vs MCI all subjects	≤ −1	0.53 (0.47, 0.60)	0.73 (0.71, 0.75)	
CU A−T− vs MCI A+T+	≤ −1	0.60 (0.33, 0.87)	0.63 (0.55, 0.70)	
MCI A−T− vs MCI A+T+	≤ −1	0.60 (0.33, 0.87)	0.50 (0.21, 0.79)	

Note. CBB = Cogstate Brief Battery. Lrn/WM = Learning/Working Memory. CU= Cognitively Unimpaired. MCI= Mild Cognitive Impairment. A = amyloid. T = tau. Cogstate Brief Battery was independent of diagnosis. Biomarker status was not considered for diagnosis.

[1] AUC values are the same regardless of cut-off applied thus are not repeated.

Table 3.

Demographic characteristics, mean performance, and frequency of low test performance across CU and MCI biomarker subgroups.

	CU A−T− (*n* = 146)	MCI A+T+ (*n* = 15)	*p*- value for CU A−T− and MCI A+T+	MCI A−T− (*n* = 14)	*p*- value for MCI A−T− and MCI A+T+
Age Mean (SD)	66.32 (12.06)	82.58 (4.28)	<0.001[a]	72.50 (12.55)	0.007[a]
Education Mean (SD)	15.24 (2.44)	14.33 (2.50)	0.173[a]	13.36 (2.37)	0.291[a]
Sex (% Male)	78 (53.4%)	9 (60.0%)	0.627[b]	8 (57.1%)	0.876[b]
Short Test of Mental Status [c]	36.18 (1.99)	32.00 (2.36)	<0.001[a]	32.14 (3.21)	0.892[a]
CBB Lrn/WM Composite Mean (z)	0.21 (0.78)	−1.49 (0.99)	<0.001[a]	−0.19 (0.75)	<0.001[a]
CBB Lrn/WM Composite			<0.001[b]		0.001[b]
Normal z > −1	138 (95.2%)	4 (26.7%)		12 (85.7%)	
z ≤ −1	7 (4.8%)	11 (73.3%)		2 (14.3%)	
CBB Attn/Psychomotor Mean (*z* score)	−0.71 (1.13)	−1.26 (1.14)	0.077[a]	−1.54 (1.35)	0.548[a]
CBB Attn/Psychomotor Composite			0.082[b]		0.588[b]
Normal z > −1	92 (63.0%)	6 (40.0%)		7 (50%)	
z ≤ −1	54 (37.0%)	9 (60.0%)		7 (50%)	

[a] p-values represent linear model ANOVAs for mean comparisons.

[b] p-values represent Pearson's Chi-squared test for frequency comparisons.

[c] Performance is considered as part of consensus diagnosis.

Note. CU = cognitively unimpaired; A = amyloid; T = tau; MCI = mild cognitive impairment; CBB = Cogstate Brief Battery; Lrn/WM = Learning / Working Memory; Attn = Attention.

Table 4.

Subtest performance and frequency of low performance across CU and MCI subgroups.

	CU (N = 2,866)	MCI (N = 226)	*p*-value
OCL accuracy Mean (SD) (*Transf))*	0.978 (0.106)	0.880 (0.111)	< 0.001[a]
OCL accuracy Mean (SD) *(Untransf, %)*	68.4 (9.7)	59.2 (10.6)	< 0.001[a]
OCL accuracy (z score) Mean (SD)	0.229 (0.827)	−0.516 (0.858)	< 0.001[a]
OCL accuracy			< 0.001[b]
Normal z > −1	2670 (93.2%)	169 (74.8%)	
z ≤ −1	196 (6.8%)	57 (25.2%)	
ONB accuracy Mean (SD) (*Transf)*	1.348 (0.190)	1.196 (0.251)	< 0.001[a]
ONB accuracy Mean (SD) *(Untransf, %)*	92.4 (10.8)	82.8 (17.4)	< 0.001[a]
ONB accuracy (z score) Mean (SD)	−0.001 (1.356)	−1.040 (1.794)	< 0.001[a]
ONB accuracy			< 0.001[b]
Normal z > −1	2356 (82.5%)	117 (52.7%)	
z ≤ −1	500 (17.5%)	105 (47.3%)	
DET Mean (SD) (*Transf)*	2.626 (0.116)	2.728 (0.157)	< 0.001[a]
DET Mean (SD) *(Untransf, ms)*	439.29 (140.87)	574.90 (250.13)	< 0.001[a]
DET (z score) Mean (SD)	−0.610 (1.061)	−1.454 (1.469)	< 0.001[a]
DET			< 0.001[b]
Normal z > −1	1959 (68.5%)	99 (44.2%)	
z ≤ −1	899 (31.5%)	125 (55.8%)	
IDN Mean (SD) (*Transf)*	2.769 (0.086)	2.847 (0.110)	< 0.001[a]
IDN Mean (SD) *(Untransf, ms)*	599.19 (129.99)	727.39 (201.93)	< 0.001[a]
IDN (z score) Mean (SD)	−0.454 (0.994)	−1.244 (1.311)	< 0.001[a]
IDN			< 0.001[b]
Normal z > −1	2145 (74.9%)	110 (48.9%)	
z ≤ −1	719 (25.1%)	115 (51.1%)	
Platform/Location			< 0.001[b]
Home (PC)	14 (0.5%)	1 (0.4%)	
PC Clinic	1799 (62.8%)	97 (42.9%)	
iPad Clinic	1053 (36.7%)	128 (56.6%)	

[a] p-values represent linear model ANOVAs for mean comparisons.

[b] p-values represent Pearson's Chi-squared test for frequency comparisons.

Note. CU = cognitively unimpaired; MCI = mild cognitive impairment; OCL = One Card Learning; ONB = One Back; DET = Detection; IDN = Identification; transf = transformed; untransf = untransformed (raw value), % = percentage correct (ONB, OCL), ms = milliseconds.

Table 5.

Subtest performance and frequency of low performance across CU and MCI biomarker subgroups.

	CU A−T− (*n* = 146)	MCI A+T+ (*n* = 15)	*p*-value CU A−T− and MCI A+T+	MCI A−T− (*n* = 14)	*p*-value MCI A−T− and MCI A+T+
OCL accuracy Mean (SD) (*Transf))*	0.987 (0.101)	0.848 (0.082)	< 0.001[a]	0.925 (0.081)	0.017[a]
OCL accuracy Mean (SD) *(Untransf, %)*	69.3 (9.3)	56.2 (8.0)	< 0.001[a]	63.6 (7.7)	0.017[a]
OCL accuracy (z score) Mean (SD)	0.285 (0.795)	−0.670 (0.629)	< 0.001[a]	−0.173 (0.634)	0.043[a]
OCL accuracy			< 0.001[b]		0.122[b]
Normal z > −1	138 (94.5%)	9 (60.0%)		12 (85.7%)	
z ≤ −1	8 (5.5%)	6 (40.0%)		2 (14.3%)	
ONB accuracy Mean (SD) (*Transf)*	1.369 (0.166)	0.996 (0.237)	< 0.001[a]	1.310 (0.185)	<0.001[a]
ONB accuracy Mean (SD) *(Untransf, %)*	93.6 (8.2)	68.7 (21.2)	< 0.001[a]	90.6 (9.4)	0.001[a]
ONB accuracy (z score) Mean (SD)	0.134 (1.214)	−2.314 (1.626)	< 0.001[a]	−0.216 (1.284)	<0.001[a]
ONB accuracy			< 0.001[b]		0.016[b]
Normal z > −1	125 (86.2%)	4 (26.7%)		10 (71.4%)	
z ≤ −1	20 (13.8%)	11 (73.3%)		4 (28.6%)	
DET Mean (SD) (*Transf)*	2.618 (0.135)	2.737 (0.145)	0.002[a]	2.722 (0.147)	0.797[a]
DET Mean (SD) *(Untrans; ms)*	437.55 (162.75)	575.326 (204.507)	0.003[a]	558.19 (208.37)	0.825[a]
DET (z score) Mean (SD)	−0.631 (1.233)	−1.383 (1.370)	0.027[a]	−1.516 (1.394)	0.798[a]
DET			0.056[b]		0.87b[a]
Normal z > −1	95 (65.1%)	6 (40.0%)		6 (42.9%)	
z ≤ −1	51 (34.9%)	9 (60.0%)		8 (57.1%)	
IDN Mean (SD) (*Transf)*	2.787 (0.105)	2.855 (0.099)	0.017[a]	2.860 (0.145)	0.919[a]
IDN Mean (SD) *(Untransf, ms)*	630.81 (164.06)	734.555 (176.678)	0.022[a]	764.40 (276.79)	0.730[a]
IDN (z score) Mean (SD)	−0.793 (1.226)	−1.129 (1.180)	0.313[a]	−1.560 (1.850)	0.458[a]
IDN			0.405[b]		0.096[b]
Normal z > −1	81 (55.5%)	10 (66.7%)		5 (35.7%)	
z ≤ −1	65 (44.5%)	5 (33.3%)		9 (64.3%)	
Platform/Location			0.071[b]		0.292[b]
Home (PC)	1 (0.7%)	0		0	
PC Clinic	38 (26.0%)	0		1 (7.1%)	
iPad Clinic	107 (73.3%)	15 (100%)		13 (92.9%)	

[a] p-values represent linear model ANOVAs for mean comparisons.

[b] p-values represent Pearson's Chi-squared test for frequency comparisons.

Note. CU = cognitively unimpaired; A = amyloid; T = tau; MCI = mild cognitive impairment; OCL = One Card Learning; ONB = One Back; DET = Detection; IDN = Identification; transf = transformed; untransf = untransformed (raw value), % = percentage correct (ONB, OCL), ms = milliseconds.