Effects of age, gender, education and race on two tests of language ability in community-based older adults

Beth E. Snitz[1,*], **Frederick W. Unverzagt**[2], **Chung-Chou H. Chang**[3,4], **Joni Vander Bilt**[5], **Sujuan Gao**[6], **Judith Saxton**[1], **Kathleen S. Hall**[2], and **Mary Ganguli**[5,7]

[1] Department of Neurology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[2] Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA

[3] Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[4] Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA

[5] Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[6] Department of Medicine, Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, USA

[7] Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA

Abstract

Background—Neuropsychological tests, including tests of language ability, are frequently used to differentiate normal from pathological cognitive aging. However, language can be particularly difficult to assess in a standardized manner in cross-cultural studies and in patients from different educational and cultural backgrounds. This study examined the effects of age, gender, education and race on performance of two language tests, the animal fluency task (AFT) and the Indiana University Token Test (IUTT). We report population-based normative data on these tests from two combined ethnically divergent, cognitively normal, representative population samples of older adults.

Methods—Participants aged ≥65 years from the Monongahela-Youghiogheny Healthy Aging Team (MYHAT) and from the Indianapolis Study of Health and Aging (ISHA) were selected based on (1) a Clinical Dementia Rating (CDR) score of 0; (2) non-missing baseline language test data; and (3) race self-reported as African American or white. The combined sample (n=1885) was 28.1 % African American. Multivariate ordinal logistic regression was used to model the effects of demographic characteristics on test scores.

Results—On both language tests, better performance was significantly associated with higher education, younger age, and white race. On the IUTT, better performance was also associated with

female gender. We found no significant interactions between age and sex, and between race and education.

Conclusions—Age and education are more potent variables than are race and gender influencing performance on these language tests. Demographically-stratified normative tables for these measures can be used to guide test interpretation and aid clinical diagnosis of impaired cognition.

Keywords

neuropsychological tests; norms; cognitive aging; verbal fluency; token test

---

Differentiation of normal from pathological cognitive aging is at the cornerstone of dementia diagnosis in older adults. Neuropsychological tests are frequently used in this process. The technical challenge associated with interpreting test performance is increased when a patient's background diverges significantly from typical or published normative reference samples. Individuals from lower-educated or ethnic minority populations often have educational and cultural characteristics that complicate interpretation of cognitive test performance (Unverzagt *et al.*, 2007; Byrd *et al.*, 2005; Manly *et al.*, 1998a; Manly *et al.*, 1998b). Ethnic minority elders are the fastest-growing proportion of older adults in developed countries such as the United States (Day, 1996), highlighting a need for cognitive tests and normative data that can more accurately identify individuals with true cognitive impairment and early dementia. Normative data can be particularly valuable when they are drawn from population-based samples of older adults from different racial/ethnic groups, as well as from different economic, educational and cultural backgrounds.

Alzheimer's disease (AD) is the most common cause of dementia in late life (Canadian Study of Health and Aging Working Group, 1994; Hendrie *et al.*, 1995). Current diagnostic criteria specify presence of memory impairment plus deficits in at least one other cognitive area, including language, motor, visual-spatial, or executive disturbances (American Psychiatric Association, 1994). Of these, language impairment may be the most difficult to assess in older adults who are demographically different than the typical normative groups which are the basis for published test reference ranges. One reason for this is that language test performance is strongly influenced by educational and cultural factors (Lezak *et al.*, 1994; Manly *et al.*, 2004; Unverzagt *et al.*, 2007).

The token test, in various forms, has a long history in neuropsychology (De Renzi and Faglioni, 1978; Benton and Hamsher, 1989). This test is characterized primarily and historically as a test of verbal comprehension, including immediate memory span for verbal sequences and capacity to understand syntax. Secondarily, the task also taps some aspects of executive ability (via working memory), especially in its more syntactically complex and longer items. Likely for this reason, the token test is highly sensitive to brain damage and to dementia (Swihart *et al.*, 1989). With slight modifications this task has been adapted to screening batteries for dementia (Unverzagt et al., 1999). The IUTT is one such modified version which has shown diagnostic utility in differentiating demented from non-demented older adults (Jones and Ayers, 2006).

The animal fluency task (AFT) has also been widely used in neuropsychology (Lezak *et al.*, 1994). This task is a type of 'category fluency' task, which in turn is a type of 'verbal fluency' task. In contrast to the token test, the AFT is an expressive language test. Factor analytic studies indicate numerous dimensions of cognitive processes involved in the AFT (and other verbal fluency tasks), including primarily language functions, such as vocabulary storage, but also speeded mental processing and executive functions (Lamar *et al.*, 2002). Again, similarly to the token test, the multidimensionality of cognitive processes, including executive functions,

likely contributes to the high sensitivity of verbal fluency tasks to global brain damage as well as to dementia.

The goals of the present study were to establish population-based norms, and to examine the effects of age, gender, education and race on language task performance in two ethnically divergent, cognitively normal, representative population samples of older adults. To allow comparison of white and African-American groups, we pooled data from two ongoing epidemiological population-based studies in the United States, one largely white and one exclusively African American. Specifically, we examined two commonly used clinical measures of language ability, verbal fluency (animal fluency task; AFT) (Rosen, 1980) and modified version of the token test: the Indiana University Token Test (Indiana University Token Test; IUTT) (Unverzagt *et al.*, 1999).

Method

General sampling and recruitment

Monongahela-Youghiogheny Healthy Aging Team (MYHAT)—The communities selected for the MYHAT study are close to the city of Pittsburgh, in the state of Pennsylvania, and surround the confluence of the Monongahela and Youghiogheny rivers. Official voter registration lists have been used as the sampling frames for our community studies for the past two decades, given the stability of the local population. The 2004 voter registration list was used as the frame for an age-stratified random sample for recruitment beginning in selected towns in 2006. Sampling aimed to achieve approximately equal numbers of participants in age-intervals 65–74, 75–84, and 85+ years. Sampling ratios were derived accordingly, with subsequent over-sampling in those aged 80–84 at study entry to compensate for small numbers of individuals 85+ years old. Community outreach and recruitment procedures were approved by the University of Pittsburgh Institutional Review Board. Outreach efforts included press releases to local newspapers, speaking engagements with community groups, and meetings with municipal officials and community leaders to explain the study and obtain local support. Selected individuals were contacted first by mail and then via mail telephone or a flyer personally delivered to the home. Recruitment criteria were (a) age 65 years or older, (b) living within the selected area, (c) not already in long-term care institution. Individuals were considered ineligible if they (d) were too ill to participate, (e) had severe vision impairment, (f) had severe hearing impairment, (g) were decisionally incapacitated (Ganguli *et al.*, In press). A total of 2036 individuals were recruited. The current report is based on data from the first (baseline) data collection wave of this ongoing longitudinal study.

Indianapolis Study of Health and Aging (ISHA)—Indianapolis is the capital of the state of Indiana. ISHA had two phases of recruitment. The initial phase began in 1992 when a random sample of residences in 29 contiguous census tracts within Indianapolis were approached by interviewers seeking residents of the address that self-identified as African American and were age 65 years or older. A total of 2,212 individuals were enrolled (85% of the eligible persons approached). In 2001, the cohort was enriched using Medicare rolls to include African Americans who were aged 70 years or older living in Indianapolis. A total of 1,892 individuals were enrolled from this phase. ISHA is a longitudinal study; data reported here were collected from the third wave onwards, corresponding to the given participant's first exposure to the IUTT test.

Assessment procedure

MYHAT—All participants provided written informed consent following procedures approved by the University of Pittsburgh's Institutional Review Board. Testing took place at either the participant's residence or the project field office. A single-stage assessment was employed.

Initially, the Mini-Mental State Examination (MMSE) was administered and scored with correction for age and education (Mungas et al., 1996). Those scoring >=21 on the age-education corrected MMSE were designated as either cognitively normal or mildly cognitively impaired. They proceeded to further evaluation, including neuropsychological assessment, clinical history, physical and neurological examination, and completion of the Clinical Dementia Rating (CDR) scale (Hughes et al., 1982). The neuropsychological battery consisted of tests tapping multiple cognitive domains and took approximately one hour to administer.

ISHA—All participants provided informed consent to participate and the project had ongoing review and approval of the IUPUI Institutional Review Board. Clinical assessments were conducted by clinicians usually in the participant's home. The study utilized a two-stage study design with screening of the entire cohort using the Community Screening Interview for Dementia (CSI-D) followed in a subgroup by a clinical assessment including neuropsychological assessment, physical and neurological examinations, standardized health examination and assessment of function, and semi-structured informant interview. CDR scores (Hughes et al 1982) were assigned in a consensus diagnostic conference. The neuropsychological assessment used a modified and expanded version of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) test battery which took 45 minutes to administer.

This report is focused on two tests of language ability that were administered in both studies. The IUTT was administered to participants in Indianapolis beginning with the third wave of the ISHA study, whereas the AFT was administered starting with the first wave. Both language tests were administered in MYHAT in the first wave.

Language tests

Animal Fluency Task (AFT; Rosen, 1980)—Participants were asked to name as many different animals as possible in 60 seconds. The raw score is the number of animal names generated, including superordinate and specific category names, male/female variants, and parent/offspring variants. This is an expressive language task tapping multiple cognitive processes including fluency of speech, integrity of semantic storage and retrieval, and self-directed search strategy (Lezak *et al.*, 1994).

Indiana University Token Test (IUTT; Unverzagt et al., 1999)—This is a test of aural comprehension and execution of two-and three-step commands. Participants are presented with an 8 ½ × 11-inch sheet of paper, that contains four rows of circles and squares that vary in shape (circles and squares), color (red, black, yellow, and green) and size (large and small). The examiner reads a series of commands (12 total) of increasing complexity (e.g., simple command: "Point to a circle;" complex command: "Before pointing to the small red circle, point to the large green square."). A correct performance receives 2 points. After an incorrect performance, the command is re-stated. Correct performance at this level receives 1 point (score range from 0–24).

Study sample and statistical analysis

For this study, participants were selected from Wave 1 of MYHAT who had a Clinical Dementia Rating (CDR) global score = 0, indicating absence of dementia and normal cognitive functioning for age. In the ISHA sample, participants with first exposure to the IUTT, CDR = 0, and diagnosis of normal were selected for the analytic sample. Participants from either site were also excluded for missing values on the outcome measures (IUTT, AFT) or the independent covariates (age, sex, education, and race).

We categorized the combined MYHAT-ISHA sample into subgroups aged less than 80 years vs. 80 or more years old, and those with less than 12 years of education vs. those with 12 or more years of education. We first examined the distribution of scores on the AFT and IUTT tests in the combined sample between the above age, sex, education, and race groups. The distribution of scores was normal on the AFT but skewed on the IUTT. For establishment of norms, we divided the combined sample into white and African-American groups and reported the mean (SD), 5th, 10th, and 50th percentile scores in each group, within further demographic subgroups.

Given that the IUTT was not normally distributed, it was not possible to treat it as a continuous variable in a linear model. Therefore, we treated both IUTT and AFT as categorical variables, grouped in 3 ordered categories (<10th %tile, 10th %tile to 50th %tile, ≥50th %tile) which we will refer to as low, medium, and high level scores. We did not break out the upper end of the distributions further as the lower percentiles were of main clinical interest (i.e., consistent with our goal of identifying cognitive impairment).

Test scores ordered at the multiple levels noted above were used as the outcome variables in multivariable analyses were performed using ordinal logistic regression modelling. This is a proportional odds model, which makes and tests the assumption that the cumulative odds ratios for any two values of a covariate are the same for each level of the outcome variable. The covariates entered in the model were age (<80, ≥80 years old), sex, education (<12 years, ≥ 12 years), and race (white, African-American). The odds ratio (OR) represents the effect of a change in level of the predictor variable on the logit (log odds). For example, an odds ratio of 1.2 for education would indicate that being in the higher education group (≥12 years) was associated with a 20% greater odds of a higher level test score compared to the lower education group. An approximate likelihood ratio test was used to verify the proportional odds assumption in the final model. To further verify the proportional odds model, we fit binary logistic regressions using IUTT cutoffs of at least 10th percentile and at least 50th percentile and compared the results with those of the ordinal regressions. Variables were considered significant if the corresponding p value of the likelihood ratio test <0.05. Interactions among the final covariates were then tested.

The sample size provided 90% power, at alpha level 0.05, to detect an odds ratio as small as 1.49 for greater education, 1.45 for younger age, 1.45 for female gender, and 1.56 for white race.

We also included a post-hoc analysis examining change in effect of race on language test performance after inclusion of a measure of reading ability (Wide Range Achievement Test Reading – 3rd ed.; WRAT-3) to estimate quality of education in the MYHAT sample, only (n = 1331; African-Americans: n = 50, 3.8%). This measure was available in the MYHAT study but not in the ISHA study, and therefore it was not included in the main a priori analyses involving the combined MYHAT/ISHA sample.

SAS 9.0 and Stata 9.0 were used for all data management and statistical analyses.

Results

Sample characteristics

MYHAT—The total size of the cohort was 2036, of whom 1982 individuals had age- and education-adjusted MMSE scores > 21 and underwent the full assessment, and 1967 completed the neuropsychological battery. The analyses reported here are restricted to those reporting their race as either African American or white, being rated as 0 on the CDR, and providing

complete data on the IUTT and AFT, resulting in an selected sample of n = 1413 (see Table 1). The number of African Americans in this sample was n = 55 (3.9 %).

ISHA—From 1993 to 2004, 516 participants were tested with the IUTT for the first time. Further restricting to CDR rating of 0, and no missing data on the animal fluency score and demographic information, 472 participants were included in this analysis (121 tested during 1993 to 1995, 98 in 1998, 146 in 2002 and 107 in 2004. All participants were African American (see Table 1).

Combined MYHAT/ISHA sample—Table 1 presents demographic characteristics of the combined study sample with CDR ratings of 0 (n = 1885). The mean age was 77.1 (SD 7.1), the mean education was 12.2 years (SD 3.0), and 10.1 % of the sample had self-reported education in the range of 0–8 years. 65.2 % were women and 28.1 % were African-American.

Influence of demographic variables and normative tables

Table 2 shows results from the multivariate ordinal logistic regression analysis showed the effects and relative importance of the covariates (predictor variables) on the two language ability test scores. Results are given as odds ratios (ORs) for an IUTT (or AFT) score which was at least at the 10th %ile (medium and high scores) compared with less than 10th %ile (low score), and at least the 50th %ile (high score) of the sample, compared with scores less than the 50th %ile (low and medium scores). Since this is a multivariable model, the OR for each covariate is adjusted for the effects of the other significant covariates. Greater age was independently associated with lower scores on the IUTT (OR, 0.50, 95% CI, 0.40–0.61). The odds ratio of 0.5 shows that compared to being younger (under age 80), being older (aged 80 or more) is associated with a 50% lower probability of a high score compared to a medium score, and of a medium score compared to a low score. Inversely stated, being younger is associated with a better score with an odds ratio of 2.0. Similarly, greater age was associated with lower scores on the AFT (OR, 0.56, 95% CI, 0.46–0.68). Female gender was associated with higher scores on the IUTT only (OR, 1.59, 95% CI, 1.28–1.96). Higher education was associated with better scores on the IUTT (OR, 3.06, 95% CI, 2.39–3.93) and with the AFT (OR, 1.91, 95% CI, 1.51–2.42). Race was associated with IUTT scores (African-American vs. white: OR, 0.74, 95% CI, 0.58–0.95) and with AFT scores (African-American vs. white: OR, 0.79, 95% CI, 0.63–0.99). Tests of the proportional odds assumption were not significant (chi-square statistic with 4 degrees of freedom =3.92, p=0.42), indicating that the proportional odds model was appropriate for these data.

Rank-ordering the effects of the different covariates, we found that for the IUTT, the odds of higher scores were associated with higher education (odds ratio 3.06 or threefold increase in odds of better score) younger age (odds ratio 2.0 or twofold increase), female gender (odds ratio 1.59), and white race (odds ratio 1.35). On the AFT, the odds of a higher score were associated with higher education (odds ratio 1.91), younger age (odds ratio 1.79, inverse of 0.56) and white race (odds ratio 1.27, inverse of 0.79).

We examined potential variation in effects across the combined sample with the use of interaction terms in the regression models. We found non-significant interactions between age and sex, and between race and education.

Table 3 provides normative data (mean, standard deviation, 5th, 10th, and 50th percentile scores) for the AFT in African American and white subgroups of the combined MYHAT-ISHA sample, combined as well as stratified by gender, education, and age. Table 4 provides similar normative data for the IUTT in African American and white subgroups of the combined MYHAT-ISHA sample, combined as well as stratified by gender, education, and age.

The post-hoc analysis of the effect of adding WRAT-3 Reading score in the regression models, in the restricted MYHAT sample, resulted in attenuation of the effect of race on both the AFT (OR=0.49 to OR=0.57) and the IUTT (OR=0.446 to OR=0.55). For both tests the effect of race on test performance in the multivariate models with reading level included became non-significant (p > .05).

Discussion

In this study, we examined the influence of demographic factors on language-based cognitive test performance among cognitively normal older adults from two U.S. population-based samples. In this ethnically diverse sample, verbal spontaneity and aural comprehension were influenced by age and education while race and gender also affected performance but to a much smaller degree. One can observe that stratifying normative tables by demographic factors does affect, by up to several points difference, the lower percentile cut-off scores at the tails of the test score distributions. Thus, these analyses formed the basis for demographically-stratified normative tables for these measures which can be used to guide test interpretation and aid clinical diagnosis.

Regarding the IUTT, we found that performance within our clinically unimpaired cohort was significantly influenced by age, gender, education, and race. These findings are consistent with other studies who found that younger adults performed better than older adults (Ivnik *et al.*, 1996); that better educated individuals performed better than those with fewer years of education (De Renzi and Faglioni, 1978); and that women performed better than men (Sarno *et al.*, 1985). We are unaware of other studies examining the effect of race on IUTT scores. With regard to race in this present study, we found that white participants performed better than African Americans. However, in the MYHAT study which included a measure of reading ability, the reading score was a significant predictor for both language test scores, and the effect of race was attenuated and became non-significant for both language tests. This post-hoc analysis is consistent with literature that suggests "years of education" may be an inadequate measure of the educational experience of multicultural elders, and that race is likely confounded by literacy and quality of education.

Regarding the AFT, we found that better performance was associated with younger age, higher education and white race. There is some variability in the literature reporting on influence of demographic factors on verbal fluency tests, with many studies consistent with our significant findings on age (e.g., Acevedo *et al.*, 2000) and education (e.g., Ivnik *et al.*, 1996; Troyer 2000), but many reporting no age effect (e.g., Bolla *et al.*, 1990) or education effect (Stuss *et al.*, 1998). Similarly, some investigators have reported gender differences in verbal fluency with female superiority (Acevedo *et al.*, 2000; Bolla *et al.*, 1990), while others have reported no gender difference (Troyer, 2000). One study reported differences in AFT performance among racial groups of male veterans, with white participants performing better than African-American participants, who in turn scored higher on average than Hispanic participants (Johnson-Selfridge *et al.*, 1998). In general, it is likely that negative studies reporting lack of significant effects of demographics on test performance have insufficient power to detect effects, and may account for inter-study differences in findings. We are unaware of any studies which contradict the direction of the associations reported here.

A principal reason for focusing on tests of language, rather than of other cognitive domains, is that language can be particularly difficult to assess in a standardized and unbiased manner in cross-cultural studies and in patients from different cultural backgrounds. While all participants in the ISHA and MYHAT studies were English-speaking Americans, our experience may be of value to clinicians and researchers working with older adults of a variety of backgrounds. Age and education (and related cohort effects) are likely to influence test

performance within any group (Unverzagt *et al.*, 1996). However, while our study did indicate statistically significant effects of education and race at the group level, the IUTT and the AFT may nevertheless be expected to be useful in truly illiterate populations. In such populations, other widely used tests such as initial letter fluency could not be used, as it requires by the very nature of the test, that the individual is literate (i.e., generating words that begin with a certain letter). In terms of face validity, both the category fluency task and the IU Token Test would appear relatively resistant to cultural bias. They primarily require naming familiar objects within a category or pointing to recognizable shapes and colors, provided the concepts of shape and color are recognized within a cultural group. For example, compared to confrontation naming tasks, there is no need to generate lists of objects of high and low frequency in a given language and culture. As well, the instructions appear straightforward and easy to translate, and scoring is also relatively easy, making training of testers less of a challenge. However, tests should only be interpreted and applied in cross-cultural and cross-national studies after adequate piloting and norming, We did this in the present study. Further, to our knowledge, this is the only study that has simultaneously examined the influence of age, education, gender and race on test performance for these language tasks. In general, our results suggest that age and education are more potent variables than are race and gender.

One limitation in the present study is that we did not have a common measure of literacy or reading level ability available with which to estimate individual differences in the quality of education, which may vary as a function of race. There is evidence that educational quality differences, as estimated by reading abilities, contribute to racial cognitive test performance differences (Manly *et al.*, 1998a; Manly *et al.*, 1998b; Manly *et al.*, 2004; Gasquoine, 1999).

Another limitation of the present study is that ethnicity is confounded with study site. It is possible that unknown, regionally specific factors may be interacting with race to contribute to the obtained pattern of results. Finally, caution is warranted when applying any normative test data to individual clinical patients who are significantly different from the normative samples on relevant demographic characteristics.

Normative studies from large community samples, such as the present one, help establish the demographic factors that affect cognitive test performance and aid in identifying cognitive impairment in older adults, especially those of minority background or of lower educational attainment. Important future research directions for similar normative studies would include introducing approaches to try to capture unmeasured variance in our demographic factors, such as "race", (i.e., quality of education, measures of acculturation, persistence of poverty) (Manly *et al.*, 1998a; Manly *et al.*, 1998b; Manly *et al.*, 2004; Gasquoine, 1999). As well, future studies should continue to expand the normative literature to more ethnic and cultural groups to keep pace with the ever increasing cultural diversity of older adults in many societies around the world.

Acknowledgments

References

Acevedo A, Loewenstein DA, Barker WW, Harwood DG, Luis C, Bravo M, Hurwitz DA, Aguero H, Greenfield L, Duara R. Category fluency test: Normative data for English- and Spanish-speaking elderly. Journal of the International Neuropsychological Society 2000;6:760–769. [PubMed: 11105466]

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Vol. 4. Washington, DC: American Psychiatric Association; 1994.

Benton, AL.; Hamsher, KD. Multilingual Aphasia Examination. Iowa City: AJA Associates; 1989.

Bolla KI, Lindgren KN, Bonaccorsy C, Bleecker ML. Predictors of verbal fluency (FAS) in the healthy elderly. Journal of Clinical Psychology 1990;46:623–8. [PubMed: 2246370]

Byrd DA, Sanchez D, Manly JJ. Neuropsychological test performance among Caribbean-born and U.S.-born African American elderly: the role of age, education and reading level. Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society 2005;27:1056–69.

Canadian Study of Health and Aging Working Group. Canadian study of health and aging: study methods and prevalence of dementia. Canadian Medical Association Journal 1994;150:899–913. [PubMed: 8131123]

Cooper DB, Lacritz LH, Weiner MF, Rosenberg RN, Cullum CM. Category fluency in mild cognitive impairment: reduced effect of practice in test-retest conditions. Alzheimer Disease & Associated Disorders 2004;18:120–2. [PubMed: 15494616]

Day, JC. Current Population Reports. US Bureau of the Census; Washington, DC: 1996. p. 25-1130.

De Renzi E, Faglioni P. Normative data and screening power of a shortened version of the Token Test. Cortex 1978;14:41–49. [PubMed: 16295108]

Ganguli, M.; Snitz, B.; Vander Bilt, J.; Chang, CCH. The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) study. International Journal of Geriatric Psychiatry; How much do depressive symptoms affect cognition at the population level?. (In press)

Gasquoine PG. Variables moderating cultural and ethnic differences in neuropsychological assessment: The case of Hispanic Americans. Clinical Neuropsychologist 1999;13:376–383. [PubMed: 10726609]

Hendrie HC, Ogunniyi A, Hall KS, Baiyewu O, Unverzagt FW, Gureje O, Gao S, Evans RM, Ogunseyinde AO, Adeyinka AO, Musick B, Hui SL. Incidence of dementia and Alzheimer disease in 2 communities: Yoruba residing in Ibadan, Nigeria, and African Americans residing in Indianapolis, Indiana. JAMA 2001;285:739–47. [PubMed: 11176911]

Hendrie HC, Osuntokun BO, Hall KS, Ogunniyi AO, Hui SL, Unverzagt FW, Gureje O, Rodenberg CA, Baiyewu O, Musick BS. Prevalence of Alzheimer's disease and dementia in two communities: Nigerian Africans and African Americans. American Journal of Psychiatry 1995;152:1485–1492. [PubMed: 7573588]

Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC. Neuropsychological tests' norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. Clinical Neuropsychologist 1996;10:262–278.

Johnson-Selfridge M, Zalewski C, Aboudarham J. The relationship between ethnicity and word fluency. Archives of Clinical Neuropsychology 1998;13:319–325. [PubMed: 14590645]

Jones SN, Ayers CR. Psychometric properties and factor structure of an expanded CERAD neuropsychological battery in an elderly VA sample. Archives of Clinical Neuropsychology 2006;21:359–365. [PubMed: 16766158]

Lamar M, Zonderman AB, Resnick S. Contribution of specific cognitive processes to executive functioning in an aging population. Neuropsychology 2002;16:156–62. [PubMed: 11949707]

Lezak, MD.; Howieson, DB.; Loring, DW.; Hannay, J.; Fischer, JS. Neuropsychological Assessment. New York: Oxford University Press; 1994.

Manly JJ, Byrd DA, Touradji P, Stern Y. Acculturation, reading level, and neuropsychological test performance among African American elders. Applied Neuropsychology 2004;11:37–46. [PubMed: 15471745]

Manly JJ, Jacobs DM, Sano M, Bell K, Merchant CA, Small SA, Stern Y. Cognitive test performance among nondemented elderly African Americans and whites. Neurology 1998a;50:1238–45. [PubMed: 9595969]

Manly JJ, Miller SW, Heaton RK, Byrd D, Reilly J, Velasquez RJ, Saccuzzo DP, Grant I. HIV-Neurobehavioral-Res-Ctr-Grp. The effect of African-American acculturation on neuropsychological test performance in normal and HIV-positive individuals. Journal of the International Neuropsychological Society 1998b;4:291–302. [PubMed: 9623004]

Rosen WG. Verbal fluency in aging and dementia. Journal of Clinical Neuropsychology 1980;2:135–146.

Sarno MT, Buonaguro A, Levita E. Gender and recovery from aphasia after stroke. Journal of Nervous and Mental Disease 1985;173:605–609. [PubMed: 2411851]

Spreen, O.; Strauss, E. A compendium of neuropsychological tests. New York: Oxford University Press; 1998.

Stuss DT, Alexander MP, Hamer L, Palumbo C, Dempster R, Binns M, Levine B, Izukawa D. The effects of focal anterior and posterior brain lesions on verbal fluency. Journal of the International Neuropsychological Society 1998;4:265–78. [PubMed: 9623001]

Swihart AA, Panisett M, Becker JT, Beyer JR, Boller F. The Token Test: Validity and diagnostic power in Alzheimer's disease. Developmental Neuropsychology 1989;5:69–78.

Troyer AK. Normative data for clustering and switching on verbal fluency tasks. Journal of Clinical and Experimental Neuropsychology 2000;22:370–378. [PubMed: 10855044]

Unverzagt FW, Gao S, Lane KA, Callahan C, Ogunniyi A, Baiyewu O, Gureje O, Hall KS, Hendrie HC. Mild cognitive dysfunction: An epidemiological perspective with an emphasis on African Americans. Journal of Geriatric Psychiatry and Neurology 2007;20:215–226. [PubMed: 18004008]

Unverzagt FW, Hall KS, Torke AM, Rediger JD, Mercado N, Gureje O, Osuntokun BO, Hendrie HC. Effects of age, education, and gender on CERAD neuropsychological test performance in an African American sample. Clinical Neuropsychologist 1996;10:180–190.

Unverzagt FW, Morgan OS, Thesiger CH, Eldemire DA, Luseko J, Pokuri S, Hui SL, Hall KS, Hendrie HC. Clinical utility of CERAD neuropsychological battery in elderly Jamaicans. Journal of the International Neuropsychological Society 1999;5:255–259. [PubMed: 10217925]

Table 1

Demographic characteristics

	ISHA (n=472)		MYHAT (n=1413)		Combined sample (n=1885)	
Age in years (mean ± sd)	78.7 ± 6.4		76.8 ± 7.3		77.2 ± 7.1	
Age 65–74 years (n, %)	125	26.3	539	38.2	664	35.2
75–84 years (n, %)	269	56.6	655	46.4	924	48.9
85–94 years (n, %)	75	15.8	212	15.0	287	15.2
≥ 95 years (n, %)	6	1.3	7	0.5	13	0.7
Gender: Female (n, %)	332	69.9	898	63.6	1230	65.2
Ethnicity: African American (n, %)	475	100.0	55	3.9	530	28.1
Education in years (mean ± sd)	9.9 ± 3.3		13.0 ± 2.4		12.2 ± 3.0	
Education 0–8 years (n, %)	152	32.1	39	2.8	191	10.1
9–12 years (n, %)	249	52.6	785	55.6	1,034	54.8
≥ 13 years (n, %)	72	15.2	589	41.7	661	35.1

Table 2

Associations of test scores with demographic characteristics (multiple ordinal logistic regression models for combined sample (N = 1851)

Covariate*	IU Token Test Score					Animal Fluency Task Score				
	Odds Ratio	Std. Err.	z	P value	95% Conf. Int.	Odds Ratio	Std. Err.	z	P value	95% Conf. Int.
Age >=80 years	0.4954	0.0525	−6.63	<0.001	0.4025 0.6097	0.5618	0.0549	−5.90	<0.001	0.4638 0.6804
Gender (Female)	1.5887	0.1722	4.27	<0.001	1.2846 1.9648	0.8633	0.0883	−1.44	0.150	0.7065 1.0550
Education >=12 years	3.0644	0.3878	8.85	<0.001	2.3913 3.9269	1.9106	0.2312	5.35	<0.001	1.5071 2.4220
Ethnicity: African American	0.7439	0.0944	−2.33	0.020	0.5802 0.9538	0.7898	0.0935	−1.99	0.046	0.6262 0.9961

Table 3

Mean (SD), median and selected percentile scores on the Animal Fluency Task score, N=1857, combined samples

Race/ethnicity: African American, n=527

Gender		Women, n=373			Men, n=154			Both genders, n=527		
Education		<HS n=209	>=HS n=164	Both education groups n=373	<HS n=94	>=HS n=60	Both education groups n=154	<HS n=303	>=HS n=224	Both education groups n=527
Age										
<80 years	Mean (SD)	13.8 (4.0)	15.0 (4.5)	14.3 (4.3)	14.4 (3.9)	16.9 (4.9)	15.2 (4.4)	14.0 (4.0)	15.5 (4.7)	14.6 (4.3)
	Median	13	15	13.5	14	16.5	15	13	15	14
	10th%tile	9	10	10	9	11	9	9	10	10
	5th%tile	8	10	8	8	9	9	8	9	8
	n	116	102	218	71	36	107	187	138	325
>=80 years	Mean (SD)	13.2 (4.2)	14.4 (4.4)	13.7 (4.3)	13.6 (3.6)	17.0 (4.1)	15.3 (4.2)	13.3 (4.1)	15.1 (4.4)	14.0 (4.3)
	Median	13	14	13	14	17	15	13	14	14
	10th%tile	8	10	8	9	13	8	8	10	9
	5th%tile	7	8	7	6	12	7	6	8	7
	n	93	62	155	23	24	47	116	86	202
Both age groups	Mean(SD)	13.5 (4.1)	14.8 (4.5)	14.1 (4.3)	14.2 (3.8)	16.9 (4.6)	15.2 (4.3)	13.7 (4.0)	15.3 (4.6)	14.4 (4.3)
	Median	13	14	13	14	17	15	13	15	14
	10th%tile	9	10	9	9	12	9	9	10	9
	5th%tile	8	8	8	8	10	8	8	9	8
	n	209	164	373	94	60	154	303	224	527

Race/ethnicity: White, n=1330

Gender		Women, n=835			Men, n=495			Both genders, n=1330		
Education		<HS n=98	>=HS n=737	Both education groups n=835	<HS n=67	>=HS n=428	Both education groups n=495	<HS n=165	>=HS n=1165	Both education groups n=1330
Age										
<80 years	Mean(SD)	15.6 (4.3)	17.8 (4.3)	17.6 (4.3)	14.8 (4.5)	17.6 (4.7)	17.3 (4.8)	15.3 (4.4)	17.7 (4.4)	17.5 (4.5)
	Median	15	18	18	15	17	17	15	17	17
	10th%tile	9	12	12	10	12	12	10	12	12
	5th%tile	9	11	11	9	11	10	9	11	11
	n	39	462	501	26	265	291	65	727	792
>=80 years	Mean (SD)	13.7 (3.9)	15.0 (4.3)	14.8 (4.2)	15.2 (5.2)	15.7 (4.4)	15.6 (4.5)	14.4 (4.5)	15.3 (4.3)	15.1 (4.3)
	Median	13	15	15	15	16	15	14	15	15
	10th%tile	8	10	10	9	10	10	9	10	10
	5th%tile	8	8	8	9	9	9	8	8	8
	n	59	275	334	41	163	204	100	438	538
Both age groups	Mean (SD)	14.5 (4.2)	16.7 (4.5)	16.5 (4.5)	15.1 (4.9)	16.9 (4.7)	16.6 (4.7)	14.7 (4.5)	16.8 (4.5)	16.5 (4.6)
	Median	14	17	16	15	17	16	14	17	16
	10th%tile	9	11	11	9	11	11	9	11	11
	5th%tile	8	10	9	9	10	9	9	10	9
	n	98	737	835	67	428	495	165	1165	1330

Table 4

Mean (SD), median and lower percentile cut-off scores on the IU Token Test Score, N=1866, combined samples

Race/ethnicity: African American, n=528

Gender		Women, n=374			Men, n=154			Both gender, n=528		
Education		<HS n=210	>=HS n=164	Both education groups n=374	<HS n=94	>=HS n=60	Both education groups n=154	<HS n=304	>=HS n=224	Both education groups n=528
Age <80 years	Mean(SD)	21.0 (3.0)	22.4 (2.2)	21.7 (2.7)	20.6 (2.7)	22.5 (1.3)	21.7 (2.7)	20.9 (2.9)	22.4 (2.0)	21.5 (2.7)
	Median	21	23	22	21	22	22	21	23	22
	10th%tile	18	20	18	17	21	18	17	21	18
	5th%tile	16	17	17	15	20	17	16	18	17
	n	*117*	*102*	*219*	*71*	*36*	*107*	*188*	*138*	*326*
>=80 years	Mean(SD)	20.8 (3.0)	21.8 (2.8)	21.2 (3.0)	20.0 (2.8)	22.4 (1.7)	21.2 (2.5)	20.6 (3.0)	22.0 (2.6)	21.2 (2.9)
	Median	22	23	22	21	23	22	21.5	23	22
	10th%tile	16	18	17	15	20	18	16	18	17
	5th%tile	15	16	16	15	19	15	15	16	16
	n	*93*	*62*	*155*	*23*	*24*	*47*	*116*	*86*	*202*
Both age groups	Mean(SD)	20.9 (3.0)	22.2 (2.5)	21.5 (2.9)	20.4 (2.7)	22.4 (1.5)	21.2 (2.5)	20.8 (2.9)	22.3 (2.3)	21.4 (2.7)
	Median	21.5	23	22	21	23	22	21	23	22
	10th%tile	17	19	18	17	21	18	17	20	18
	5th%tile	16	17	16	15	20	16	15	17	16
	n	*210*	*164*	*374*	*94*	*60*	*154*	*304*	*224*	*528*

Race/ethnicity: White, n=1338

Gender		Women, n=841			Men, n=497			Both gender, n=1338		
Education		<HS n=99	>=HS n=742	Both education groups n=841	<HS n=67	>=HS n=430	Both education groups n=497	<HS n=166	>=HS n=1172	Both education groups n=1338
Age <80 years	Mean(SD)	23.2 (1.1)	23.5 (0.9)	23.4 (0.9)	21.5 (1.9)	23.1 (1.5)	23.0 (1.6)	22.5 (1.7)	23.3 (1.2)	23.3 (1.2)
	Median	23.5	24	24	22	24	24	23	24	24
	10th%tile	22	22	22	19	21	21	20	22	22
	5th%tile	21	22	22	18	21	20	19	21	21
	n	*40*	*463*	*503*	*26*	*266*	*292*	*66*	*729*	*795*
>=80 years	Mean(SD)	22.2 (1.7)	22.8 (1.5)	22.7 (1.5)	21.8 (2.4)	22.6 (1.4)	22.5 (1.7)	22.0 (2.0)	22.7 (1.5)	22.6 (1.6)
	Median	22	23	23	22	23	23	22	23	23
	10th%tile	20	20	20	20	21	21	20	21	20
	5th%tile	19	20	20	19	20	20	19	20	20
	n	*59*	*279*	*338*	*41*	*164*	*205*	*100*	*443*	*543*
Both age groups	Mean(SD)	22.6 (1.5)	23.2 (1.2)	23.1 (1.3)	21.7 (2.2)	22.9 (1.5)	22.8 (1.7)	22.2 (1.9)	23.1 (1.3)	23.0 (1.4)
	Median	23	24	24	23	23	23	22	24	24
	10th%tile	20	22	21	19	21	21	20	21	21
	5th%tile	20	20	20	18	20	20	19	21	20
	n	*99*	*742*	*841*	*67*	*430*	*497*	*166*	*1172*	*1338*