

Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old

M.S. Beeri, PhD; J. Schmeidler, PhD; M. Sano, PhD; J. Wang, BA; R. Lally, BA; H. Grossman, MD; and J.M. Silverman, PhD

Abstract—Objective: To evaluate the performance of nondemented subjects 85 years and older on the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) neuropsychological battery, and to assess its relationship with sociodemographic variables. **Methods:** We studied 196 subjects enrolled in an Alzheimer’s Disease Research Center study who had a complete CERAD neuropsychological assessment. We used multiple regression analysis to predict performance on the neuropsychological tests from age, education, and sex. Eight representative hypothetical individuals were created (for example, an 87-year-old man, with high education). For each test, estimates of performance at the 10th, 25th, 50th, and 75th percentiles were reported for the eight representative hypothetical individuals. **Results:** Mean age was 89.2 years (SD = 3.2), mean years of education was 14.9 (SD = 3.2), and 66% of the sample were women. For 11 of the 14 neuropsychological tests, there was a significant multiple regression model using education, age, and sex as predictors. Neither the models nor the predictors used individually were significant for Delayed Recall, Savings, or correct Recognition. Among the significant results, seven had education as the strongest predictor. Lower age and higher education were associated with better performance. Women performed better than men in three of four tests with significant results for sex. **Conclusions:** In a sample of oldest old whose primary language is English, neuropsychological testing is influenced mainly by education and age. Cutoff scores based on younger populations and applied to the oldest old might lead to increased false-positive misclassifications.

NEUROLOGY 2006;67:1006–1010

The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) developed a relatively brief neuropsychological battery for the assessment of patients with Alzheimer disease (AD). Norms for the battery have been published for several cultures, populations, and settings.^{1–5} Norms have also been published for a broad range of ages^{6–8} where the highest range is generally referred to as “80+.” From the mean and SD of these older age categories, most of the subjects in these studies are probably close to age 80 years.

Age as well as sex and education are associated with the scores of the tests of the CERAD neuropsychological battery.^{2,8} Without a clear indication of the normal range of performance expected from nondemented very old individuals, and the understanding of the associations of demographic variables with

these scores, interpretation of test results becomes subjective, prone to error, and when basing the interpretation on norms for younger cohorts, at great risk for false positives for impairment. In this study we evaluated the performance on the CERAD neuropsychological battery in 196 nondemented elderly, whose primary language is English, with ages ranging from 85 to 101 years, participating in a study examining cardiovascular risk factors for cognitive decline and dementia in the oldest old.

Methods. Subjects. Data are from an ongoing, NIA-funded, longitudinal research project of Mount Sinai’s ADRC investigating cardiovascular risk factors in the oldest old. Subjects were healthy volunteers. Investigators were blind to subjects’ cardiovascular status at the time of recruitment. The study was approved by the Institutional Review Board of the Mount Sinai School of Medicine, NY, and all subjects signed informed consent.

Subjects were recruited after talks on memory at senior centers in the tri-state area (New York, New Jersey, and Connecticut), or through newspaper advertisements. In addition, subjects were asked to invite acquaintances to participate in the study. Volunteers who were 85 years and older and who reported having no memory problems were visited and assessed at their residences (none of which were institutions, although this was not a requirement). After written informed consent was obtained, a Clinical

Additional material related to this article can be found on the *Neurology* Web site. Go to www.neurology.org and scroll down the Table of Contents for the September 26 issue to find the title link for this article.

From the Departments of Psychiatry (M.S.B., J.S., M.S., J.W., R.L., H.G., J.M.S.) and Biomathematical Sciences (J.S.), Mount Sinai School of Medicine; and Bronx Veterans Affairs Medical Center (M.S., H.G., J.M.S.), New York, NY.

Supported by NIA grants 1 K01 AG023515-01A2 (M.S.B.), P50-AG05138 (M.S.), and P01-AG02219 (V.H.).

Disclosure: The authors report no conflicts of interest.

Received August 9, 2005. Accepted in final form May 10, 2006.

Address correspondence and reprint requests to Dr. Beeri, Mount Sinai School of Medicine, Department of Psychiatry, One Gustave Levy Place, Box 1230, New York, NY 10029; e-mail: michal.beeri@mssm.edu

1006 Copyright © 2006 by AAN Enterprises, Inc.

Copyright © by AAN Enterprises, Inc. Unauthorized reproduction of this article is prohibited.

Dementia Rating scale (CDR)⁹ score was obtained based on information from both the subject and an informant. Only subjects with CDR = 0 (not demented) were included in the study. Those with CDR \geq 0.5 (questionable dementia or dementia) were invited to participate in other ADRC projects. Because the number of enrolled African American and Hispanic subjects was too small at the time of data analysis to support normative inferences, only the normative data for the white subjects are presented here. Eighty-eight percent of the participants were born in the United States and the rest lived in the United States since early childhood. There were no significant differences in neuropsychological scores between these two groups. The 196 subjects who 1) satisfied strict entry criteria, i.e., exclusion of serious neurologic, medical, and psychiatric disorders that could affect cognition, 2) who were 85 years and older at entry into the study, and 3) spoke English as their first language, participated in this study.

Clinical assessments. The interview included neuropsychological testing, medical and family history, drug inventory, sociodemographic, lifestyle, and dietary information, the Geriatric Depression Scale, and a blood draw. Those with CDR score equal to zero but with a discrepancy between informant and subject's report, raising doubt about their cognitive intactness, were referred to the ADRC Clinical Core for a full dementia workup. All subjects so referred (n = 6) had their non-dementia status (CDR = 0) confirmed.

Neuropsychological testing. CERAD neuropsychological assessment battery was administered to each subject by a certified psychometrician. The battery was designed to assess the basic cognitive functions affected in AD and included the various measures listed below, presented in the order of administration within the battery. The subtest designations that will be used throughout the text and tables are also provided.

Verbal Fluency Test (Fluency). This test measures verbal production, semantic memory, and language.¹⁰ It requires the subject to name as many examples of the category "animal" as possible in 1 minute.

Boston Naming Test (Boston). This test measures visual naming and presents 15 line drawings of common objects from the Boston Naming Test.¹¹ These items are stratified into three groups of five items each, representing objects of high (easy to name), medium, and low (hard to name) frequency of occurrence in the English language. The maximum score is 15.

Mini-Mental State Examination (MMSE). This is a general cognitive screening test that measures orientation, language, concentration, constructional praxis, and memory.² The maximum score on the test is 30.

Word list memory (Trial 1, Trial 2, Trial 3, Total trials 1-3). This is a free recall memory test that assesses learning ability for new verbal information. Participants are presented 10 unrelated items to remember on printed cards. They are instructed to read aloud each word as it is presented. Immediately following presentation of the 10 words, the participant is asked to recall as many items as possible. On each of the three learning trials, the 10 words are presented in a different order. The maximum score on each trial is 10. The maximum total score is 30.

Constructional Praxis (Praxis). This task is part of the Alzheimer's Disease Assessment Scale (ADAS¹²). It measures visuospatial and constructional abilities and requires the subject to copy four line drawings presented in order of increasing complexity (circle, diamond, overlapping rectangles, and cube). The total possible score is 11.

Word list recall (Delayed). This test assesses the ability to recall, after 15 minutes, the 10 words given in the word list memory test. The maximum number of correct responses is 10. For each subject, a saving score (Savings) can be calculated and is presented as a percentage reflecting the relative amount of verbal information retained over the delay interval ($(\text{Delay/Trial } 3) \times 100 = \text{Savings}$).

Word list recognition. This test counts the number of 10 words presented in the word list memory task correctly recognized (Rec-yes). These words are presented among 10 distractor words. The number of distractor words correctly identified (Rec-no) is also counted. The maximum score for each is 10.

Data collection for these subjects was performed in the context of a longitudinal ADRC study that was initiated using the original CERAD neuropsychological battery. Although Constructional Praxis Recall was subsequently added to the CERAD battery, it was not included in the version used in this study.

Trail Making Test. Since the CERAD battery lacked speed and flexibility components, we also assessed the Trail Making Tests. The Trail Making tests¹³ measure timed attention, visual scanning, and sequencing. Part A (Trails A) entails connecting randomly ordered numbers by drawing a line in sequence and has a strong motor speed and agility component. Part B (Trails B) entails connecting numbers and letters in alternating order and adds a strong complex set shifting component reflecting mental flexibility. The psychometrician stopped each test after 5 minutes. The times to complete each of the tasks were used as the two measures for analyses. Although the numbers of errors were also recorded, they are not analyzed here.

Statistical analysis. To extend the 5-year age intervals reported in other normative studies,¹³ the sample was divided into subgroups, 85 to 90 years old and 90 and above. Subjects of the current study tended to have high levels of education, similar to other research studies of this nature.² For this reason the sample was also divided by those with only a high school education or less (12 years of education and below) and those with at least some post high school education (13 years of education and above). We also divided the sample by sex. Comparison of the variables for descriptive purposes was done using *t* test. In order to provide norms, these three dichotomies were used to create eight categories described below.

Half of the variables had distributions that satisfied the normality assumption. Rec-no had over 90% of its observations at the ceiling, so transformations would not normalize its distribution. For simplicity, no transformation was applied to this variable. Its percentiles (reported in table E-1 on the *Neurology* Web site at www.neurology.org) were censored to a maximum of 10.0 reflecting its ceiling. Square root or logarithmic transformations were applied to the other variables.

For each neuropsychological test, we performed multiple regression analyses using as independent variables continuous age and education, and sex. Use of continuous age and education exploits the full range of these variables rather than using arbitrary groups for tests of significance. In a preliminary step, the interactions of these three predictors were tested by stepwise regression. Only three of the interactions were statistically significant (two pertained to the problematic variable Rec-no). Since it is likely that three results pass the 0.05 *p* value threshold by chance when performing 56 tests of significance of interactions, it was concluded that interactions would not be included when fitting the norms. The residuals from all regression analyses had skewness and kurtosis below 1 in absolute value, except for Rec-no.

To produce norms for subgroups of the population, it is necessary to categorize the continuous variables. Age and education were dichotomized as described above, and used with sex to create eight categories. Thus, three pairs of values were created: young age (mean = 87.2) and old age (mean = 92.7), male and female, and low education (mean = 11.4) and high education (mean = 16.8). Combinations of the three pairs of values were created as eight representative hypothetical individuals: 1) low education, young age, male; 2) low education, young age, female; 3) low education, old age, male; 4) low education, old age, female; 5) high education, low age, male; 6) high education, low age, female; 7) high education, high age, male; 8) high education, high age, female.

For each neuropsychological test, transformed if necessary, the regression equation was used to estimate prediction intervals for the scores of these eight representative hypothetical individuals. The 10th, 25th, 50th, and 75th percentiles reported in table E-1 were obtained from these prediction intervals. For the transformed variables, their estimated percentiles were subjected to the inverse transformation to provide percentiles in the scales of the original distributions. No variable except Rec-no had a ceiling that affected the reported percentile. All statistical analyses were done using SPSS version 13.0.

Results. Characteristics of the sample. Table 1 presents the demographic data for the high and low education groups. Within each of the education groups, men and women did not differ in their age and education. Average ages did not differ between the low and high education

Table 1 Demographic variables for total sample

	≤12 Years of education			>12 Years of education			Total sample, n = 196
	Men, n = 16	Women, n = 51	Total for low education, n = 67	Men, n = 50	Women, n = 79	Total for high education, n = 129	
Age, y							
Mean (SD)	89 (3.9)	89 (2.8)	89 (3.1)	89 (2.8)	90 (3.5)	89 (3.2)	89 (3.2)
Range	85–99	85–97	85–99	85–99	85–101	85–101	85–101
Education, y							
Mean (SD)	10.8 (1.8)	11.6 (1.3)	11.4 (1.4)	17.1 (1.7)	16.6 (2.3)	16.8 (2.1)	14.9 (3.2)
Range	6–12	5–12	5–12	13–20	13–20	13–20	5–20

groups for men, women, and all subjects. The two education groups differed by 5.4 years on average.

Multiple regression analyses results. Table 2 presents, for each of the neuropsychological tests, the mean and SD of the whole sample for the original variables, and the multiple regression results, using transformed variables when necessary. For 11 out of the 14 neuropsychological tests, the multiple correlation model consisting of education, age, and sex as predictors was significant (see third column, table 2). Among these significant tests, eight were significant for education (MMSE, Trial 1, Trial 2, Trial 3, total Trials 1–3, fluency, praxis, and Trails B), seven for age (Trial 1, Trial 2, total Trials 1–3, rec-no, Boston, Trails A and Trails B), and four for sex (MMSE, Trial 1, total Trials 1–3, and Trails B); for seven tests, education had the largest β magnitude. In all significant results, lower age and higher education were associated with better performance. Among the tests with significant results by sex, women performed better than men, except for Trails B (for which a positive β describes higher response times and thus worse performance for women).

Table E-1 presents the estimated percentiles of neuropsychological tests for the oldest old by combinations of education, age, and sex. For example, for men with low education, 85 thru 89 years old, the estimated performance at the 10th percentile for the MMSE is 24.5, at the 25th percentile is 26.2, at the 50th percentile is 27.7, and at the

75th percentile is 28.8. For women with high education, 90 years old and above, the estimated performance at the 10th percentile of the Trails B is 281 seconds, at the 25th percentile is 228 seconds, at the 50th percentile is 177 seconds, and at the 75th percentile is 131 seconds. It should be noted that over 90% of the subjects had the ceiling value of 10 on Rec-no. This violation of the assumption of normality impairs the validity of the tests of significance in table 2 and the percentiles in table E-1 for this variable.

Discussion. Neuropsychological tests measured in the CERAD battery as well as Trail Making tests A and B are affected differentially by age, sex, and education. The sample is highly educated with an average of 15 years of education and many of those in the lower education group had at least some high school education. Thus, these performance levels should be interpreted with caution when applying them to epidemiologic studies, especially since education was clearly the most influencing variable for the neuropsychological test scores in this very elderly sample. However, the mean and SD of education for the current sample are essentially identical to those of 85+ subjects of the National Alzheimer's Coordinating Center (NACC; mean = 15.1; SD =

Table 2 Means (SD) and multiple regressions predicting neuropsychological tests from dichotomized age, sex, and education

Neuropsychological test* (range)	Mean (SD)	R	β for age	β for sex	β for education
MMSE (0–30)	28.0 (1.5)	0.27†	–0.12	0.15‡	0.23†
Trial 1 (0–10)	4.8 (1.6)	0.32†	–0.20†	0.21†	0.20†
Trial 2 (0–10)	6.7 (1.5)	0.29†	–0.18‡	0.08	0.25†
Trial 3 (0–10)	7.4 (1.4)	0.21‡	0.04	0.11	0.19‡
Total Trials 1–3 (0–30)	19.0 (3.8)	0.30†	–0.14‡	0.16‡	0.25†
Delayed recall (0–10)	5.8 (1.7)	0.16	0.03	0.11	0.13
Savings (0–)§	78.8 (24.4)	0.09	–0.03	0.08	0.03
Rec–yes (0–10)	9.3 (1.3)	0.16	–0.02	0.03	0.16‡
Rec–no (0–10)	9.9 (0.31)	0.24‡	–0.22†	0.00	0.11
Fluency (0–)	15.1 (4.7)	0.35†	–0.10	–0.01	0.34†
Boston (0–15)	13.4 (1.4)	0.21‡	–0.15‡	–0.08	0.12
Praxis (0–11)	9.2 (1.8)	0.28†	0.02	–0.03	0.27†
Trails A (0–300 sec)	69.0 (34.8)	0.23‡	0.21†	0.02	–0.11
Trails B (0–300 sec)	166.7 (73.1)	0.31†	0.21†	0.15‡	–0.18‡

* Square root transformations were applied to Mini-Mental State Examination (MMSE), Fluency, Praxis, Rec–yes, and Trails B. Logarithmic transformations were applied to Savings and Trails A.

§ Savings = (Delayed/Trial 3) * 100. Since subjects might recall more words on Delayed Recall than in Trial 3, this measure might exceed 100%.

† $p < 0.005$; ‡ $p < 0.05$.

3.5¹⁴). NACC entrants are usually recruited from Memory Clinic clients, so these results may be useful also in tertiary care clinical settings.

Immediate recall (measured by Trial 1, Trial 2, Trial 3, and total Trials 1–3) was affected mainly by education, but also by age and sex, in contrast to delayed recall and savings. Delayed recall and savings are associated with subsequent conversion to AD.^{15,16} Thus, these memory measures, which have been found to be most sensitive in identifying dementia and AD, are unaffected by demographic characteristics in the oldest old. This is consistent with results found for younger samples, with broader ranges of age and education.^{2,17}

The lack of an age effect on the Fluency measure conflicts with some reports indicating age-related decline in verbal production^{6,18} but is in agreement with others¹⁹ including the CERAD report.² One study¹⁸ had a substantially larger age range, possibly increasing the opportunity of finding an age effect. However, other studies,^{2,19} that had a larger age range than that of the current study, found no age effect. All studies used exactly the same procedure. It is possible that Fluency, a strongly language-oriented measure, behaves differently in different languages. Accordingly, the older Japanese samples (above the age of 80) that had similar levels of education to our low education sample had a lower weighted average Fluency score (11.1) than the median value (12.9) of the representative hypothetical individual that scored, in most instances, the lowest scores of the sample in the current study (90 years and above male with a low level of education). This may reflect the effects of culture on the performance of this test. Since the sample of the current study includes only white oldest old whose primary language is English, extrapolation of our findings to oldest old from different ethnic groups may be imprecise.

Trail Making tests A and B were included to reinforce both the psychomotor component of the battery (Trails A) as well as the executive functions component (Trails B). Both parts of the test were age dependent, i.e., the older the subjects, the longer the time to complete the task. Also, men were faster than women in both parts of the test. However, only Trails B was influenced also by education. Recently, a reduction in the maximum time allowed to perform the test from 5 minutes to 3 minutes has been suggested¹⁴ in order to limit the frustration of subjects as well as the costs and total administration time in large-scale studies. Based on the current study's result (where the median estimates of Trails B range from 150 to 200 seconds), decreasing the time to finish the test would result in a substantial increase of a ceiling effect.

Over 90% of the subjects had a ceiling score of 10 for Rec-no. Although the percentiles presented in table E-1 are problematic, the rarity of a score of 9 or below suggests potential impairment.

The estimates for any representative hypothetical

individual (such as young-old male with low education) were not based on the subgroup of subjects in this study with that combination of characteristics. Instead it was based on the fitted regression model for a hypothetical subject with that combination of characteristics. Since the model was estimated using all 196 subjects, the estimation for subgroups with smaller sizes, especially men with no more than high school, were more stable than if estimates for each group were based only on subjects in that group. However, a limitation of this procedure is that it includes only linear effects of age and education.

Our findings suggest that neuropsychological testing is influenced mainly by education and age despite the relatively high education levels of the sample and the very old and relatively narrow age range. Cutoff scores based on younger populations and applied to the oldest old might lead to increased false-positive misclassifications. The thresholds for cognitive impairment are consistently lower for the older age ranges described in this study compared to the previously published CERAD norms for younger age ranges. For example, a 90-year-old man with a high level of education would be still considered unimpaired with a score of 11 in the Fluency test, based on the current results (using the 10th percentile), but would be considered impaired based on norms for highly educated but slightly younger individuals.² Similarly, an 85-year-old woman with a high level of education would be considered unimpaired based on the current results with a score of 55% in the Savings measure (using the 10th percentile) but would be considered impaired by the original CERAD norms.²

References

1. Whyte SR, Cullum CM, Hynan LS, Lacritz LH, Rosenberg RN, Weiner MF. Performance of elderly Native Americans and Caucasians on the CERAD Neuropsychological Battery. *Alzheimer Dis Assoc Disord* 2005; 19:74–78.
2. Welsh KA, Butters N, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* 1994;44:609–614.
3. McCurry SM, Gibbons LE, Uomoto JM, et al. Neuropsychological test performance in a cognitively intact sample of older Japanese American adults. *Arch Clin Neuropsychol* 2001;16:447–459.
4. Bertolucci PH, Okamoto IH, Brucki SM, Siviero MO, Toniolo NJ, Ramos LR. Applicability of the CERAD neuropsychological battery to Brazilian elderly. *Arq Neuropsiquiatr* 2001;59:532–536.
5. Guruje O, Unverzagt FW, Osuntokun BO, et al. The CERAD Neuropsychological Test Battery: norms from a Yoruba-speaking Nigerian sample. *West Afr J Med* 1995;14:29–33.
6. Fillenbaum GG, McCurry SM, Kuchibhatla M, et al. Performance on the CERAD neuropsychology battery of two samples of Japanese-American elders: norms for persons with and without dementia. *J Int Neuropsychol Soc* 2005;11:192–201.
7. Fillenbaum GG, Heyman A, Huber MS, Ganguli M, Unverzagt FW. Performance of elderly African American and White community residents on the CERAD Neuropsychological Battery. *J Int Neuropsychol Soc* 2001;7:502–509.
8. Ganguli M, Ratcliff G, Huff FJ, et al. Effects of age, gender, and education on cognitive tests in a rural elderly community sample: norms from the Monongahela Valley Independent Elders Survey. *Neuroepidemiology* 1991;10:42–52.
9. Fillenbaum GG, Peterson B, Morris JC. Estimating the validity of the clinical Dementia Rating Scale: the CERAD experience. Consortium to Establish a Registry for Alzheimer's Disease. *Aging (Milano)* 1996;8: 379–385.
10. Newcombe F. *Missile wounds of the brain*. London: Oxford University Press; 1969.

11. Kaplan E, Goodglass H, Weintraub S. The Boston Naming Test. Philadelphia: Lea and Febiger; 2005.
12. Mohs RC, Rosen WG, Davis KL. The Alzheimer's disease assessment scale: an instrument for assessing treatment efficacy. *Psychopharmacol Bull* 1983;19:448-450.
13. Spreen O, Strauss E. A compendium of neuropsychological tests: administration, norms, and commentary. Second edition. New York: Oxford University Press, 1998.
14. National Alzheimer's Coordinating Center. Available at: www.alz.washington.edu
15. Welsh KA, Butters N, Hughes JP, Mohs RC, Heyman A. Detection and staging of dementia in Alzheimer's disease. Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Arch Neurol* 1992;49:448-452.
16. Tierney MC, Yao C, Kiss A, McDowell I. Neuropsychological tests accurately predict incident Alzheimer disease after 5 and 10 years. *Neurology* 2005;64:1853-1859.
17. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA* 1993;269:2386-2391.
18. Wiederholt WC, Cahn D, Butters NM, Salmon DP, Kritz-Silverstein D, Barrett-Connor E. Effects of age, gender and education on selected neuropsychological tests in an elderly community cohort. *J Am Geriatr Soc* 1993;41:639-647.
19. Stricks L, Pittman J, Jacobs DM, Sano M, Stern Y. Normative data for a brief neuropsychological battery administered to English- and Spanish-speaking community-dwelling elders. *J Int Neuropsychol Soc* 1998;4:311-318.

ELECTRONIC HEALTH RECORDS VENDOR REPORT AVAILABLE

The AAN has scrutinized the field of electronic health records (EHR) systems and prepared a report on the vendors that meet the Academy's criteria. Designed as a guide for small and solo neurology practices, the report is available at www.aan.com/ehrvendors.