Received: 5 October 2020 | Accepted: 5 October 2020 | Published online: 19 November 2020

DOI: 10.1002/trc2.12103

Alzheimer's & Dementia
Translational Research & Clinical Interventions

RESEARCH ARTICLE

Differentiating among stages of cognitive impairment in aging: Version 3 of the Uniform Data Set (UDS) neuropsychological test battery and MoCA index scores

Hiroko H. Dodge[1] | **Felicia C. Goldstein**[2] | **Nicole I Wakim**[3] | **Tamar Gefen**[4] | **Merilee Teylan**[5] | **Kwun C.G. Chan**[5] | **Walter A. Kukull**[5,6] | **Lisa L. Barnes**[7] | **Bruno Giordani**[8] | **Timothy M. Hughes**[9] | **Joel H. Kramer**[10] | **David A. Loewenstein**[11] | **Daniel C. Marson**[12] | **Dan M. Mungas**[13] | **Nora Mattek**[1] | **Bonnie C. Sachs**[9] | **David P. Salmon**[14] | **Monica Willis-Parker**[2] | **Kathleen A. Welsh-Bohmer**[15] | **Katherine V. Wild**[1] | **John C. Morris**[16] | **Sandra Weintraub**[4] | **National Alzheimer's Coordinating Center (NACC)**

[1] Layton Aging and Alzheimer's Disease Center, Oregon Health & Science University, Portland, Oregon, USA

[2] Goizueta Alzheimer's Disease Research Center, Emory University School of Medicine, Atlanta, Georgia, USA

[3] Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA

[4] The Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

[5] National Alzheimer's Coordinating Center, University of Washington, Seattle, Washington, USA

[6] Department of Epidemiology, University of Washington, Seattle, Washington, USA

[7] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA

[8] Michigan Alzheimer's Disease Center, University of Michigan, Ann Arbor, Michigan, USA

[9] The Wake Forest Alzheimer's Disease Research Center, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA

[10] Memory and Aging Center, University of California San Francisco, San Francisco, California, USA

[11] Center for Cognitive Neuroscience and Aging, University of Miami Miller School of Medicine, Miami, Florida, USA

[12] Alzheimer's Disease Center, University of Alabama at Birmingham, Birmingham, Alabama, USA

[13] The UC Davis Alzheimer's Disease Research Center, University of California, Davis, Sacramento, California, USA

[14] The Shiley-Marcos Alzheimer's Disease Research Center, University of California, San Diego, La Jolla, California, USA

[15] Alzheimer's Disease Research Center, Duke University, Durham, North Carolina, USA

[16] Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, Missouri, USA

Correspondence
Hiroko H. Dodge, Department of Neurology, Oregon Health & Science University, CR 131, 3181 SW Sam Jackson Park Road, Portland, OR 97239.
Email: dodgeh@ohsu.edu

Abstract

Introduction: Federally funded Alzheimer's Disease Centers in the United States have been using a standardized neuropsychological test battery as part of the National Alzheimer's Coordinating Center Uniform Data Set (UDS) since 2005. Version 3 (V3) of the UDS replaced the previous version (V2) in 2015. We compared V2 and V3 neu-

ropsychological tests with respect to their ability to distinguish among the Clinical Dementia Rating (CDR) global scores of 0, 0.5, and 1.

Methods: First, we matched participants receiving V2 tests (V2 cohort) and V3 tests (V3 cohort) in their cognitive functions using tests common to both versions. Then, we compared receiver-operating characteristic (ROC) area under the curve in differentiating CDRs for the remaining tests.

Results: Some V3 tests performed better than V2 tests in differentiating between CDR 0.5 and 0, but the improvement was limited to Caucasian participants.

Discussion: Further efforts to improve the ability for early identification of cognitive decline among diverse racial groups are required.

KEYWORDS

differentiating CDR, National Alzheimer's Coordinating Center Uniform Data Set (NACC UDS), optimal cut-point, racial differences, receiver-operating characteristic area under the curve (ROC-AUC), validity

1 | INTRODUCTION

The National Alzheimer's Coordinating Center (NACC) at the University of Washington maintains a repository of valuable neuropsychological information from the Uniform Data Set (UDS) on participants from all of the National Institute on Aging (NIA)–funded Alzheimer's Disease Centers (ADCs) in the United States. There are more than 30 past and present ADCs. The UDS consists of data collection protocols administered systematically to participants enrolled into the Clinical Cores of each ADC.[1–3] Participants are recruited, enrolled, and followed on an annual basis, thereby generating center-specific longitudinal cohorts. These participants include individuals with clinical syndromic diagnoses of normal cognition (NC), mild cognitive impairment (MCI), or cognitive impairment that does not meet clinical MCI criteria, and dementia of various etiologies, including Alzheimer's disease (AD). Consent is obtained at the individual ADCs, as approved by their institutional review boards (IRBs). The UDS data include demographics, medical history, medication use, physical and neurological exam findings, clinical ratings of dementia severity (Clinical Dementia Rating [CDR] Dementia Staging Instrument),[4] and neuropsychological test scores. Systematic guidelines for clinical diagnosis are based on the most up to date published diagnostic research criteria.[5–8]

All UDS data collection instruments were constructed with the guidance and approval of the ADC Clinical Task Force (CTF), a group formed originally by the NIA to develop standardized methods for collecting longitudinal data that would encourage and support collaboration across the ADCs.[1,9] Further information on the NACC database may be found at: https://www.alz.washington.edu.html.

The first version of the UDS was available in 2005. A second version, UDS V2, was implemented in 2008, with slight revisions to neuropsychological test instructions and to other data collection forms. In 2015, UDS V3 was implemented to overcome challenges in the UDS V2 neuropsychological test battery.[10] Briefly, V3 was introduced with the aim (1) to reduce significant practice effects in longitudinal follow-up, especially in episodic memory tasks; (2) to use non-proprietary instruments; (3) to use instruments that are potentially more sensitive to earlier stages of cognitive decline; and (4) to add a visuoconstructional measure and a visual memory test, which were missing in UDS V2.

UDS V2 neuropsychological tests[2] included a measure of overall dementia severity: the Mini-Mental State Examination (MMSE)[11]; measures of attention: Trail Making Test Part A[12] and Digit Span Forward (Wechsler Memory Scale-Revised)[13]; a measure of visuomotor processing speed: Digit Symbol (Wechsler Adult Intelligence Scale-Revised)[13]; measures of executive functioning: Digit Span Backward (Wechsler Memory Scale-Revised)[13] and Trail Making Test Part B[12]; measures of episodic memory: Logical Memory Story A Immediate and Delayed Recall (Wechsler Memory Scale-Revised)[13]; and language measures: semantic fluency (Animals, Vegetables),[14] short version of the Boston Naming Test.[15,16]

In UDS V3, the Mini-Mental Status Exam (MMSE) was replaced with the Montreal Cognitive Assessment (MoCA).[17] MoCA index scores can be calculated for subsets of items in the domains of attention, memory, orientation, language, executive function, and visuospatial function (see Method and Table 1 in Appendix for more information).[17,18] Logical Memory Immediate and Delayed Recall was replaced with the Craft Story Immediate and Delayed Recall[19]; Digit Span Forward and Backward with the Number Span Forward and Backward Test; and the Boston Naming Test (BNT) with the Multilingual Naming Test (MINT).[20] The Benson Complex Figure[21] was added as a test of visuoconstructional ability (Copy condition) and as a measure of delayed visual recall. Timed phonemic fluency for the letters F and L was also added.[22] Therefore, four tests remained consistent across the two batteries: category fluency animals and vegetables and Trailmaking Tests A and B.

A crosswalk study that assessed equivalent test scores between V2 and their replacements in V3 listed above was published previously.[23]

This study examined whether the newly implemented neuropsychological tests in UDS V3 have the same validity as V2 in differentiating among global Clinical Dementia Rating (CDRs) scores of 0 (no cognitive impairment), 0.5 (questionable or mild cognitive impairment), and 1 (mild dementia). We also generated composite scores for each cognitive domain using tests in V3, and compared them with MoCA Index scores[18] in terms of their abilities to differentiate between CDR scores.

RESEARCH IN CONTEXT

1. Systematic Review: Federally funded Alzheimer's Disease Centers in the United States have been using a standardized neuropsychological test battery as part of the National Alzheimer's Coordinating Center Uniform Data Set (UDS) since 2005. Version 3 (V3) of the UDS replaced the previous version (V2) in 2015. The process for selecting a new set of neuropsychological tests for V3 and their normative test scores was published previously. A crosswalk study providing tables that allow scores on the new tests to be converted to equivalent scores was also published.
2. Interpretation: In this study, we compared V2 and V3 neuropsychological tests with respect to their ability to distinguish between the Clinical Dementia Rating (CDR) global scores of 0, 0.5, and 1. The same tests administered in both versions showed different discriminatory abilities in distinguishing CDR score of 0 and 0.5, indicating that the V2 cohort (participants assessed between 2005 and 2015) and the V3 cohort (participants assessed between 2015 and 2018) differ in their levels of cognitive abilities within CDR score of 0.5. This necessitated matching two cohorts before comparing the discriminatory abilities of remaining tests. After matching two cohorts on their cognitive abilities, the test battery in V3 improved the abilities to differentiate between CDR 0 and 0.5 in some tests, but the magnitude of these improvements was relatively small and the gain was limited to the Caucasian participants.
3. Future Directions: Data continue to accumulate. Neuropsychological test scores are available to researchers via requests to NACC, where researchers can link neuropsychological test scores with other data including clinical variables, MRI, and autopsy records. Further research with sufficient numbers of under-represented groups is required to address racial- and ethnic-specific normative scores and optimal cut-points to improve diagnostic accuracy across diverse groups.

2 | METHODS

2.1 | Data

Each AD Center in the United States enrolls participants in an NACC research cohort according to center-specific priorities; the records of the participants will be uploaded to NACC. In general, most participants come from clinician referral, self-referral by patients or family members, active recruitment through community organizations, and volunteers who wish to contribute to research. Most centers also enroll volunteers with normal cognition. Therefore, NACC participants are not an epidemiologically based sample of the U.S. population with or without dementia. Rather, they are best regarded as a referral-based or volunteer case series selected based on each center's research focus.

To reduce the influence of practice effects, the sample was restricted to individuals who received the UDS V2 or V3 neuropsychological test battery at their initial visit. Individuals had none, questionable, or MCI corresponding to a global CDR score of 0, 0.5, or 1, respectively. Data were obtained from the September 2018 data freeze.

2.2 | Statistical methods

For each individual test and composite score at baseline for V2 and V3, we fit 3 logistic regression models: CDR 0 versus 0.5; CDR 0.5 versus 1; and CDR 0 versus 1, adjusted for age, sex, years of education, and race. We examined ROC-AUCs as an indicator of overall performance in differentiating CDR groups and compared ROC-AUCs between V2 and V3 test batteries. We chose CDR global scores as the outcome measure because it is a clinical and functional assessment and is mostly determined independently from neuropsychological test results, thereby avoiding the inherent circularity of studying subjects diagnosed using neuropsychological tests. Confidence intervals for each AUC were calculated using bootstrap methods.[24] We also assessed ROC-AUCs stratified by race. Race stratification analysis was limited to Caucasian versus African American (AA) participants because other racial ethnic groups did not have a sufficient sample size.

2.3 | Sample selection process to match V2 and V3 cohorts

In recent years, the ADCs have been more interested in recruiting patients in earlier stages of cognitive decline for their center-specific longitudinal NACC research cohorts, because current research is increasingly focused on these subjects for clinical trials and primary prevention. Thus the CDR 0.5 group of the V3 cohort (enrolled after 2015) may be made up of a larger proportion of cases at an earlier stage of decline, which could potentially make it more difficult to discriminate between healthy volunteers (CDR = 0) and patients with MCI (CDR = 0.5) in the V3 cohort (enrolled after 2015) than with the CDR 0.5 group of the V2 cohort (enrolled between 2008 and 2015). Therefore, before comparing each of the V2 and V3 cognitive tests, we first matched participants in the V2 and V3 cohorts with respect to their

cognitive abilities by using the four neuropsychological tests common to both batteries. A detailed description of statistical matching methods used is found in the Appendix.

2.4 | UDS V3 domain-specific composite scores

We generated the composite scores for each cognitive domain in V3 by first calculating the Z-score for each test using the baseline distribution of mean and SD and taking the average of the Z-scores included in each domain.

The following composite scores were generated. The National Alzheimer's Disease Coordinating Center (NACC) Data Dictionary variable names are indicated in brackets below for reference:

1. Memory Composite Score: Immediate Craft Story Recall (paraphrase scoring) [CRAFTURS], Delayed Craft Story Recall (paraphrase scoring) [CRAFTDRE], Total score for delayed recall of Benson figure [UDSBENTD].
2. Language Composite Score: Category Fluency (animals) [ANIMALS], Category Fluency (vegetables) [VEG], Multilingual Naming Test (MINT) (total score) [MINTTOTS], Number of correct F-words and L-words [UDSVERTN].
3. Attention Score: Trail Making Test Part A [TRAILA (reversed Z-score)], Forward Number Span Test (# of correct trials) [DIGFORCT].
4. Executive Function Score: Trail Making Test Part B [TRAILB (reversed Z-score)], Backward Number Span Test (# of correct trials) [DIGBACCT].
5. Visuospatial Score: Total Score for copy of Benson figure [UDSBENTC].

In addition, the Global Composite score was generated for V3 in two ways. First, we took the average of the Z-scores of the five domains listed above (henceforth called UDS Global Composite-5 score). Second, in an attempt to create a consolidated version of composite scores, we generated a brief Composite Score using the average of the Z-scores in the following three tests: MoCA Total Score, Delayed Craft Story Recall, and Trail Making Test Part B; this combination of scores will henceforth be called UDS Global Composite-3 Score. These tests were chosen based on their proven ability to measure global cognition, memory, and executive dysfunction, which are hallmark domains affected by dementia and are therefore assumed to be declining continuously, even at an earlier stage of the disease spectrum.

2.5 | MoCA domain-specific index scores

We generated MoCA domain-specific index scores using an algorithm described previously.[17,18,25] The algorithm for calculating MoCA Index scores is also summarized later in the Appendix.

3 | RESULTS

UDS V2 and V3 contained 13,119 and 4894 unique subjects, respectively, who were 60 years or older with CDR score ≤1 and provided information on education and race. After excluding those with missing data from one of the four tests (Category Fluency Animals and Vegetables, Trails Making Tests A and B) used for matching in V2 and V3 cohorts, we had 4318 participants in V3. The same number of subjects was selected from V2 (detailed methods in Appendix). The comparison between selected and non-selected V2 participants did not differ in the distribution of age, sex, and education, and the correlation matrices of cognitive tests did not differ between the selected and non-selected groups, which is discussed in the Appendix. Demographic and clinical characteristics (age, sex, education, race, and cognitive test scores) for participants used for the subsequent analyses are presented in Table 1. By design, the four tests that are common in both V2 and V3 batteries and used to match the two cohorts are almost identical in their means and SDs (Table 1). In addition, in Table 1, we show MMSE scores converted from MoCA using previously published cross-walk results.[23] For the Trail Making measures, we also calculated connections per second (ie, number of correct connections/unit time to complete), which has less of a ceiling effect compared to completion time alone because, in some instances, participants exceeded the maximum cut-off time.

3.1 | V2 and V3 test battery comparisons

Table 2 shows ROC-AUC values for each cognitive test in V2 and V3 in discriminating CDR classes. In the full cohort (ie, combining race categories), Craft Story Delayed Recall (paraphrase scoring) in V3 showed higher discriminative ability in distinguishing between CDR 0.5 and 1 than the corresponding Logical Memory Delayed Recall in V2 (AUC = 0.73 for Logical Memory Delayed Recall, AUC = 0.78 for Craft Story Delayed Recall; $P < 0.01$). Among Caucasian participants, MoCA total score in V3 showed higher discriminative ability in distinguishing between CDR 0 and 0.5 than the corresponding MMSE in V2 (AUC = 0.78 for MMSE, AUC = 0.81 for MoCA; $P < 0.01$). Among African American participants, Backward Number Span (V3) showed higher discriminative ability than Digit Span Backward (V2) in distinguishing between CDR 0 and 1 (AUC = 0.69 for Digit Span Backward, AUC = 0.87 for Backward Number Span; $P < 0.05$), as well as the ability to discriminate between CDR 0.5 and 1 (AUC = 0.56 and AUC = 0.77, respectively, $P < 0.05$). No other tests differed significantly between V2 and V3 in their discriminatory abilities.

3.2 | Within version analysis in V3 tests

MoCA Index scores versus V3 individual test. For differentiating CDR 0 versus 0.5, MoCA Memory Index score showed similar ROC-AUC (AUC = 0.77) as Craft Story Delayed Recall (0.77 for verbatim scoring and 0.78 for paraphrase scoring). Similarly, MoCA Executive Index

TABLE 1 Demographic characteristics and cognitive test scores of study participants (age ≥60, non-missing education and race/ethnicity, non-missing Trail A, Trail B, category fluency [animals], and category fluency [vegetables])

	Version 2 (N = 4318)*			Version 3 (N = 4318)		
	CDR = 0N = 2274	CDR = 0.5N = 1722	CDR = 1N = 322	CDR = 0N = 2274	CDR = 0.5N = 1722	CDR = 1N = 322
Age, y (mean, SD)	71.6 (7.4)	73.2 (7.5)	74.7 (7.8)	71.2 (6.5)	72.6 (7.0)	74.8 (7.6)
Education, y (mean, SD)	16.0 (2.9)	15.7 (3.3)	15.2 (3.1)	16.3 (2.7)	16.0 (2.9)	15.5 (3.1)
Female (%)	64.0	45.2	45.7	65.3	50.6	46.3
Race (%)						
Caucasian	76.4	82.2	88.8	76.4	82.2	88.8
African American	16.2	12.5	6.2	16.2	12.5	6.2
American Indian or Alaska Native	0.6	0.8	1.2	1.2	0.7	0
Native Hawaiian or Pacific Islander	0	0.1	0.3	0.1	0	0.3
Asian	2.7	1.5	1.6	3.5	2.3	3.4
Multiracial	4.1	2.9	1.9	2.6	2.3	1.2
Cognitive Test Scores						
Common tests across two versions **(V2 and V3 are matched using the following 4 tests and therefore by design the distributions are similar between V2 and V3 cohorts)**						
Animals	20.7 (5.5)	16.4 (5.4)	12.2 (5.1)	20.7 (5.6)	16.5 (5.5)	12.2 (5.1)
Vegetables	14.4 (4.0)	10.7 (4.2)	7.2 (3.6)	14.4 (4.1)	10.8 (4.3)	7.3 (3.7)
Trail A	33.7 (13.6)	43.2 (22.4)	59.6 (32.8)	33.7 (13.9)	43.1 (22.7)	59.7 (32.0)
Trail B	90.0 (47.0)	138.1 (80.2)	210.9 (89.3)	90.5 (48.1)	138.2 (80.1)	209.1 (88.8)
V2 specific tests						
MMSE	28.9 (1.3)	26.7 (2.8)	23.1 (3.7)			
Logical memory immediate recall	13.1 (3.8)	8.9 (4.4)	5.3 (3.9)			
Logical memory delayed recall	11.9 (4.0)	6.7 (4.9)	3.2 (4.1)			
Digit span forward	8.5 (2.0)	7.8 (2.0)	7.4 (2.1)			
Digit span backward	6.7 (2.2)	5.8 (2.1)	4.9 (2.0)			
Boston naming test	27.1 (3.1)	24.7 (5.0)	21.4 (6.9)			
Digit symbol	47.1 (11.3)	38.2 (12.2)	28.8 (12.7)			

(Continues)

TABLE 1 (Continued)

	Version 2 (N = 4318)*			Version 3 (N = 4318)		
	CDR = 0N = 2274	CDR = 0.5N = 1722	CDR = 1N = 322	CDR = 0N = 2274	CDR = 0.5N = 1722	CDR = 1N = 322
V3 specific tests						
MoCA				25.9 (2.8)	22.1 (4.1)	17.1 (4.5)
Craft Story immediate recall (verbatim scoring)				21.4 (6.6)	15.1 (7.3)	9.5 (5.9)
Craft Story delayed recall (verbatim scoring)				18.5 (6.6)	11.1 (7.8)	4.2 (5.5)
Number span forward test (correct trials)				8.1 (2.3)	7.5 (2.3)	7.0 (2.3)
Number span backward test (correct trials)				6.9 (2.2)	5.9 (2.1)	5.0 (2.1)
Multilingual naming test (MINT) (total score)				29.8 (2.5)	27.8 (4.7)	24.7 (6.2)
Copy of Benson Figure (total score)				15.5 (1.4)	15.0 (2.0)	14.1 (3.1)
Delayed drawing of Benson figure (total score)				11.1 (3.0)	7.3 (4.4)	3.5 (3.9)
MoCA in V3 converted to MMSE using Cross Walk results (Monsell, et al)[23]	28.9 (1.4)	26.7 (2.8)	23.0 (3.8)	29.1 (1.3)	26.9 (3.0)	23.0 (4.0)

(0.70) and Attention Index (0.68) showed a discriminative ability similar to that of the respective cognitive tests tapping the same domains (0.71 for Trail making B for executive; 0.66 for Digit Span Backward for attention). For the language domain, Animal Fluency (0.72) and MINT (0.70) showed better abilities to differentiate CDR 0 versus 0.5 than MoCA Language index (0.65, $P < 0.05$) (Table 3). Similar patterns were found for CDR 0.5 versus 1 comparisons.

MoCA Index score versus UDS Composite Scores. The UDS Memory Composite was better able to differentiate CDR 0 versus 0.5 than the MoCA Memory Index (0.80 for UDS Memory Composite versus 0.77 for MoCA Memory Index, $P < 0.05$). In addition, across all CDR comparisons ($P < 0.01$), UDS Language Composite performed better than MoCA Langue Index score. On the other hand, MoCA Executive (0.70), Attention (0.68), and Visuospatial (0.67) Indexes performed better than UDS Executive (0.61), Attention (0.61), and Visuospatial (0.63) composites, respectively, in differentiating CDR 0 versus 0.5 ($P < 0.01$). Similar patterns were found in differentiating CDR 0 versus 1.

3.3 | UDS Global Composite-5 score versus UDS Global Composite-3 score

The UDS Global Composite-5 score (average of five domains) and Composite-3 score (average of three domains) showed similar ROC-AUCs across all CDR comparisons.

Racial differences in MoCA index and UDS composite scores: In all MoCA Indexes except orientation, it was easier to differentiate CDR 0 versus 0.5 among Caucasian than AA participants (0.79 versus 0.69, respectively, for memory, $P < 0.05$; 0.71 versus 0.66 for executive function, $P < 0.01$; 0.66 versus 0.63 for language, $P < 0.01$; 0.70 versus 0.66 for attention, $P < 0.01$; 0.68 versus 0.63 for visuospatial, $P < 0.01$). On the other hand, in all MoCA Indexes except memory and orientation, it was easier to differentiate CDR 0 versus 1 among AA than Caucasian participants (0.94 versus 0.87, respectively, for executive function, 0.85 versus 0.79 for language; 0.90 versus 0.83 for attention; 0.90 versus 0.82 for visuospatial, $P < 0.01$ for all comparisons). It was also easier to differentiate CDR 0.5 versus 1 among AA than Caucasian participants for executive function (0.84 vs 0.73, respectively, $P < 0.05$), language (0.71 vs 0.66, $P < 0.05$), and attention (0.77 vs 0.69, $P < 0.05$).

Similarly we found differences in the discriminative ability of UDS composites. It was easier to differentiate CDR 0 versus 0.5 among Caucasian than AA participants in Memory (0.82 vs 0.71, $P < 0.01$) and Language (0.75 vs 0.68, $P < 0.01$). On the other hand, it was easier to differentiate between CDR 0 versus 1 in AA participants as compared to Caucasian participants in Attention (0.85 vs 0.70, $P < 0.01$), as well as Visuospatial composite (0.86 vs 0.73, $P < 0.01$). Likewise, Attention and Visuospatial composites were easier to differentiate between CDR 0.5 versus 1 in AA participants as compared to Caucasian participants (0.74 vs 0.63 for attention, $P < 0.05$; and 0.74 vs 0.64 for visuospatial, $P < 0.05$). The UDS Composite-5 score showed racial difference in discriminative ability between CDR 0 versus 0.5 (0.76 among Caucasians versus 0.66 among AA participants, $P < 0.01$).

TABLE 2 Comparison of ROC-AUC between V2 and V3 cognitive tests for the entire sample and stratified by Race

All Subjects	Version 2 (n = 4318)				Version 3 (n = 4318)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
NACCMMSE	0.780	0.956	0.803	MOCATOTS	0.794	0.966	0.814
Total mini mental state exam (MMSE) score	(0.765, 0.794)	(0.944, 0.969)	(0.777, 0.829)	MoCA total raw score	(0.779, 0.808)	(0.956, 0.976)	(0.789, 0.838)
LOGIMEM	0.777	0.927	0.741	CRAFTVRS	0.750	0.918	0.749
Immediate story recall	(0.763, 0.792)	(0.911, 0.944)	(0.711, 0.77)	Immediate craft story recall (verbatim scoring)	(0.734, 0.765)	(0.902, 0.934)	(0.72, 0.777)
				CRAFTURS	0.764	0.921	0.751
				Immediate craft story recall (paraphrase scoring)	(0.749, 0.779)	(0.905, 0.938)	(0.722, 0.779)
MEMUNITS	0.796	0.940	0.735	CRAFTDVR	0.772	0.947	0.784
Delayed story recall	(0.781, 0.81)	(0.924, 0.957)	(0.706, 0.764)	Delayed craft story recall (verbatim scoring)	(0.757, 0.787)	(0.932, 0.962)	(0.757, 0.811)
				CRAFTDRE	0.782	0.949	0.788**
				Delayed craft story recall (paraphrase scoring)	(0.767, 0.797)	(0.935, 0.964)	(0.761, 0.815)
DIGIF	0.655	0.738	0.609	DIGFORCT	0.631	0.737	0.634
Digit span forward (correct trials)	(0.638, 0.672)	(0.71, 0.766)	(0.576, 0.641)	Forward number span test (correct trials)	(0.613, 0.649)	(0.708, 0.766)	(0.600, 0.668)
				DIGFORSL	0.633	0.734	0.629
				Forward number span test (longest span)	(0.615, 0.651)	(0.704, 0.763)	(0.595, 0.664)
DIGIB	0.669	0.786	0.644	DIGBACCT	0.666	0.799	0.668
Digit span backward (correct trials)	(0.653, 0.686)	(0.76, 0.812)	(0.612, 0.677)	Backward number span test (correct trials)	(0.648, 0.683)	(0.773, 0.825)	(0.635, 0.701)
				DIGBACLS	0.653	0.790	0.667
				Backward number span test (longest span)	(0.636, 0.671)	(0.763, 0.816)	(0.634, 0.700)
BOSTON	0.715	0.836	0.671	MINTTOTS	0.703	0.854	0.695
Boston Naming Test (total score)	(0.699, 0.731)	(0.812, 0.86)	(0.638, 0.704)	Multilingual naming test (MINT) (total score)	(0.686, 0.719)	(0.831, 0.877)	(0.662, 0.729)
WAIS	0.721	0.877	0.722	UDSVERFC	0.643	0.776	0.658
WAIS-R Digit Symbol	(0.705, 0.737)	(0.856, 0.898)	(0.691, 0.753)	Number of correct F-words generated	(0.625, 0.66)	(0.749, 0.802)	(0.626, 0.691)
				UDSVERLC	0.638	0.778	0.661
				Number of correct L-words generated	(0.62, 0.655)	(0.751, 0.805)	(0.628, 0.693)
				UDSVERTN	0.644	0.783	0.661
				Number of correct F-words and L-words	(0.627, 0.662)	(0.756, 0.809)	(0.629, 0.694)
				UDSBENTC	0.633	0.755	0.647
				Total Score for copy of Benson figure	(0.615, 0.651)	(0.726, 0.784)	(0.613, 0.682)

(Continues)

TABLE 2 (Continued)

All Subjects	Version 2 (n = 4318)				Version 3 (n = 4318)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
				UDSBENTD	0.770	0.933	0.759
				Total score for delayed drawing of Benson figure	(0.754, 0.785)	(0.916, 0.95)	(0.729, 0.789)
ANIMALS	0.741	0.892	0.724	ANIMALS	0.725	0.892	0.731
Category Fluency (animals)	(0.725, 0.756)	(0.872, 0.911)	(0.693, 0.754)	Category fluency (animals)	(0.709, 0.74)	(0.874, 0.911)	(0.702, 0.761)
VEG	0.745	0.911	0.740	VEG	0.734	0.910	0.746
Category Fluency (vegetables)	(0.730, 0.760)	(0.894, 0.929)	(0.711, 0.769)	Category fluency (vegetables)	(0.719, 0.75)	(0.893, 0.928)	(0.717, 0.774)
TRAILA	0.688	0.839	0.693	TRAILA	0.674	0.842	0.704
Trail Making Test Part A	(0.671, 0.704)	(0.815, 0.863)	(0.661, 0.725)	Trail making test Part A	(0.657, 0.691)	(0.819, 0.864)	(0.674, 0.734)
TRAILA_CONNECT_SEC	0.688	0.839	0.700	TRAILA_CONNECT_SEC	0.672	0.843	0.707
Trail A Lines per second	(0.671, 0.704)	(0.816, 0.863)	(0.669, 0.731)	Trail A Lines per second	(0.655, 0.688)	(0.82, 0.865)	(0.678, 0.737)
TRAILB	0.730	0.896	0.740	TRAILB	0.719	0.891	0.742
Trail Making Test Part B	(0.714, 0.745)	(0.876, 0.916)	(0.710, 0.770)	Trail Making Test Part B	(0.703, 0.735)	(0.871, 0.912)	(0.714, 0.771)
TRAILB_CONNECT_SEC	0.731	0.893	0.746	TRAILB_CONNECT_SEC	0.720	0.890	0.742
Trail B Lines per second	(0.715, 0.746)	(0.871, 0.915)	(0.716, 0.777)	Trail A Lines per second	(0.704, 0.736)	(0.868, 0.911)	(0.713, 0.772)
African American subjects							
	Version 2 (n = 604)				Version 3 (n = 604)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
NACCMMSE	0.750	0.962	0.837	MOCATOTS	0.702	0.972	0.858
Total mini mental state exam (MMSE) score	(0.707, 0.792)	(0.925, 0.998)	(0.717, 0.958)	MoCA total raw score	(0.656, 0.748)	(0.939, 1)	(0.775, 0.941)
LOGIMEM	0.714	0.937	0.811	CRAFTVRS	0.686	0.937	0.813
Immediate story recall	(0.67, 0.758)	(0.90, 0.974)	(0.722, 0.901)	Immediate craft story recall (verbatim scoring)	(0.64, 0.732)	(0.874, 1.00)	(0.716, 0.911)
				CRAFTURS	0.693	0.927	0.790
				Immediate craft story recall (paraphrase scoring)	(0.648, 0.739)	(0.864, 0.991)	(0.681, 0.899)
MEMUNITS	0.729	0.983	0.847	CRAFTDVR	0.697	0.945	0.834
Delayed story recall	(0.685, 0.774)	(0.97, 0.995)	(0.784, 0.91)	Delayed craft story recall (verbatim scoring)	(0.652, 0.742)	(0.885, 1.00)	(0.739, 0.929)
				CRAFTDRE	0.698	0.925	0.806
				Delayed craft story recall (paraphrase scoring)	(0.653, 0.743)	(0.852, 0.998)	(0.696, 0.916)
DIGIF	0.617	0.697	0.588	DIGFORCT	0.636	0.798	0.681

(Continues)

TABLE 2 (Continued)

African American subjects							
	Version 2 (n = 604)				Version 3 (n = 604)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
Digit span forward (correct trials)	(0.568, 0.666)	(0.582, 0.812)	(0.467, 0.709)	Forward number span test (correct trials)	(0.588, 0.683)	(0.678, 0.918)	(0.544, 0.818)
				DIGFORSL	0.639	0.801	0.679
				Forward number span test (longest span)	(0.591, 0.686)	(0.679, 0.923)	(0.539, 0.818)
DIGIB	0.620	0.690	0.563	DIGBACCT	0.642	0.879*	0.772*
Digit span backward (correct trials)	(0.571, 0.669)	(0.561, 0.818)	(0.42, 0.705)	Backward number span test (correct trials)	(0.594, 0.689)	(0.809, 0.949)	(0.662, 0.882)
				DIGBACLS	0.636	0.865	0.737
				Backward number span test (longest span)	(0.588, 0.683)	(0.795, 0.935)	(0.62, 0.855)
BOSTON	0.694	0.839	0.689	MINTTOTS	0.669	0.888	0.755
Boston Naming Test (total score)	(0.648, 0.739)	(0.739, 0.938)	(0.562, 0.815)	Multilingual naming test (MINT) (total score)	(0.624, 0.715)	(0.812, 0.964)	(0.646, 0.864)
WAIS	0.700	0.897	0.757	UDSVERFC	0.643	0.860	0.749
WAIS-R Digit Symbol	(0.655, 0.745)	(0.836, 0.958)	(0.662, 0.853)	Number of correct F-words generated	(0.595, 0.692)	(0.775, 0.945)	(0.627, 0.87)
				UDSVERLC	0.634	0.866	0.749
				Number of correct L-words generated	(0.585, 0.682)	(0.785, 0.946)	(0.647, 0.852)
				UDSVERTN	0.644	0.865	0.758
				Number of correct F-words and L-words	(0.595, 0.693)	(0.782, 0.948)	(0.646, 0.869)
				UDSBENTC	0.618	0.864	0.745
				Total Score for copy of Benson figure	(0.569, 0.667)	(0.792, 0.936)	(0.649, 0.841)
				UDSBENTD	0.681	0.950	0.814
				Total score for delayed drawing of Benson figure	(0.634, 0.729)	(0.909, 0.991)	(0.735, 0.894)
ANIMALS	0.675	0.829	0.682	ANIMALS	0.652	0.883	0.756
Category Fluency (animals)	(0.63, 0.721)	(0.728, 0.93)	(0.549, 0.815)	Category fluency (animals)	(0.605, 0.699)	(0.812, 0.954)	(0.645, 0.866)
VEG	0.677	0.893	0.773	VEG	0.656	0.881	0.774
Category Fluency (vegetables)	(0.631, 0.724)	(0.824, 0.962)	(0.657, 0.889)	Category fluency (vegetables)	(0.609, 0.704)	(0.792, 0.971)	(0.66, 0.888)
TRAILA	0.651	0.805	0.710	TRAILA	0.655	0.888	0.769
Trail Making Test Part A	(0.604, 0.699)	(0.691, 0.918)	(0.595, 0.825)	Trail making test Part A	(0.608, 0.702)	(0.808, 0.968)	(0.671, 0.867)
TRAILA_CONNECT_SEC	0.658	0.804	0.690	TRAILA_CONNECT_SEC	0.659	0.891	0.755
Trail A Lines per second	(0.611, 0.705)	(0.69, 0.919)	(0.565, 0.816)	Trail A Lines per second	(0.613, 0.706)	(0.822, 0.959)	(0.654, 0.856)
TRAILB	0.679	0.919	0.809	TRAILB	0.672	0.939	0.809

(Continues)

TABLE 2 (Continued)

African American subjects							
	Version 2 (n = 604)				Version 3 (n = 604)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
Trail Making Test Part B	(0.632, 0.726)	(0.848, 0.989)	(0.709, 0.909)	Trail making test Part B	(0.625, 0.719)	(0.893, 0.985)	(0.723, 0.895)
TRAILB_CONNECT_SEC	0.675	0.925	0.839	TRAILB_CONNECT_SEC	0.677	0.943	0.810
Trail B Lines per second	(0.628, 0.722)	(0.845, 1.000)	(0.735, 0.943)	Trail A Lines per second	(0.63, 0.724)	(0.900, 0.986)	(0.724, 0.896)
Caucasian subjects							
	Version 2 (n = 3439)				Version 3 (n = 3439)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
NACCMMSE	0.784	0.955	0.794	MOCATOTS	0.817**	0.966	0.804
Total mini mental state exam (MMSE) score	(0.768, 0.8)	(0.941, 0.969)	(0.766, 0.822)	MoCA total raw score	(0.802, 0.832)	(0.955, 0.977)	(0.776, 0.832)
LOGIMEM	0.783	0.924	0.732	CRAFTVRS	0.760	0.914	0.739
Immediate story recall	(0.766, 0.799)	(0.906, 0.942)	(0.7, 0.764)	Immediate craft story recall (verbatim scoring)	(0.743, 0.778)	(0.896, 0.932)	(0.707, 0.771)
				CRAFTURS	0.776	0.920	0.743
				Immediate craft story recall (paraphrase scoring)	(0.759, 0.792)	(0.902, 0.938)	(0.711, 0.774)
MEMUNITS	0.802	0.938	0.725	CRAFTDVR	0.785	0.945	0.776
Delayed story recall	(0.786, 0.817)	(0.92, 0.955)	(0.694, 0.757)	Delayed craft story recall (verbatim scoring)	(0.768, 0.801)	(0.929, 0.962)	(0.746, 0.806)
				CRAFTDRE	0.798	0.952	0.782**
				Delayed craft story recall (paraphrase scoring)	(0.782, 0.814)	(0.936, 0.967)	(0.753, 0.812)
DIGIF	0.658*	0.724	0.594	DIGFORCT	0.627	0.719	0.626
Digit span forward (correct trials)	(0.639, 0.677)	(0.693, 0.754)	(0.559, 0.629)	Forward number span test (correct trials)	(0.607, 0.647)	(0.686, 0.752)	(0.588, 0.663)
				DIGFORSL	0.628	0.712	0.618
				Forward number span test (longest span)	(0.608, 0.648)	(0.679, 0.745)	(0.58, 0.656)
DIGIB	0.674	0.777	0.635	DIGBACCT	0.669	0.785	0.657
Digit span backward (correct trials)	(0.655, 0.692)	(0.749, 0.805)	(0.599, 0.67)	Backward number span test (correct trials)	(0.65, 0.689)	(0.756, 0.814)	(0.621, 0.694)
				DIGBACLS	0.657	0.777	0.655
				Backward number span test (longest span)	(0.637, 0.676)	(0.747, 0.806)	(0.619, 0.692)

(Continues)

TABLE 2 (Continued)

Caucasian subjects							
	Version 2 (n = 3439)				Version 3 (n = 3439)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
BOSTON	0.730	0.838	0.658	MINTTOTS	0.719	0.855	0.679
Boston Naming Test (total score)	(0.712, 0.747)	(0.812, 0.864)	(0.622, 0.694)	Multilingual naming test (MINT) (total score)	(0.701, 0.737)	(0.83, 0.88)	(0.641, 0.716)
WAIS	0.725	0.868	0.711	UDSVERFC	0.637	0.753	0.645
WAIS-R Digit Symbol	(0.707, 0.742)	(0.844, 0.892)	(0.678, 0.745)	Number of correct F-words generated	(0.617, 0.656)	(0.723, 0.784)	(0.609, 0.681)
				UDSVERLC	0.632	0.760	0.651
				Number of correct L-words generated	(0.613, 0.652)	(0.73, 0.79)	(0.615, 0.686)
				UDSVERTN	0.638	0.762	0.649
				Number of correct F-words and L-words	(0.618, 0.658)	(0.732, 0.793)	(0.613, 0.684)
				UDSBENTC	0.627	0.733	0.640
				Total Score for copy of Benson figure	(0.607, 0.647)	(0.7, 0.767)	(0.602, 0.678)
				UDSBENTD	0.785	0.929	0.750
				Total score for delayed drawing of Benson figure	(0.769, 0.802)	(0.91, 0.949)	(0.717, 0.783)
ANIMALS	0.755	0.894	0.714	ANIMALS	0.740	0.889	0.719
Category Fluency (animals)	(0.739, 0.772)	(0.874, 0.914)	(0.681, 0.747)	Category fluency (animals)	(0.723, 0.757)	(0.87, 0.909)	(0.687, 0.751)
VEG	0.756	0.909	0.725	VEG	0.746	0.906	0.729
Category Fluency (vegetables)	(0.739, 0.773)	(0.891, 0.928)	(0.693, 0.757)	Category fluency (vegetables)	(0.729, 0.764)	(0.887, 0.925)	(0.698, 0.76)
TRAILA	0.693	0.836	0.680	TRAILA	0.676	0.835	0.696
Trail Making Test Part A	(0.675, 0.712)	(0.811, 0.861)	(0.645, 0.714)	Trail making test Part A	(0.658, 0.695)	(0.81, 0.859)	(0.663, 0.729)
TRAILA_CONNECT_SEC	0.692	0.834	0.689	TRAILA_CONNECT_SEC	0.674	0.834	0.698
Trail A Lines per second	(0.673, 0.71)	(0.809, 0.859)	(0.655, 0.722)	Trail A Lines per second	(0.655, 0.692)	(0.809, 0.858)	(0.666, 0.73)
TRAILB	0.739	0.887	0.727	TRAILB	0.727	0.884	0.734
Trail Making Test Part B	(0.721, 0.756)	(0.864, 0.91)	(0.694, 0.761)	Trail making test Part B	(0.71, 0.745)	(0.86, 0.907)	(0.702, 0.766)
TRAILB_CONNECT_SEC	0.739	0.886	0.735	TRAILB_CONNECT_SEC	0.727	0.883	0.734
Trail B Lines per second	(0.721, 0.756)	(0.862, 0.911)	(0.702, 0.769)	Trail A Lines per second	(0.709, 0.745)	(0.858, 0.907)	(0.702, 0.767)

Adjusted for age, gender and years of education. For models for all subjects, race was also adjusted. Subjects missing years of education are excluded.
Shaded tests are used to match V2 and V3 cohorts. Therefore, by design, ROC-AUC of these tests are similar between the V2 and V3 cohorts.
*: significant evidence that version 2 ROC-AUC not equal to version 3 AUC (at 0.05 significance level, two tailed)
The test which showed better performance is indicted by * or **.
Comparisons made between following 10 pairs: MMSE and MOCATOTS, LOGIMEM and CRAFTURS, MEMUNITS and CRAFTDRE, DIGIB and DIGBACCT, DIGIF and DIGFORSL, BOSTON and MINTTOTS. TRAILA, TRAILB, ANIMALS, and VEG were used to match the V2 and V3 cohorts. When there is more than one test in V 3 to compare to V 2, we chose the one with the higher ROC-AUC for comparisons.

(Continues)

TABLE 2 (Continued)

Adjusted for age, gender and years of education. Subjects missing years of education are excluded.
Shaded tests are used to match V2 and V3 cohorts. Therefore, by design, ROC-AUC of these tests are similar between the V2 and V3 cohorts.
*: significant evidence that version 2 ROC-AUC not equal to version 3 AUC (at 0.05 significance level, two tailed).
**: significant evidence that version 2 ROC-AUC not equal to version 3 AUC (at 0.01 significance level, two tailed).
The test which showed better performance is indicted by * or **.
Comparisons made between following 10 pairs: MMSE and MOCATOTS, LOGIMEM and CRAFTURS, MEMUNITS and CRAFTDRE, DIGIB and DIGBACCT, DIGIF and DIGFORSL, BOSTON and MINTTOTS. TRAILA, TRAILB, ANIMALS, and VEG were used to match the V2 and V3 cohorts. When there is more than one test in V 3 to compare to V 2, we chose the one with the higher ROC-AUC for comparisons.
Adjusted for age, gender and years of education. Subjects missing years of education are excluded.
Shaded tests are used to match V2 and V3 cohorts. Therefore, by design, ROC-AUC of these tests are similar between the V2 and V3 cohorts.
*: significant evidence that version 2 ROC-AUC not equal to version 3 AUC (at 0.05 significance level, two tailed)
**: significant evidence that version 2 ROC-AUC not equal to version 3 AUC (at 0.01 significance level, two tailed)
The test which showed better performance is indicted by * or **.
Comparisons made between following 10 pairs: MMSE and MOCATOTS, LOGIMEM and CRAFTURS, MEMUNITS and CRAFTDRE, DIGIB and DIGBACCT, DIGIF and DIGFORSL, BOSTON and MINTTOTS. TRAILA, TRAILB, ANIMALS, and VEG were used to match the V2 and V3 cohorts. When there is more than one test in V 3 to compare to V 2, we chose the one with the higher ROC-AUC for comparisons.

TABLE 3 ROC-AUC for MoCA index scores and domain specific composite scores in Version 3: comparisons between (1) MoCA Index vs Composite and (2) Caucasian participants vs African American participants

	All subjects (n = 4318)			African American subjects (n = 604)			Caucasian subjects (n = 3439)		
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
MoCA memory index	0.779	0.945	0.760	0.692	0.923	0.791	0.795*	0.945	0.749
	(0.764, 0.794)	(0.931, 0.958)	(0.733, 0.788)	(0.647, 0.737)	(0.849, 0.997)	(0.69, 0.893)	(0.779, 0.811)	(0.931, 0.958)	(0.718, 0.78)
MoCA executive index	0.704##	0.888##	0.743#	0.669	0.948**	0.845*	0.711**	0.878	0.731
	(0.687, 0.72)	(0.868, 0.908)	(0.713, 0.773)	(0.622, 0.716)	(0.908, 0.987)	(0.769, 0.921)	(0.692, 0.729)	(0.855, 0.902)	(0.698, 0.765)
MoCA language index	0.658	0.803	0.672	0.633	0.856**	0.719*	0.667**	0.794	0.661
	(0.64, 0.675)	(0.777, 0.829)	(0.639, 0.704)	(0.586, 0.681)	(0.769, 0.943)	(0.599, 0.84)	(0.647, 0.686)	(0.765, 0.823)	(0.625, 0.697)
MoCA attention index	0.688##	0.848##	0.709#	0.661	0.905**	0.775*	0.700**	0.838	0.698
	(0.671, 0.705)	(0.824, 0.872)	(0.677, 0.741)	(0.614, 0.708)	(0.852, 0.959)	(0.677, 0.872)	(0.681, 0.718)	(0.811, 0.866)	(0.663, 0.733)
MoCA visuospatial index	0.679##	0.837##	0.690	0.631	0.901**	0.793	0.688**	0.825	0.678
	(0.662, 0.696)	(0.814, 0.861)	(0.658, 0.722)	(0.582, 0.679)	(0.842, 0.96)	(0.708, 0.878)	(0.669, 0.707)	(0.798, 0.852)	(0.643, 0.714)
MoCA orientation index	0.704	0.919	0.788	0.667	0.943	0.833	0.711	0.916	0.780
	(0.688, 0.721)	(0.90, 0.938)	(0.76, 0.817)	(0.619, 0.715)	(0.889, 0.996)	(0.744, 0.922)	(0.693, 0.73)	(0.895, 0.937)	(0.749, 0.811)
Memory composite score	0.805#	0.959	0.797	0.715	0.952	0.827	0.822**	0.960	0.790
	(0.791, 0.819)	(0.945, 0.973)	(0.770, 0.824)	(0.670, 0.759)	(0.900, 1.000)	(0.730, 0.924)	(0.807, 0.837)	(0.945, 0.974)	(0.760, 0.820)
Executive function composite score	0.617	0.771	0.682	0.631	0.831	0.700	0.605	0.748	0.673
	(0.599, 0.634)	(0.743, 0.799)	(0.650, 0.714)	(0.582, 0.680)	(0.733, 0.928)	(0.587, 0.812)	(0.585, 0.626)	(0.716, 0.780)	(0.639, 0.708)
Language composite score	0.740##	0.913##	0.743##	0.683	0.926	0.792	0.755**	0.910	0.727
	(0.724, 0.756)	(0.895, 0.930)	(0.714, 0.773)	(0.637, 0.730)	(0.879, 0.974)	(0.691, 0.894)	(0.737, 0.772)	(0.891, 0.929)	(0.695, 0.760)
Attention composite score	0.614	0.738	0.649	0.618	0.853**	0.746**	0.605	0.709	0.634
	(0.596, 0.632)	(0.708, 0.767)	(0.615, 0.683)	(0.569, 0.667)	(0.769, 0.936)	(0.655, 0.837)	(0.585, 0.625)	(0.674, 0.743)	(0.596, 0.672)
Visuospatial composite score	0.633	0.755	0.647	0.618	0.864**	0.745**	0.627	0.733	0.640

(Continues)

TABLE 3 (Continued)

Test	All subjects (n = 4318)			African American subjects (n = 604)			Caucasian subjects (n = 3439)		
	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
	(0.615, 0.651)	(0.726, 0.784)	(0.613, 0.682)	(0.569, 0.667)	(0.792, 0.936)	(0.649, 0.841)	(0.607, 0.647)	(0.700, 0.767)	(0.602, 0.678)
UDS-3 Global composite score	0.747	0.908	0.722	0.653	0.875	0.756	0.767**	0.910	0.712
	(0.731, 0.763)	(0.888, 0.929)	(0.691, 0.753)	(0.605, 0.700)	(0.778, 0.972)	(0.640, 0.872)	(0.749, 0.784)	(0.889, 0.932)	(0.678, 0.745)
Global composite score	0.748	0.905	0.724	0.660	0.913	0.796	0.765**	0.907	0.717
	(0.732, 0.764)	(0.887, 0.923)	(0.692, 0.755)	(0.613, 0.707)	(0.863, 0.963)	(0.705, 0.887)	(0.748, 0.782)	(0.887, 0.926)	(0.682, 0.751)

*: P < 0.05, **: P < 0.01. Statistically significant difference in comparing between African American subjects and Caucasians

#: P < 0.05, ## : P < 0.01. Statistically significant difference in comparing between MoCA Index and UDS Composites

The test which showed better ability is indicted by * or ** (comparison between racial groups), or by # or ## (comparison between MoCA Index and UDS composite scores).

Adjusted for age, gender, and years of education. For regression involving all subjects, additional adjustment for race.

MoCA Memory Index Score is composed of the number of words recalled in delayed free, category-cued, and multiple-choice conditions.

MoCA Executive Index Score is composed of trail-making, clock, digit span, letter A tapping, serial 7 subtraction, letter fluency, and abstraction.

MoCA Visuospatial Index Score is composed of cube copy, clock, and naming.

MoCA Language Index Score is composed of naming, sentence repetition, and letter fluency.

MoCA Attention Index Score is composed of digit span, letter A tapping, serial 7 subtraction, sentence repetition, and words recalled in both immediate recall trials.

MoCA Orientation Index Score is composed of all orientation items.

*: *P* < 0.05, **: *P* < 0.01. Statistically significant difference in comparing between African American subjects and Caucasians

#: *P* < 0.05, ## : *P* < 0.01. Statistically significant difference in comparing between MoCA Index and UDS Composites

The test which showed better performance is indicted by * or ** (comparison between racial groups) or by # or ## (comparison between MoCA Index and UDS composite scores).

Adjusted for age, gender, and years of education. For regression involving all subjects, additional adjustment for race.

UDS Memory Composite Score is composed of immediate craft story recall (paraphrase scoring), delayed craft story recall (paraphrase scoring), total score for delayed drawing of Benson figure, and MoCA memory index score.

UDS Language Composite Score is composed of category fluency (animals), category fluency (vegetables), MINT total score, number of correct F-words and L-words, and MoCA language index.

UDS Attention Score is composed of trail making test part A, forward number span test (correct trials), and MoCA attentional index.

UDS Executive Function Score is composed of trail making test part B, backward number span test (correct trials), and MoCA executive functions index.

UDS Visuospatial Score is composed of total score for copy of Benson figure and MoCA Visuospatial Index.

UDS Global Composite-5 Score is the average of 5 domain specific composite scores.

UDS Global Composite-3 Score is composed of MoCA total score, delayed craft story recall, and trail making test part B.

4 | DISCUSSION

This study examined whether the newly implemented neuropsychological tests in UDS V3 have the same validity as V2 in differentiating among global Clinical Dementia Ratings of 0 (no cognitive impairment), 0.5 (questionable or mild cognitive impairment), and 1 (mild dementia). There are two major findings. First, among Caucasian participants, MoCA total score in V3 showed better ability than MMSE in V2 in differentiating between CDR 0 and 0.5. Craft Story Delayed Recall (paraphrase scoring) in V3 also showed better ability than Logical Memory Delayed Recall in V2 in differentiating between CDR 0.5 and 1. However, the magnitudes of these improvements are not large, with fewer than 5 units of improvement in both cases, even though they were statistically significant.

Second, among AA participants, we saw a large improvement in differentiating between CDR 1 both from CDR 0 and CDR 0.5, with V3 gaining more than 20 units over V2 in ROC-AUCs in Backward Number Span. (eg, AUC = 0.56 for Digit Span Backward in V2; AUC = 0.77 for Backward Number Span in V3, $P < 0.05$, for the CDR 0.5 vs 1 comparison). Because these digit number tests are similar between V2 and V3, this finding could suggest a possible difference in the characteristics of AA participants with CDR = 1 between the cohorts, rather than gains due to the modified tests. This speculation is also supported by the fact that most MoCA Index scores and some UDS Composite scores showed better ability to distinguish CDR 0.5 from CDR 1 among AA participants than Caucasian participants; it is likely that AA participants in the V3 cohort had later stage CDR 1. It is difficult to speculate further about any potential reasons for the above findings, given the small number of those with CDR = 1 among the AA participants.

Some other noteworthy findings include the results of comparisons between the UDS Global Composite-5 score (average of five domains) and the Composite-3 score (average of three domains); both had similar ROC-AUCs across all CDR comparisons, suggesting that the UDS Global Composite-3 can be a shorter alternative to the Global Composite-5, which requires scores in all five domains. Further reproducibility studies using more diversified participants are needed to validate the finding. In addition, this study clearly showed a large shift in cohort characteristics in NACC participants over time (see Appendix for details). In the group of participants recruited more recently (ie, the V3 cohort), there was a larger proportion of participants in an earlier stage of MCI *within* CDR = 0.5 than that in the V2 cohort; this was confirmed by the fact that tests common to both batteries showed better ability to discriminate between participants with CDR 0 versus CDR 0.5 in the V2 cohort than in the V3 cohort ($P < 0.01$). Participants with CDR = 0.5 can be heterogeneous, covering a wide range of cognitive functions from early to late MCI. Conversion rates from MCI to AD, as well as longitudinal trajectories of cognitive and functional outcomes, are expected to be sensitive to this within-group variability. Depending on the hypothesis, future studies that combine V2 and V3 cohorts should pay careful attention to the potential impact of cohort differences on their clinical outcomes.

There are several methodological limitations in this study. First, we matched V2 and V3 cohorts using the four tests administered in both batteries. These tests capture only executive and language domains. Although the correlation matrices of those selected versus not-selected from the V2 cohort showed no systematic selection bias, harmonizing two cohorts across domains that are more comprehensive might improve generalizability of this study result. The MoCA and MMSE measure global functions and therefore would be better suited for use as anchors for matching the two cohorts using the previous crosswalk study results.[23] However, the ceiling effect of MMSE (eg, MoCA scores of 27, 28, 29, and 30 are all equivalent with MMSE of 30) did not allow us to use these tests as anchors. Second, although the CDR is supposed to be scored independently from neuropsychological tests, clinicians might use cognitive test scores to guide CDR scoring, especially when only limited information is available (eg, no collateral information). Third, NACC comprises a multi-site clinical case series, and thus population inferences should not be drawn directly. That is, ROC-AUCs reported in this study might not be generalizable to population-based probability samples. Fourth, the main aim of this article was to compare discriminative abilities across CDR scores between V2 and V3 neuropsychological batteries. Due to the large number of comparisons, we compared global measures only on combined sensitivity and specificity using ROC-AUCs, rather than addressing sensitivity, specificity, and accuracy separately. Depending on the study aim (eg, high specificity is preferred over high sensitivity in order to decrease the chance of following false positive subjects in clinical trials),[26] different cut-points will need to be re-assessed. Fifth, the current study did not confirm etiologies. Future research should evaluate the sensitivity of V3 to preclinical and prodromal AD and other types of dementia. Finally, we note that MoCA became non-proprietary in 2019, although its use is authorized to NACC ADCs free of charge until the end of 2025.[27]

In summary, the current study showed that the test battery in V3 improved the abilities in differentiating between CDR 0 and 0.5 in some tests, but the magnitude of these improvements was relatively small and the gain was limited to the Caucasian participants. Due to its wide use among ADCs in the United States, the UDS cognitive test battery or its composites can provide effective tools to enrich clinical trials.[28] However, further research with sufficient numbers of underrepresented groups is required to address racial- and ethnic-specific normative scores and optimal cut-points to improve diagnostic accuracy across diverse groups.[29,30,31–34,35]

ACKNOWLEDGMENTS

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD,

MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD), and P30 AG053760 (PI Henry Paulson, MD).

The Uniform Data Set Neuropsychology Work Group also wishes to thank the following individuals for giving their permission to include tests they created in the UDSNB 3.0: Zaid Nasreddine, M.D. (MoCA); Suzanne Craft, PhD ("Craft Story" and scoring method); Andrew Saykin, PhD for providing data on the equivalence of forms of the Craft Story to Logical Memory; Tamar Gollan, PhD for the Multilingual Naming Test (MINT); Joel Kramer for providing alternate number strings for the Number Span test; and Kate Possin and Joel Kramer for the Benson Complex Figure Test.

CONFLICTS OF INTEREST

Hiroko H Dodge, Felicia C. Goldstein, Nicole I Wakim, Tamar Gefen, Merilee Teylan, Kwun CG Chan, Walter A. Kukull, Lisa L. Barnes, Bruno Giordani, Timothy M. Hughes, Joel H. Kramer, David A. Loewenstein, Daniel C Marson, Dan M Mungas, Bonnie C. Sachs, Monica Willis-Parker, Katherine V Wild, John C. Morris, and Sandra Weintraub have no conflicts of interest to declare. David Salmon is a paid consultant for Aptinyx, Inc. and Biogen, Inc. Kathleen Welsh-Bohmer received funding from Verasci as a contract through Duke University and funding from Takeda over the last 5 years (until May 2019) as part of a contract with Duke.

REFERENCES

1. Beekly DL, Ramos EM, Lee WW, et al. The National Alzheimer's Coordinating Center (NACC) database: the Uniform Data Set. *Alzheimer Dis Assoc Disord*. 2007;21(3):249-258.
2. Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. *Alzheimer Dis Assoc Disord*. 2009;23(2):91-101.
3. 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.
4. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993;43(11):2412-2414.
5. Jack CR Jr, Albert MS, Knopman DS, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):257-262.
6. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia 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):263-269.
7. Albert MS, DeKosky ST, Dickson D, 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):270-279.
8. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of 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):280-292.
9. Beekly DL, Ramos EM, van Belle G, et al. The National Alzheimer's Coordinating Center (NACC) Database: an Alzheimer disease database. *Alzheimer Dis Assoc Disord*. 2004;18(4):270-277.
10. Weintraub S, Besser L, Dodge HH, et al. Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). *Alzheimer Dis Assoc Disord*. 2018;32(1):10-17.
11. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198.
12. Reitan RM. Validity of the Trail-making Tests as an indication of organic brain damage. *Percept Mot Skills*. 1985;8:271-276.
13. Wechsler D. *Wechsler Memory Scale Revised*. San Antonio, TX: The Psychological Corporation; 1987.
14. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989;39(9):1159-1165.
15. Mack WJ, Freed DM, Williams BW, Henderson VW. Boston Naming Test: shortened versions for use in Alzheimer's disease. *J Gerontol*. 1992;47(3):P154-158.
16. Kaplan E, Goodglass H, Weintraub S. *The Boston Naming Test*. Philadelphia: Lea & Febiger; 1983.
17. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.
18. Goldstein FC, Milloy A, Loring DW, for the Alzheimer's Disease Neuroimaging I. Incremental validity of Montreal cognitive assessment index scores in mild cognitive impairment and Alzheimer disease. *Dement Geriatr Cogn Disord*. 2018;45(1-2):49-55.
19. Craft S, Newcomer J, Kanne S, et al. Memory improvement following induced hyperinsulinemia in Alzheimer's disease. *Neurobiol Aging*. 1996;17(1):123-130.
20. Ivanova I, Salmon DP, Gollan TH. The multilingual naming test in Alzheimer's disease: clues to the origin of naming impairments. *J Int Neuropsychol Soc*. 2013;19(3):272-283.
21. Possin KL, Laluz VR, Alcantar OZ, Miller BL, Kramer JH. Distinct neuroanatomical substrates and cognitive mechanisms of figure copy performance in Alzheimer's disease and behavioral variant frontotemporal dementia. *Neuropsychologia*. 2011;49(1):43-48.
22. Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. New York: Oxford University Press; 2012.
23. Monsell SE, Dodge HH, Zhou XH, et al. Results From the NACC uniform data set neuropsychological battery crosswalk study. *Alzheimer Dis Assoc Disord*. 2015;30(2):134-139.
24. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med*. 2000;19(9):1141-1164.
25. Julayanont P, Brousseau M, Chertkow H, Phillips N, Nasreddine ZS. Montreal Cognitive Assessment Memory Index Score (MoCA-MIS) as a predictor of conversion from mild cognitive impairment to Alzheimer's disease. *J Am Geriatr Soc*. 2014;62(4):679-684.
26. Breitner JCS. How can we really improve screening methods for AD prevention trials? *Alzheimers Dement (N Y)*. 2016;2(1):45-47.
27. Borson S, Sehgal M, Chodosh J. Monetizing the MoCA: what now? *J Am Geriatr Soc*. 2019;67(11):2229-2231.
28. Lin M, Gong P, Yang T, Ye J, Albin RL, Dodge HH. Big data analytical approaches to the NACC dataset: aiding preclinical trial enrichment. *Alzheimer Dis Assoc Disord*. 2018;32(1):18-27.

29. Milani SA, Marsiske M, Cottler LB, Chen X, Striley CW. Optimal cutoffs for the Montreal Cognitive Assessment vary by race and ethnicity. *Alzheimers Dement (Amst)*. 2018;10:773-781.
30. Goldstein FC, Ashley AV, Miller E, Alexeeva O, Zanders L, King V. Validity of the montreal cognitive assessment as a screen for mild cognitive impairment and dementia in African Americans. *J Geriatr Psychiatry Neurol*. 2014;27(3):199-203.
31. Sink KM, Craft S, Smith SC, et al. Montreal Cognitive Assessment and Modified Mini Mental State Examination in African Americans. *J Aging Res*. 2015;2015:872018.
32. Rossetti HC, Lacritz LH, Cullum CM, Weiner MF. Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology*. 2011;77(13):1272-1275.
33. Rossetti HC, Smith EE, Hynan LS, et al. Detection of Mild cognitive impairment among community-dwelling African Americans using the montreal cognitive assessment. *Arch Clin Neuropsychol*. 2019;34(6):809-813.
34. Rossetti HC, Lacritz LH, Hynan LS, Cullum CM, Van Wright A, Weiner MF. Montreal cognitive assessment performance among community-dwelling African Americans. *Arch Clin Neuropsychol*. 2017;32(2):238-244.
35. Sachs BC, Steenland K, Zhao L, et al. Expanded Demographic Norms for Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set. *Alzheimer Dis Assoc Disord*. 2020;34(3):191-197.
36. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. *J Stat Softw*. 2011;42(8):1-28.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Dodge HH, Goldstein FC, Wakim NI, et al. Differentiating among stages of cognitive impairment in aging: Version 3 of the Uniform Data Set (UDS) neuropsychological test battery and MoCA index scores. *Alzheimer's Dement*. 2020;6:e12103. https://doi.org/10.1002/trc2.12103

APPENDIX 1: XXXX

In this section, we describe the steps we took to match the cognitive functions of 2 (V2 and V3) cohorts in order to make a fair comparison of V2 and V3 test batteries with respect to their abilities in differentiating across CDRs. We also provide detailed descriptions of how to generate MoCA Index scores and optimal cut-points. We provide the latter to aid clinicians in determining the cognitive status of the participants.

Sample selection process to match V2 and V3 cohorts

To compare the discriminative ability of the two test batteries, it was desirable to have the same number of subjects in each CDR level from each cohort. In addition, we were concerned that the two cohorts could vary in their cognitive abilities, especially within the CDR = 0.5 group. As discussed in Part 1, in recent years, the ADCs have been more interested in recruiting patients with earlier stages of cognitive decline because current research is focused increasingly on these subjects for clinical trials and primary prevention. Thus there may be a larger proportion of cases at an earlier stage of decline in the V3 cohort's CDR 0.5 group, which could make it more difficult to discriminate between healthy volunteers (CDR = 0) and patients with MCI (CDR = 0.5), even when they are administered the same tests. Our preliminary studies demonstrated that this is the case.

We conducted two preliminary analyses. First, we selected the same number of participants from both V2 and V3 cohorts, using participants from V2 who were recruited close to the date when V2 was first introduced (recruited between September 2005 and February 2011; preliminary analyses [1]). Second, we selected participants from V3 who were recruited close to the date when V3 was introduced (recruited between December 2011 and June 2015; preliminary analyses [2]). In both preliminary analyses 1 and 2, we found that three of four tests common to both batteries (Category Fluency Animals, Trails Making tests A and B) showed better ability to discriminate between participants with CDR 0 versus CDR 0.5 in the V2 cohort than in the V3 cohort ($P < 0.01$). We also saw better test scores in V3 than in V2 in all CDR groups on these three tests, but the improvement was most striking in the CDR 0.5 group. Therefore, to make a fair comparison of the remaining tests in V2 and V3 batteries, we matched the V2 and V3 cohorts using the tests common to both batteries. We used the matchit function in R.[36] We started with V3 participants aged 60 years or older, who were not missing values for race, education, or any of the four tests (Category Fluency Animals and Vegetables, Trails Making Tests A and B) used for matching; this resulted in a sample size of 4318. We then stratified our samples by CDR scores (three variables) and race (two variables, African American vs Caucasian participants) in order to match participants between V2 and V3 cohorts within each combination of these variables (ie, each of $3 \times 2 = 6$ cells). We used nearest neighbor matching based on participants' scores from the four previously mentioned tests. This match method selected the individual in V2 who had the smallest distance from each individual in V3 in terms of the four test scores. Distance was calculated using the "mahalonobis" option, which is similar to Euclidean distance in three dimensions. However, the four tests created four dimensions to measure distance. Once an individual in V2 is matched with one from V3, they are no longer considered a possible match for others in V3. With this matching approach, ROC-AUCs of the same tests used in both versions were no longer different in discriminating across CDR scores. In this article, we report the results in the last matched cohort analysis. In **Appendix Table 1**, we show the demographic characteristics of each of the selected V2 cohorts used in our preliminary analyses 1 and 2, as well as the entire sample and the test scores and AUC-ROCs of the four common tests. In addition, **Appendix Figure 1** shows the correlation matrices of test scores for those who were selected versus not selected from V2. The correlations between the two groups are almost identical, indicating no systematic bias in the selection process.

TABLE A1 (A) Demographic characteristics of selected and entire participants in V2 and the participants in V3 and scores of the common tests in V2 and V3 (the data used for preliminary analyses [1] and [2]). (B) ROC-AUC results under the Preliminary Analyses 1 for tests used in both V2 and V3 cohorts

(A)	Version 2 Entire Sample (N = 11,060) Dates of Assessment: 9/9/2005 – 3/13/2015			Version 2 (A) (N = 4320)* Used for Preliminary Analysis 1 Dates of Assessment: 9/9/2005 – 4/19/2011			Version 2 (B) (N = 4320)* Used for Preliminary Analysis 2 Dates of Assessment: 12/20/2011 – 3/13/2015			Version 3 (N = 4318) Dates of Assessment: 3/16/2015 – 8/8/2018		
	CDR = 0 N = 4960	CDR = 0.5 N = 4684	CDR = 1 N = 1416	CDR = 0 N = 2274	CDR = 0.5 N = 1724	CDR = 1 N = 322	CDR = 0 N = 2274	CDR = 0.5 N = 1724	CDR = 1 N = 322	CDR = 0 N = 2274	CDR = 0.5 N = 1722	CDR = 1 N = 322
Age, y (mean, SD)	71.8 (7.4)	73.5 (7.6)	75.2 (8.0)	71.8 (7.5)	74.1 (7.5)	75.5 (7.7)	71.8 (7.3)	73.1 (7.7)	74.1 (8.0)	71.2 (6.5)	72.6 (7.0)	74.8 (7.6)
Education, y (mean, SD)	15.9 (2.9)	15.4 (3.3)	14.8 (3.5)	15.6 (3.1)	14.9 (3.4)	14.6 (3.6)	16.2 (2.8)	15.6 (3.2)	15.3 (3.2)	16.3 (2.7)	16.0 (2.9)	15.5 (3.1)
Female (%)	65.3	47.7	47.2	66.1	48.5	49.1	64.2	45.9	43.5	65.3	50.6	46.3
Race (%)												
Caucasian	77.5	81.0	84.1	75.1	81.4	82.9	79.3	79.6	86.3	76.4	82.2	88.8
AfricanAmerican	17.0	11.6	9.7	19.1	11.3	11.5	15.5	12.4	7.1	16.2	12.5	6.2
American Indian or AlaskaNative	0.4	1.1	1.0	0.7	1.0	0.6	0.2	1.5	1.6	1.2	0.7	0
Native Hawaiian or PacificIslander	0.1	0.1	0.2	0.1	0.2	0	0.1	0.2	0.3	0.1	0	0.3
Asian	1.8	2.4	1.3	1.1	1.9	1.6	2.6	3.3	0.6	3.5	2.3	3.4
Multiracial	3.2	3.6	3.5	4.0	4.2	3.4	2.3	2.9	4.0	2.6	2.3	1.2
Cognitive Test Scores												
Common tests across two versions												
Animals	20.1 (5.6)	15.7 (5.5)	11.4 (4.8)	19.8 (5.7)	15.3 (5.2)	10.9 (4.3)	20.5 (5.5)	15.9 (5.7)	11.9 (5.0)	20.6 (5.6)	16.5 (5.5)	12.2 (5.1)

(Continues)

TABLE A1 (Continued)

(A)	Version 2 Entire Sample (N = 11,060) Dates of Assessment: 9/9/2005 – 3/13/2015			Version 2 (A) (N = 4320)* Used for Preliminary Analysis 1 Dates of Assessment: 9/9/2005 – 4/19/2011			Version 2 (B) (N = 4320)* Used for Preliminary Analysis 2 Dates of Assessment: 12/20/2011 – 3/13/2015			Version 3 (N = 4318) Dates of Assessment: 3/16/2015 – 8/8/2018		
	CDR = 0 N = 4960	CDR = 0.5 N = 4684	CDR = 1 N = 1416	CDR = 0 N = 2274	CDR = 0.5 N = 1724	CDR = 1 N = 322	CDR = 0 N = 2274	CDR = 0.5 N = 1724	CDR = 1 N = 322	CDR = 0 N = 2274	CDR = 0.5 N = 1722	CDR = 1 N = 322
Vegetables	14.6 (4.3)	10.6 (4.2)	7.3 (3.6)	14.4 (4.3)	10.6 (4.1)	7.4 (3.4)	14.7 (4.2)	10.5 (4.1)	7.5 (3.7)	14.4 (4.1)	10.8 (4.3)	7.3 (3.7)
Trail A	35.1 (15.8)	45.3 (23.6)	64.3 (35.6)	36.8 (18.0)	47.1 (24.8)	65.0 (35.2)	33.3 (13.3)	44.4 (22.7)	61.9 (34.6)	33.7 (14.0)	43.1 (22.7)	59.7 (32.0)
Trail B	95.3 (53.3)	149.7 (83.4)	223.8 (85.4)	100.2 (58.1)	155.9 (84.5)	225.9 (83.2)	90.9 (49.0)	147.8 (82.7)	213.1 (88.7)	90.5 (48.1)	138.2 (80.1)	209.4 (88.8)

(B) ROC-AUC comparisons between V2 cohort (A) and V3 cohort							
Version 2				Version 3			
Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1	Test	CDR 0 vs 0.5	CDR 0 vs 1	CDR 0.5 vs 1
ANIMALS	0.767**	0.923	0.764	ANIMALS	0.734	0.920	0.777
Category Fluency (animals)	(0.753, 0.782)	(0.911, 0.935)	(0.742, 0.786)	Category Fluency (animals)	(0.719, 0.749)	(0.907, 0.932)	(0.755, 0.798)
VEG	0.765	0.919	0.755	VEG	0.745	0.932	0.780
Category Fluency (vegetables)	(0.75, 0.779)	(0.907, 0.932)	(0.733, 0.778)	Category Fluency (vegetables)	(0.73, 0.76)	(0.92, 0.944)	(0.759, 0.801)
TRAILA	0.716**	0.860	0.711	TRAILA	0.686	0.878	0.749
Trail A Lines per second	(0.701, 0.733)	(0.84, 0.877)	(0.684, 0.735)	Trail A Lines per second	(0.667, 0.7)	(0.859, 0.892)	(0.723, 0.769)
TRAILB	0.754**	0.908	0.735	TRAILB	0.719	0.892	0.742
Trail Making Test Part B	(0.739, 0.77)	(0.892, 0.925)	(0.708, 0.762)	Trail Making Test Part B	(0.703, 0.735)	(0.872, 0.912)	(0.714, 0.771)

**The same tests used in both batteries (Category Fluency Animals, Trails A and B) showed statistically significant better performance in differentiating CDR of 0 and 0.5 when used among V2 cohort, indicating the potential difference in cognitive abilities among those with CDR of 0 and 0.5 between the two cohorts.

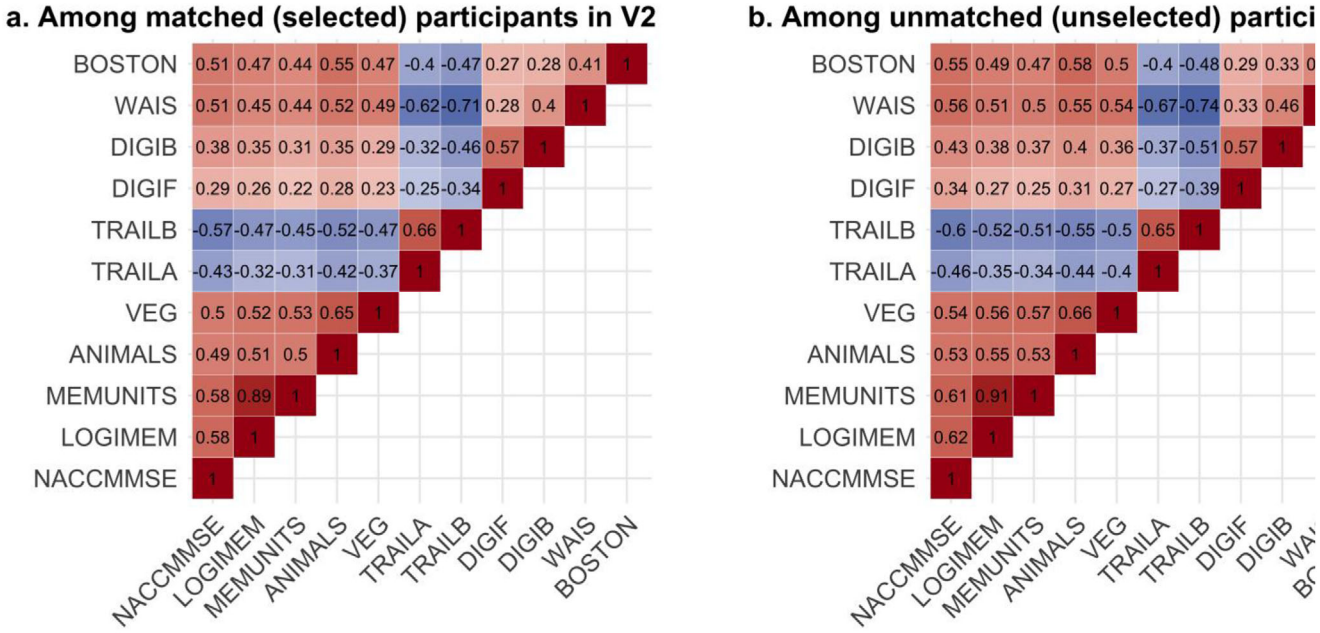

FIGURE A1 Correlation metrics of cognitive test scores among those who were selected (matched with V3) versus those who were not selected

MoCA domain-specific index scores

We generated MoCA domain-specific index scores using an algorithm described previously.[17,18,25] The Memory Index Score consisted of the number of words recalled in delayed-free, category-cued, and multiple-choice conditions, multiplied by 3, 2, and 1, respectively (0-15 points). The Executive Index Score included Trail-Making, Clock Drawing, Digit Span, Letter Tapping, Serial 7 Subtraction, Letter Fluency, and Abstraction (0-13 points). The Visuospatial Index Score consisted of Cube Copy, Clock Drawing, and Naming (0-7 points). The Language Index Score included Naming, Sentence Repetition and Letter Fluency (0-6 points). The Attention Index Score comprised Digit Span, Letter Tapping, Serial 7 Subtraction, Sentence Repetition, and Words Recalled in both Registration Trials (0-18 points). The Orientation Index Score included all Orientation items (0-6 points). The rules for calculating MoCA Index scores are also summarized in **Appendix Table 2**.

Optimal cut-points

To aid clinicians in providing a diagnosis based on NACC UDS V3 test scores, we also provided optimal cut-points for each individual test and composite score **in the Supplemental section**. These optimal cut-points were determined through the largest Youden's J value in the current paper. For each point on the ROC curve, Youden's J is defined as one subtracted from the sum of the specificity and sensitivity. The maximum Youden's is then mapped back to its original measure, resulting in the optimal cut-point to differentiate the two CDR groups. SAS Version 9.3 and R Version 3.4.2 (packages: base, pROC, dplyr) were used for the analysis.

TABLE A2 MoCA index score calculation table

RULES FOR SINGLE ITEM SCORING AND INDEX SCORE CALCULATIONS FOR THE MOCA[1,2]							
		MOCA INDEX DOMAINS					
MOCA ITEMS in the order of testing/scoring	Points Toward Total Score	Memory[3]	Executive	Attention/ Concentration	Language	Visuospatial	Orientation
Trail Making Task	1	–	1	–	–		–
Cube Copy	1	–	–	–	–	1	–
Clock Circle	1	–	1	–	–	1	–
Clock Hands	1	–	1	–	–	1	–
Clock Time	1	–	1	–	–	1	–
Language: Naming 3 objects	3	–	–	–	3	3	–
Memory: Registration (2 learning trials, total possible = 10)[3]	0	–	–	Immediate recall 2 trials total/10	–	–	–
Attention Digits	2	–	2	2	–	–	–
Attention Letter A	1	–	1	1	–	–	–
Att: serial 7′s	3	–	3	3	–	–	–
Lang: Repetition	2	–	–	2	2	–	–
Lang: Fluency	1	–	1	–	1	–	–
Abstraction	2	–	2	–	–	–	–
Delayed Recall with no cue[3]	5	=3X number words recalled free (max = 15)	–	–	–	–	–
Delayed Recall Category cue[3]	0	=2X number of words retrieved with category cue (max = 10)	–	–	–	–	–
Delayed Recall Recognition[3]	0	=1X number of words recognized (max 5)	–	–	–	–	–
Orient: Date	1	–	–	–	–	–	1
Orient: Month	1	–	–	–	–	–	1
Orient: Year	1	–	–	–	–	–	1
Orient: Day	1	–	–	–	–	–	1
Orient: Place (Name)	1	–	–	–	–	–	1
Orient: City	1	–	–	–	–	–	1
TOTALS	30	15	13	18	6	7	6

[1]© Ziad Nasreddine MD 2004. MoCA is a registered trademark property of Neurosearch Développements Inc. and is used under license. Form created as part of the National Alzheimer's Coordinating Center Uniform Data Set, copyright 2013 University of Washington.

[2]Julayanont P, Brousseau M, CHertkow H, Phillips N, Nasreddine, ZS Montreal Cognitive Assessment Memory Index Score (MoCA-MIS) as a predictor of conversion from mild cognitive impairment to Alzheimer's disease. Journal of the American Geriatrics Society 62: 679-684, 2014.

[3]The standard administration of the MoCA does not score category and recognition responses, even if administered. To maximize the yield from this test item, the following strategy was adopted: If all five words are freely recalled, cued and category recall are not administered and the total score would be 15 (3 points for each word recalled freely). After free recall, category cues are given *only for items not recalled*. Each word correct with category cue is awarded 2 points. After category cues, only words not correct are then tested with recognition. Award 1 point for each word correct on recognition (max = 5). The following is an example: individuals gets two items on free recall, two items on the three items cued, and recognizes the 5th item on multiple choice. Memory Index Score: $(2 \times 3) + (2 \times 2) + (1 \times 1) = 11/15$.