Expanded Demographic Norms for Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set

Bonnie C. Sachs, PhD,† Kyle Steenland, PhD,‡ Liping Zhao, MSPH,§ Timothy M. Hughes, PhD,† Sandra Weintraub, PhD,‖ Hiroko H. Dodge, PhD,¶ Lisa L. Barnes, PhD,# Suzanne Craft, PhD,† Monica L. Parker, MD,** and Felicia C. Goldstein, PhD***

Background: Norms for the Uniform Data Set Version 3 Neuropsychological Battery are available for cognitively normal individuals based on age, education, and sex; however, these norms do not include race. We provide expanded norms for African Americans and whites.

Methods: Data from 32 Alzheimer's Disease Centers (ADCs) and ADC affiliated cohorts with global Clinical Dementia Rating Scale (CDR) Dementia Staging Instrument scores of 0 were included. Descriptive statistics for each test were calculated by age, sex, race, and education. Multiple linear regressions were conducted to estimate the effect of each demographic variable; squared semipartial correlation coefficients measured the relative importance of variables.

Results: There were 8313 participants (16% African American) with complete demographic information, ranging from 6600 to 7885 depending on the test. Lower scores were found for older and less educated groups, and African Americans versus whites. Education was the strongest predictor for most tests, followed in order by age, race, and sex. Quadratic terms were significant for age and education, indicating some nonlinearity, but did not substantially increase R^2.

Conclusions: Although race-based norms represent incomplete proxies for other sociocultural variables, the appropriate application of these norms is important given the potential to improve diagnostic accuracy and to reduce misclassification bias in cognitive disorders of aging such as Alzheimer disease.

Key Words: uniform data set, National Alzheimer Coordinating Center, cognitive test norms, neuropsychology

(*Alzheimer Dis Assoc Disord* 2020;34:191–197)

The Uniform Data Set Version 3 Neuropsychological Battery (UDSNB 3.0) of the Alzheimer's Disease Centers (ADCs) was introduced in 2015 to replace some of the measures used in earlier versions to evaluate cognition, as well as to expand domains that were not already being assessed. The Neuropsychology Work Group of the National Institutes of Health-National Institute on Aging Clinical Task Force was charged with selecting measures that would reduce practice effects and maintain continuity with domains for which there was longitudinal data spanning a decade. In addition, there was a need to use existing nonproprietary measures or develop new ones to allow the sharing of data with nonparticipating centers. These alternative measures, described in detail in Weintraub et al,[1] were chosen to assess

Received for publication February 24, 2020; accepted April 19, 2020.

From the *Department of Neurology; †Section on Gerontology and Geriatric Medicine, Department of Internal Medicine, Wake Forest Alzheimer's Disease Research Center, Wake Forest School of Medicine, Winston-Salem, NC; ‡Department of Environmental and Occupational Health; §Department of Biostatistics and Bioinformatics, School of Public Health; **Department of Neurology, School of Medicine, Goizueta Alzheimer's Disease Research Center, Emory University, Atlanta, GA; ‖Departments of Psychiatry, Neurology and Alzheimer's Disease Center, Northwestern University Feinberg School of Medicine; #Departments of Neurological Sciences and Behavioral Sciences, Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; and ¶Department of Neurology, Layton Aging, Alzheimer's Disease Center, Oregon Health and Science University, Portland, OR.

Supported by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). This work was directly supported by P30AG049638, R01AG054069 and R01AG058969. A full list of participating MESA investigators and institutions can be found at www.mesa-nhlbi.org.

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), P30 AG062428-01 (PI James Leverenz, 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), P30 AG062421-01 (PI Bradley Hyman, MD, PhD), P30 AG062422-01 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), 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), P30 AG062429-01(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), P30 AG062715-01 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

The authors declare no conflicts of interest.

Reprints: Felicia C. Goldstein, PhD, Department of Neurology, Emory University, 12 Executive Park Drive NE, Atlanta, GA 30329 (e-mail: fgoldst@emory.edu).

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.alzheimerjournal.com.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

global cognitive status (Montreal Cognitive Assessment[2]) as well as specific domains including attention/working memory (Number Span), verbal episodic memory (Craft Story[3]), and language (Multilingual Naming Test[4]). In addition, new measures were added to evaluate nonpreviously assessed areas of visuoconstructional ability (Benson Complex Figure Copy[5]), visual episodic memory (Benson Complex Figure—Immediate and Delayed Recall[5]), and timed phonemic fluency (Letters "F" and "L"). A Crosswalk study was subsequently conducted to determine the degree of harmonization between the 2 batteries, and the results indicated good correlations between the earlier and current measures.[6]

Published norms for UDSNB 3.0 are available for 3602 cognitively normal participants tested between March 2015 and November 2016,[1] and an online calculator is available on the National Alzheimer's Coordinating Center (NACC) website that can be used to derive regression based normative scores based on an individual's age, education, and sex (www.alz.washington.edu). However, the norms did not include a breakdown of performance by race due to the small sample of African Americans represented at that time in the database. Whereas 2990 participants in the normative sample were white, only 504 were African American. As such, norms for a combined sample, albeit largely white, were provided. Weintraub et al[1] recommended expansion of the norms in the future to include under-represented groups. Since the clinician diagnoses of normal cognition, mild cognitive impairment, and dementia in the NACC database are determined, in large part, via the results of this cognitive battery and the associated norms, it is crucial to establish norms based on complete demographic information. Although some investigators have noted the importance of developing norms that include race to avoid misdiagnosis, others have emphasized that "race" is only a proxy for individual differences in literacy, cultural exposures and quality of education on performance.[7–11] The source of the differences notwithstanding, in the absence of full demographic normative data including age, education, and race, individuals may be misclassified as impaired; this has been the case in numerous studies using the Montreal Cognitive Assessment (MoCA), which is part of the current battery.[12–14] Thus, norms that account for multiple demographic factors including race have become routine. Demographic norms have been adopted for some of the most commonly administered neuropsychological tests, including the Boston Naming Test (BNT), Dementia Rating Scale-2nd Edition, Digit Symbol Substitution Task, Trail Making Test (TMT), Rey Auditory Verbal Learning Test (RAVLT), the Hopkins Verbal Learning Test-Revised, the Brief Visuospatial Memory Test-Revised, and the California Verbal Learning Test.[15–21] Current demographically adjusted normative data sets include the Mayo Older African Americans Normative Studies,[17,18,21] Halstead-Reitan Battery norms (Heaton norms),[16] and the Wechsler Adult Intelligence and Memory Scales[22,23] (Advanced Clinical Solutions).[24]

In the current study, we provide expanded norms for UDSNB 3.0 that include race. The larger sample allows for more stable norms as well as an examination of the relative contributions of all 4 demographic variables to test performance.

METHODS

Participants

The study participants were enrolled in the National Institutes of Health-National Institute on Aging supported ADRCs, a nationwide consortium of academic research sites (www.alz.washington.edu). We used the information available from 32 ADCs from March 2015 through May 2019. Written consent was obtained for all participants using forms approved by the institutional review boards at each site. Participants included persons with at least one UDSNB 3.0 test, and a global Clinical Dementia Rating Scale (CDR)[25] score of 0. We used the test scores from each participant's first administration of UDSNB 3.0 to avoid practice effects from repeated administrations. We did not exclude individuals who had previously received UDSNB 2.0. Similar to Weintraub et al,[1] we did not exclude participants based on their test scores, even if they fell in the "impaired" range (eg, MoCA < 20 points) to avoid any prejudgment of test values. In addition, participants were not required to have a diagnosis of "normal cognition" since this determination is based on interpretation of the cognitive test scores using currently available normative data.

To maximize the number of African American individuals in all age and education ranges, we also included individuals in a selected ADC affiliated cohort, namely, the Multi-Ethnic Study of Atherosclerosis (MESA).[26] MESA is a diverse, multisite, longitudinal observational study of older adults from 6 sites in the United States (https://mesa-nhlbi.org) that began implementing the UDSNB 3.0 in 2016. This analysis includes MESA participants who were evaluated at the Wake Forest ADRC (50% African American) aged 60 to 93, with a CDR global score of 0. The decision to limit inclusion to African Americans enrolled in MESA at the Wake Forest site was because only Wake Forest MESA participants had received the UDSNB 3.0 at the time of these analyses. All UDS version 3 components (questionnaires and UDSNB 3.0) were administered in standardized fashion, using ADRC-trained and certified staff, and were reviewed and adjudicated with all other Wake Forest Clinical Core ADRC participants by the same adjudication committee.

Measures

A full description of the UDSNB 3.0 neuropsychological measures may be found in Weintraub et al.[1] Briefly, the MoCA consists of items evaluating memory, language, visuomotor ability, attention, and executive functioning, yielding a total score of 30 points.[2] The Craft Story (Version 21)[3] evaluates verbal episodic memory by reading a paragraph out loud to the study participant, and separately scoring the number of verbatim units and the paraphrased units recalled immediately and following a 30 minute delay. The Multilingual Naming Test[4] requires oral naming of line drawings. The Benson Complex Figure[5] assesses visuoconstructional ability by having the participant copy a complex geometric figure, followed by recall of the design after 10 to 15 minutes. Number Span involves reading a string of numbers out loud to the study participant and asking for verbatim recall (Forward Condition) or reversal of the string of numbers (Backward Condition). Phonemic and semantic fluency are assessed by having the participant generate words beginning with a specified letter (Letters "F" and "L") and a category (Animals and Vegetables), each in 1 minute. We also calculated norms for Trail Making Parts A and B[27] which, while not new measures for UDSNB 3.0, allowed for co-norming of this test with the rest of the battery.

Statistical Analysis

Descriptive statistics were tabulated by age, education, sex, and race. Consistent with procedures used by Weintraub et al,[1]

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

age was divided into 4 groups (below 60, 60 to 69, 70 to 79, 80 y and above), as was education (≤ 12, 13 to 15, 16, ≥ 17 y). The mean and SD of each test were calculated by age, sex, race, and education groups. Multiple linear regression models were run with all 4 demographics included to estimate the effect of age (continuous), education (continuous, years of formal schooling), and sex and race (both categorical) on each neuropsychological measure. To determine the relative importance of the 4 variables for each test, we calculated partial correlation coefficients based on squared semipartial correlation which measured incremental value in R^2.

To detect a curvilinear relationship, in Supplemental Analyses (Supplemental Digital Content 1, http://links.lww.com/WAD/A273) we added quadratic terms of centered age and education, and compared R^2 and residual plots between linear and quadratic regression models. Quadratic terms were often significant at the 0.05 level due to the large sample size, but they did not markedly improve R^2. We present as our main results the linear term without the quadratic one to be consistent with the previous publication of norms by Weintraub et al[1] but present the fuller quadratic model in Supplementary Tables (Supplemental Digital Content 1, http://links.lww.com/WAD/A273).

To examine the residual distribution and evaluate the fit of linear regression models, histograms of residuals for each of the tests were plotted. Because the distributions of the copy condition of the Benson Complex Figure, MINT, and Trail Making Part A and Part B were highly skewed (calculation for Trail Making as the number of correct answers divided by seconds avoided this problem, but this metric is not commonly used by clinicians), we applied nonparametric median regression to estimate the conditional median of the cognitive measures as a sensitivity analysis. Goodness-of-fit statistics "R1" were calculated. Median regression is more robust to outliers than least squares regression and makes no assumptions about the distribution of the residuals. It should be noted that the regression coefficients from linear regressions are unbiased even if the residuals are not normal, but hypotheses tests are biased.

Statistical analyses were performed using SAS 9.4 (SAS Institute Inc.) or R software (www.R-project.org).

RESULTS

There were 8313 participants (8077 from NACC and 236 from MESA) with complete information on age, education, sex, and race, of whom 16.4% of the participants were African American (16% in NACC, 47% in MESA). In our analyses by race, subjects from NACC were either classified as African American or white, which resulted in the exclusion of 6% of the subjects with CDR = 0 who were in another racial group (American Indian, native Hawaiian, Asian, other). Sensitivity analyses revealed no significant differences in test scores between MESA and NACC participants after controlling for age, education, sex, and race (data not shown). Table 1 shows a breakdown of the sample sizes by age, education sex, and race. Sample sizes for some groups of African Americans in the youngest and oldest age groups were sometimes small; overall 6 of 32 age/education/race/sex groups had <10 subjects, and all but one of these groups was comprised of African American men who were below 60 years old or 80 years and above and had education levels > 12 years.

Table 2 shows the means, SD, medians, 25th and 75th percentiles, and ranges of the cognitive scores for the entire sample. The sample sizes ranged from 6600 to 7885, depending on the specific test. Tables 3 and 4 show the data by age and education. It is clear from Tables 3 and 4 that there are marked differences in scores, with lower scores for older versus younger groups and for less educated versus more educated groups. Table 5 shows linear regression results for scores regressed on sex, age (continuous), education (continuous, self-reported years of schooling), and race. Model R^2 values were modest, ranging from 0.06 to 0.26, indicating the importance of other variables besides demographic ones for predicting test results. Semipartial correlation coefficients for different tests indicate that overall education is the strongest predictor (top factor on 12/20 tests), while sex (top factor on Vegetable naming) is the weakest, with age (top factor on 7 tests and second on 6 tests) and race (top factor on the MINT and second on 11 tests) intermediate between the 2.

Supplemental Table 1 (Supplemental Digital Content 1, http://links.lww.com/WAD/A273) also shows marked differences between African Americans versus whites. Supplemental Table 2 (Supplemental Digital Content 1, http://links.lww.com/WAD/A273) shows the R^2 of models with and without race, indicating those tests for which the race adjustment is most important. Supplemental Figure 1 (Supplemental Digital Content 1, http://links.lww.com/WAD/A273) shows the distribution of the residuals for the linear regressions in Table 5. Although generally conforming to a normal distribution, the distributions for 4 tests (Benson Complex Figure copy, total score MINT, Trail Making Tests A and B) are clearly not normal. For these we ran quantile regression for the median, with results shown in Supplemental Table 3 (Supplemental Digital Content 1, http://links.lww.

TABLE 1. Sample Distribution by Race, Sex, Education, and Age

Age (y)	Education	Sex	White	African American	Total
< 60	≤ 12	Female	52	4	56
		Male	24	8	32
	13-15	Female	81	15	96
		Male	40	10	50
	16	Female	142	19	161
		Male	82	5	87
	≥ 17	Female	157	19	176
		Male	95	4	99
60-69	≤ 12	Female	102	63	165
		Male	48	28	76
	13-15	Female	239	96	335
		Male	88	26	114
	16	Female	324	77	401
		Male	211	33	244
	≥ 17	Female	586	105	691
		Male	342	31	373
70-79	≤ 12	Female	225	103	328
		Male	90	23	113
	13-15	Female	310	114	424
		Male	115	27	142
	16	Female	424	68	492
		Male	231	33	264
	≥ 17	Female	746	133	879
		Male	580	32	612
≥ 80	≤ 12	Female	167	84	251
		Male	55	20	75
	13-15	Female	201	64	265
		Male	72	9	81
	16	Female	248	22	270
		Male	169	10	179
	≥ 17	Female	370	69	439
		Male	334	9	343

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 2. Summary Statistics for Cognitively Normal UDSNB 3.0 Participants

UDS Version 3 Neuropsychological Test*	Domain	Maximum Score	Sample's Scores			
			N	Mean (SD)	Q25, Q50, Q75	Range
MoCA, total score	Dementia severity/Global Cognition	30	6600	26.0 (2.9)	24, 27, 28	13-30
Craft Story 21 recall immed. verbatim, total units	Memory	44	6607	21.6 (6.7)	17, 22, 26	0-41
Craft Story 21 recall immed. paraphrase, total units	Memory	25	6607	16.0 (4.0)	13, 16, 19	0-25
Craft Story 21 recall delay. verbatim, total units	Memory	44	6604	18.8 (6.7)	14, 19, 24	0-40
Craft Story 21 recall delay. paraphrase, total units	Memory	25	6604	14.9 (4.3)	12, 15, 18	0-25
Benson complex figure copy, total score	Visuospatial	17	7593	15.5 (1.4)	15, 16, 17	9-17
Benson complex figure recall, total score	Visuospatial/ memory	17	7590	11.1 (3.1)	9, 11, 13	0-17
Number span test forward, total correct trials	Attention	14	6622	8.2 (2.3)	6, 8, 10	0-14
Number span test forward, longest span	Attention	9	6621	6.6 (1.3)	6, 7, 8	0-9
Number span test backward, total correct trials	Attention	14	6617	7.0 (2.3)	6, 7, 8	0-14
Number span test backward, longest span	Attention	8	6617	5.0 (1.3)	4, 5, 6	0-8
MINT, total score	Lang. naming	32	6560	29.9 (2.3)	29, 30, 32	19-32
Phonemic test, F-words total in 60 s	Lang. verbal fluency	40	7647	14.6 (4.7)	11, 14, 18	0-35
Phonemic test, L words total in 60 s	Lang. verbal fluency	40	7619	13.7 (4.5)	11, 14, 17	0-35
Phonemic test, total F- and L-words	Lang. verbal fluency	80	7380	28.3 (8.6)	22, 28, 34	1-64
Animals list generation, total in 60 s	Lang. category fluency	77	7885	20.9 (5.7)	17, 21, 25	0-49
Vegetables list generation, total in 60 s	Lang. category fluency	77	7869	14.7 (4.3)	12, 15, 17	0-36
Trail making test part A, time (s)	Processing speed	150	7792	31.8 (12.2)	23, 29, 38	9-108
Trail making test part B, time (s)	Executive function	300	7475	86.7 (51.1)	55, 71, 102	13-300

*Higher scores indicate better scores except for the Trail Making test parts A and B.
Delay indicates delayed; Immed, immediate; Lang, language; MINT, multilingual naming test; MoCA, Montreal Cognitive Assessment; UDS, Uniform Data Set; UDSNB 3.0, Uniform Data Set Version 3 Neuropsychological Battery.

com/WAD/A273); model R^2 for these regressions with all 4 variables and for each one singly are found in Supplemental Table 4 (Supplemental Digital Content 1, http://links.lww.com/WAD/A273). Supplemental Table 5 (Supplemental Digital Content 1, http://links.lww.com/WAD/A273) shows the results when adding quadratic terms for age and education, which were almost always statistically significant mostly due to the large sample size. The quadratic effects indicate that the negative effects of increasing age, and the beneficial effects of more education, are not linear. The R^2 of the models improved with the addition of quadratic terms, but not dramatically.

TABLE 3. Mean Neuropsychological Test Scores by Age Group

UDS Version 3 Neuropsychological Test*	Mean (SD)			
	< 60 y	60-69 y	70-79 y	≥ 80 y
MoCA, total score	27.2 (2.3)	26.5 (2.7)	25.9 (2.8)	24.7 (3.3)
Craft Story 21 recall immed. verbatim, total units	22.5 (6.7)	22.6 (6.6)	21.7 (6.5)	19.3 (6.6)
Craft Story 21 recall immed. paraphrase, total units	16.5 (4.1)	16.6 (3.8)	16.0 (3.9)	14.5 (4.2)
Craft Story 21 recall delay. verbatim, total units	20.2 (6.8)	19.7 (6.5)	18.8 (6.6)	16.2 (6.6)
Craft Story 21 recall delay. paraphrase, total units	15.8 (4.4)	15.5 (4.1)	14.9 (4.2)	13.1 (4.5)
Benson complex figure copy, total score	15.7 (1.2)	15.6 (1.3)	15.5 (1.3)	15.3 (1.5)
Benson complex figure recall, total score	12.6 (2.6)	11.7 (2.8)	10.9 (3.0)	9.8 (3.3)
Number span test forward, total correct trials	8.7 (2.4)	8.3 (2.4)	8.2 (2.3)	7.9 (2.2)
Number span test forward, longest span	6.8 (1.3)	6.7 (1.3)	6.6 (1.3)	6.5 (1.3)
Number span test backward, total correct trials	7.7 (2.4)	7.2 (2.3)	7.0 (2.2)	6.5 (2.2)
Number span test backward, longest span	5.4 (1.3)	5.1 (1.3)	5.0 (1.3)	4.8 (1.3)
MINT, total score	29.9 (2.1)	30.2 (2.1)	29.9 (2.2)	29.3 (2.7)
Phonemic test, F-words total in 60 s	15.7 (4.5)	14.9 (4.7)	14.4 (4.7)	13.8 (4.8)
Phonemic test, L words total in 60 s	14.9 (4.4)	14.2 (4.5)	13.6 (4.4)	13.0 (4.5)
Phonemic test, total F- and L-words	30.6 (8.2)	29.2 (8.5)	28.0 (8.5)	26.9 (8.7)
Animals list generation, total in 60 s	23.5 (5.5)	22.0 (5.8)	20.7 (5.4)	18.6 (5.5)
Vegetables list generation, total in 60 s	15.8 (4.1)	15.4 (4.3)	14.7 (4.2)	13.3 (4.2)
Trail making test part A, time (s)	23.6 (9.4)	29.0 (10.8)	32.4 (11.2)	38.1 (13.6)
Trail making test part A, correct lines/time (s)	1.1 (0.4)	0.9 (0.3)	0.8 (0.3)	0.7 (0.2)
Trail making test part B, time (s)	59.8 (33.8)	76.1 (42.1)	88.3 (48.6)	111 (62.1)
Trail making test part B, correct lines/time (s)	0.5 (0.2)	0.4 (0.1)	0.3 (0.1)	0.3 (0.1)

*Higher scores indicate better scores except for the Trail Making test parts A and B.
Delay indicates delayed; Immed, immediate; MINT, multilingual naming test; MoCA, Montreal Cognitive Assessment; UDS, Uniform Data Set.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.
Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 4. Mean Neuropsychological Test Scores by Education Group

	Mean (SD)			
UDS Version 3 Neuropsychological Test*	**≤ 12 y**	**13-15 y**	**16 y**	**≥ 17 y**
MoCA, total score	23.7 (3.7)	25.3 (3.0)	26.4 (2.5)	26.8 (2.4)
Craft Story 21 recall immed. verbatim, total units	19.0 (7.0)	20.7 (6.5)	21.7 (6.5)	22.7 (6.5)
Craft Story 21 recall immed. paraphrase, total units	14.1 (4.5)	15.4 (3.9)	16.1 (4.0)	16.7 (3.8)
Craft Story 21 recall delay. verbatim, total units	16.2 (7.0)	17.8 (6.6)	18.9 (6.6)	19.8 (6.5)
Craft Story 21 recall delay. paraphrase, total units	12.9 (4.8)	14.1 (4.2)	15.0 (4.2)	15.6 (4.0)
Benson complex figure copy, total score	15.0 (1.6)	15.3 (1.4)	15.5 (1.3)	15.7 (1.2)
Benson complex figure recall, total score	10.1 (3.4)	10.8 (3.1)	11.2 (3.0)	11.4 (3.0)
Number span test forward, total correct trials	7.3 (2.3)	7.8 (2.2)	8.3 (2.3)	8.6 (2.4)
Number span test forward, longest span	6.1 (1.3)	6.4 (1.3)	6.7 (1.3)	6.8 (1.3)
Number span test backward, total correct trials	6.0 (2.2)	6.6 (2.1)	7.2 (2.2)	7.4 (2.2)
Number span test backward, longest span	4.4 (1.3)	4.8 (1.3)	5.1 (1.3)	5.2 (1.3)
MINT, total score	28.5 (2.7)	29.3 (2.4)	30.1 (2.0)	30.4 (2.0)
Phonemic test, F-words total in 60 s	12.4 (4.7)	13.5 (4.5)	14.6 (4.6)	15.6 (4.6)
Phonemic test, L words total in 60 s	11.6 (4.5)	12.7 (4.3)	13.7 (4.3)	14.8 (4.3)
Phonemic test, total F- and L-words	24.2 (8.6)	26.2 (8.3)	28.3 (8.2)	30.4 (8.3)
Animals list generation, total in 60 s	17.9 (5.2)	19.6 (5.4)	21.1 (5.4)	22.3 (5.8)
Vegetables list generation, total in 60 s	13.3 (3.9)	14.3 (4.1)	14.7 (4.2)	15.3 (4.5)
Trail making test part A, time (s)	37.1 (14.9)	33.3 (12.4)	31.0 (11.8)	30.0 (10.8)
Trail making test part A, correct lines/time (s)	0.7 (0.3)	0.8 (0.3)	0.9 (0.3)	0.9 (0.3)
Trail making test part B, time (s)	120 (73.9)	95.6 (55.8)	82.4 (45.3)	75.4 (37.1)
Trail making test part B, correct lines/time (s)	0.3 (0.1)	0.3 (0.1)	0.4 (0.1)	0.4 (0.1)

*Higher scores indicate better scores except for the trail making test parts A and B.
Delay indicates delayed; Immed, immediate; MINT, multilingual naming test; MoCA, Montreal Cognitive Assessment; UDS, Uniform Data Set.

Supplemental File 1 (Supplemental Digital Content 1, http://links.lww.com/WAD/A273) gives the means and SDs of all tests for each demographic subgroup, enabling clinicians to use these data for norms.

DISCUSSION

The Uniform Data Set Version 3 introduced by NACC in 2015 standardized collection of various data elements across ADCs including demographic, medical history, clinical, and neuropsychological data. The UDSNB 3.0 contains almost all new cognitive tests when compared with Versions 1 and 2. Preliminary normative data have been provided for these tests, but full demographically corrected data have been lacking. In this analysis of cognitively and functionally unimpaired individuals based on CDR scores of 0 in the NACC database, we provide stratified normative data and regression based norms accounting for variation due to age, sex, education, and race for racial/ethnic groups who are most often included in research in the United States (white and African American). The results of normative tables and regression based analyses demonstrate that educational attainment has the greatest impact on individual test performance, followed by age, race, and sex. Although race, and to some extent education-based norms are controversial as they represent incomplete proxies for a host of important variables such as socioeconomic status, educational quality, experiences of discrimination, and acculturation, the effect of race on UDSNB 3.0 test performance is important, given the potential to improve diagnostic accuracy and to reduce misclassification bias.

These findings point to a modest amount of residual sociocultural bias by observed racial differences in UDSNB 3.0 tests. Currently, there are adequate NACC UDSNB 3.0 data on African American and white participants to create race-specific normative data as a first step to reduce this bias; however, more UDSNB 3.0 data from diverse populations in the United States are required to further attenuate residual bias in normative data. As shown herein, additional recruitment of under-represented groups (eg, lower educational attainment, oldest old, racial/ethnic minority groups, geographic diversity) of the United States is needed. Therefore, we propose that the creation of normative reference data for standardized batteries such as the UDSNB 3.0 be an iterative and dynamic process that represents the recruitment and research priorities of the National Institutes of Health, ADRCs, and the NACC to include under-represented groups in ADRD research. This work also underscores the need for future research to determine the social determinants of performance on the UDSNB 3.0 tests to minimize bias in the adjudication of cognitive performance in nonhomogenous populations.

Although this paper provides an important update to the available normative data for the UDSNB 3.0, there are limitations of the current project. First, as noted by Manly and colleagues,[9,28,29] race and self-reported years of education are proxies for a myriad of other sociocultural and sociodemographic variables including quality of education, SES, immigrant status, and area deprivation. It is possible that these factors, more so than the crude "race" and "education" variables, play an important role in determining cognitive test performance. Unfortunately, those variables were not available to us as part of these analyses. Second, as is the case with other cohort studies, sample sizes are smaller for groups with lower education, especially those older than 80 years, although the patterns are not as consistent for African Americans. For groups with small sample size, our estimates of means are less precise and potentially therefore less accurate. Third, we attempted to replicate methods used in the original normative study by Weintraub et al[1] by classifying participants as cognitively normal who had a CDR of 0 and not excluding persons with suspected low scores on the tests

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 5. Multivariable Linear Regression Coefficients and 95% CIs for Sex, Age, Education and Race

	Coefficient (95% CI)					
UDS Version 3 Neuropsychological Test*	**Female**	**Age (10 y)†**	**Education (y)**	**African American**	R^2	**Order of Variable Importance‡**
MoCA, total score	**0.67** (0.54-0.80)	**−0.68** (−0.74 to −0.62)	**0.33** (0.30-0.35)	**−2.23** (−2.40 to −2.07)	0.26	Education, race, age, sex
Craft Story 21 recall immed. verbatim, total units	**1.51** (1.18-1.84)	**−0.94** (−1.09 to −0.79)	**0.44** (0.38-0.49)	**−2.00** (−2.43 to −1.57)	0.08	Education, age, race, sex
Craft Story 21 recall immed. paraphrase, total units	**0.92** (0.72-1.11)	**−0.60** (−0.69 to −0.52)	**0.29** (0.26-0.33)	**−1.53** (−1.79 to −1.28)	0.10	Education, age, race, sex
Craft Story 21 recall delay. verbatim, total units	**1.52** (1.19-1.85)	**−1.13** (−1.27 to −0.98)	**0.43** (0.37-0.49)	**−2.57** (−3.00 to −2.14)	0.10	Age, education, race, sex
Craft Story 21 recall delay. paraphrase, total units	**0.98** (0.77-1.19)	**−0.77** (−0.86 to −0.68)	**0.31** (0.28-0.35)	**−2.02** (−2.29 to −1.75)	0.12	Education, age, race, sex
Benson complex figure copy, total score	**0.16** (0.10-0.22)	**−0.12** (−0.15 to −0.09)	**0.07** (0.06-0.09)	**−0.51** (−0.59 to −0.43)	0.06	Education, race, age, sex
Benson complex figure recall, total score	**−0.26** (−0.40 to −0.12)	**−0.86** (−0.92 to −0.80)	**0.14** (0.11-0.16)	**−0.90** (−1.08 to −0.72)	0.12	Age, education, race, sex
Number span test forward, total correct trials	**−0.22** (−0.33 to −0.10)	**−0.23** (−0.28 to −0.17)	**0.15** (0.13-0.17)	**−0.59** (−0.74 to −0.44)	0.06	Education, age, race, sex
Number span test forward, longest span	**−0.11** (−0.18 to −0.05)	**−0.11** (−0.14 to −0.08)	**0.08** (0.07-0.10)	**−0.31** (−0.40 to −0.23)	0.06	Education, age, race, sex
Number span test backward, total correct trials	0.00 (−0.11 to 0.11)	**−0.32** (−0.37 to −0.27)	**0.16** (0.14-0.18)	**−1.07** (−1.21 to −0.93)	0.10	Education, race, age, sex
Number span test backward, longest span	0.00 (−0.06 to 0.07)	**−0.18** (−0.21 to −0.15)	**0.09** (0.08-0.10)	**−0.58** (−0.66 to −0.50)	0.10	Education, race, age, sex
MINT, total score	**−0.60** (−0.70 to −0.49)	**−0.18** (−0.22 to −0.13)	**0.17** (0.15-0.19)	**−2.02** (−2.16 to −1.89)	0.21	Race, education, sex, age
Phonemic test, F-words total in 60 s	**1.01** (0.79-1.22)	**−0.45** (−0.55 to −0.36)	**0.39** (0.35-0.43)	**−1.51** (−1.80 to −1.23)	0.09	Education, race, age, sex
Phonemic test, L words total in 60 s	**0.94** (0.73-1.14)	**−0.49** (−0.58 to −0.40)	**0.40** (0.36-0.44)	**−1.72** (−1.99 to −1.45)	0.10	Education, race, age, sex
Phonemic test, total F- and L-words	**1.96** (1.57-2.36)	**−0.93** (−1.10 to −0.75)	**0.77** (0.70-0.84)	**−3.13** (−3.67 to −2.59)	0.10	Education, race, age, sex
Animals list generation, total in 60 s	**0.72** (0.47-0.96)	**−1.38** (−1.49 to −1.27)	**0.48** (0.43-0.52)	**−3.46** (−3.77 to −3.14)	0.19	Age, education, race, sex
Vegetables list generation, total in 60 s	**2.83** (2.64-3.02)	**−0.69** (−0.78 to −0.61)	**0.26** (0.23-0.29)	**−1.31** (−1.55 to −1.07)	0.15	Sex, age, education, race
Trail making test part A, time (s)	**−1.30** (−1.81 to −0.78)	**4.05** (3.81-4.28)	**−0.71** (−0.80 to −0.62)	**7.53** (6.86-8.19)	0.21	Age, race, education, sex
Trail making test part A, correct lines/time (s)	**0.03** (0.02-0.04)	**−0.12** (−0.12 to −0.11)	**0.01** (0.01-0.02)	**−0.17** (−0.19 to −0.15)	0.24	Age, race, education, sex
Trail making test part B, time (s)	**−3.45** (−5.59 to −1.30)	**14.48** (13.52-15.45)	**−4.77** (−5.15 to −4.39)	**35.40** (32.61-38.20)	0.24	Age, race, education, sex
Trail making test part B, correct lines/time (s)	**0.01** (0.00-0.02)	**−0.05** (−0.06 to −0.05)	**0.01** (0.01-0.01)	**−0.09** (−0.10 to −0.09)	0.26	Age, race, education, sex

Bold values indicate statistical significance at $P < 0.01$.

*Higher scores indicate better scores except for the trail making test parts A and B.

†Age scaled by dividing by 10.

‡Variable importance was based on squared semipartial correlation which measured incremental value in R^2 after adding each variable to a model with the other 3 already in the model. Variables are ordered by importance based on which variable added the most increase in the R^2.

Delay indicates delayed; Immed, immediate; MINT, multilingual naming test; MoCA, Montreal Cognitive Assessment; UDS, Uniform Data Set.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

(eg, MoCA). We used these approaches to limit any circularity of examining test scores and clinical impressions, which were no doubt based in part, on current normative data. Although we continue to feel this is the most prudent approach, it is possible that some participants deemed to be cognitively normal on the CDR may have had the earliest stages of cognitive impairment.

In conclusion, the creation of standardized UDSNB 3.0 and its implementation by ADRCs within a diverse segment of the population enables the production of more precise normative data for use in defining cognitive performance and impairment in the increasingly diverse population of the United States. These normative data may function to reduce the misclassification of impairment in studies of age-related cognitive disorders such as Alzheimer disease and related dementias.

ACKNOWLEDGMENTS

The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions.

REFERENCES

1. 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;32:10–17.
2. 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;53:695–699.
3. Craft S, Newcomer J, Kanne S, et al. Memory improvement following induced hyperinsulinemia in Alzheimer's disease. *Neurobiol Aging*. 1996;17:123–130.
4. 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;19:272–283.
5. Possin KL, Laluz VR, Alcantar OZ, et al. Distinct neuroanatomical substrates and cognitive mechanisms of figure copy performance in Alzheimer's disease and behavioral variant frontotemporal dementia. *Neuropsychologia*. 2011;49:43–48.
6. Monsell SE, Dodge HH, Zhou X-H, et al. Results from the NACC Uniform Data Set neuropsychological battery crosswalk study. *Alzheimer Dis Assoc Disord*. 2016;30:134–139.
7. Dotson VM, Kitner-Triolo M, Evans MK, et al. Literacy-based normative data for low socioeconmic status African Americans. *Clin Neuropsychol*. 2008;22:989–1017.
8. Manly JJ. Advantages and disadvantages of separate norms for African Americans. *Clin Neuropsychol*. 2005;19:270–275.
9. Sisco S, Gross AL, Shih RA, et al. The role of early-life educational quality and literacy in explaining racial disparities in cognition in late life. *J Gerontol B Psychol Sci Soc Sci*. 2015;70:557–567.
10. Stasenko A, Jacobs DM, Salmon DP, et al. The Multilingual Naming Test (MINT) as a measure of picture naming ability in Alzheimer's disease. *J Int Neuropsychol Soc*. 2019;25:821–833.
11. Werry AE, Daniel M, Bergstrom B. Group differences in normal neuropsychological test performance for older non-Hispanic White and Black/African American adults. *Neuropsychology*. 2019;33:1089–1100.
12. Carson N, Leach L, Murphy KJ. A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *Int J Geriatr Psychiatry*. 2018;33:379–388.
13. Goldstein FC, Ashley AV, Miller E, et al. Validity of the Montreal Cognitive Assessment as a screen for mild cognitive impairment and dementia in African Americans. *J Geriatr Psychiatry*. 2014;27:199–203.
14. Waldron-Perrine B, Axelrod BN. Determining an appropriate cutting score for indication of impairment on the Montreal Cognitive Assessment. *Int J Geriatr Psychiatry*. 2012;27:1189–94.
15. Ferman TJ, Lucas JA, Ivnik RJ, et al. Mayo's Older African American Normative Studies: Auditory Verbal Learning Test norms for African American and Caucasian elders. *Clin Neuropsychol*. 2005;19:214–228.
16. Heaton RK, Miller SW, Taylor MJ, et al. Revised comprehensive norms for an expanded Halstead-Reitan battery: demographically adjusted neuropsychological norms for African American and Caucasians adults. Lutz, Fl: Psychological Assessment Resources Inc.; 2004.
17. Lucas JA, Ivnik RJ, Smith GE, et al. Mayo's Older Adrican Americans Normative Studies: Norms for Boston Naming Test, Controlled Oral Word Association, Category Fluency, Animal Naming, Token Test, WRAT-3 Reading, Trail Making Test, Stroop Test, and Judgment of Line Orientation. *Clin Neuropsychol*. 2005;19:189–213.
18. Lucas JA, Ivnik RJ, Smith GE, et al. Mayo's Older African Americans Normative Studies: normative data for commonly used clinical neuropsychological measures. *Clin Neuropsychol*. 2005;19:162–183.
19. Norman MA, Moore DJ, Taylor MJ, et al. Demographically corrected norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. *J Clin Exp Neuropsychol*. 2011;33:793–804.
20. Schneider AL, Sharrett AR, Gottesman RF, et al. Normative data for 8 neuropsychological tests in older blacks and whites from the atherosclerosis risk in communities (ARIC) study. *Alzheimer Dis Assoc Disord*. 2015;29:32–44.
21. Rilling LM, Lucas JA, Ivnik RJ, et al. Mayo's Older African American Normative Studies: Norms for the Mattis Dementia Rating Scale. *Clin Neuropsychol*. 2005;19:229–242.
22. Wechsler D. Wechsler Adult Intelligence Scale (4th ed.). San Antonio: Pearson; 2008.
23. Wechsler D. Wechsler Memory Scale (4th ed.). San Antonio: Pearson; 2009.
24. Advanced Clinical Solutions for WAIS-IV and WMS-IV. San Antonio: Pearson; 2009.
25. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993;43:2412–2414.
26. Carr JJ, Nelson JC, Wong ND, et al. Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) study. *Radiology*. 2005;234:35–43.
27. Army Individual Test Battery. Manual of Directions and Scoring. U.S. War Department AGsO, Washington, DC; 1944.
28. Manly JJ. Deconstructing race and ethnicity: implications for measurement of health outcomes. *Med Care*. 2006;44:S10–S16.
29. Manly JJ, Byrd DA, Touradji P, et al. Acculturation, reading level, and neuropsychological test performance among African American elders. *Appl Neuropsychol*. 2004;11:37–46.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.