

UC Davis

UC Davis Previously Published Works

Title

Effects of Education and Race on Cognitive Decline: An Integrative Study of Generalizability Versus Study-Specific Results

Permalink

<https://escholarship.org/uc/item/0t62w93c>

Journal

Psychology and Aging, 30(4)

ISSN

0882-7974

Authors

Gross, Alden L
Mungas, Dan M
Crane, Paul K
[et al.](#)

Publication Date

2015-12-01

DOI

10.1037/pag0000032

Peer reviewed



Published in final edited form as:

Psychol Aging. 2015 December ; 30(4): 863–880. doi:10.1037/pag0000032.

Effects of education and race on cognitive decline: An integrative analysis of generalizability versus study-specific results

Alden L. Gross^{1,*}, Dan M. Mungas², Paul K. Crane³, Laura E. Gibbons³, Anna MacKay-Brandt⁴, Jennifer J. Manly⁵, Shubhabrata Mukherjee³, Heather Romero⁶, Bonnie Sachs⁷, Michael Thomas², Guy G. Potter⁶, and Richard N. Jones⁸

¹ Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; Center on Aging and Health, Johns Hopkins University, Baltimore, MD, USA

² Department of Psychiatry, University of California, San Diego, CA, USA

³ Department of General Internal Medicine, University of Washington, Seattle, WA, USA

⁴ Department of Neurology, Division of Cognitive Neuroscience, Columbia University, New York, NY, USA; Department of Psychiatry, Columbia University, New York, NY, USA

⁵ Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Department of Neurology, and the Gertrude H. Sergievsky Center, College of Physicians and Surgeons, Columbia University, New York, NY, USA

⁶ Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA

⁷ Department of Neurology, Wake Forest Baptist Medical Center, Winston Salem, NC, USA

⁸ Departments of Psychiatry and Human Behavior and Neurology, Warren Alpert Medical School, Brown University, Providence, RI, USA

Abstract

Objective—To examine variability across multiple prospective cohort studies in level and rate of cognitive decline by race/ethnicity and years of education.

Method—To compare data across studies, we harmonized estimates of common latent factors representing overall or general cognitive performance, memory, and executive function derived from the: 1) Washington Heights, Hamilton Heights, Inwood Columbia Aging Project (N=4,115), 2) Spanish and English Neuropsychological Assessment Scales (N=525), 3) Duke Memory, Health, and Aging study (N=578), and 4) Neurocognitive Outcomes of Depression in the Elderly (N=585). We modeled cognitive change over age for cognitive outcomes by race, education, and study. We adjusted models for sex, dementia status, and study-specific characteristics.

Results—For baseline levels of overall cognitive performance, memory, and executive function, differences in race and education tended to be larger than between-study differences and

*Corresponding author: Alden L. Gross, agross14@jhu.edu; Phone: 410-474-3386; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 616 N. Wolfe St., Baltimore, MD 21205, USA..

consistent across studies. This pattern did not hold for rate of cognitive decline: effects of education and race/ethnicity on cognitive change were not consistently observed across studies, and when present were small, with racial/ethnic minorities and those with lower education declining at faster rates.

Discussion—In this diverse set of datasets, non-Hispanic whites and those with higher education had substantially higher baseline cognitive test scores. However, differences in the rate of cognitive decline by race/ethnicity and education did not follow this pattern. This study suggests that baseline test scores and longitudinal change have different determinants, and future studies to examine similarities and differences of causes of cognitive decline in racially/ethnically and educationally diverse older groups is needed.

Keywords

cognitive performance; item response theory; confirmatory factor analysis; harmonization; cognitive trajectory; race/ethnicity; education

Introduction

Cognitive impairment in older adults is a public health challenge that is increasing in magnitude as the older population grows. Understanding the risk factors for cognitive decline is critical for minimizing the impact of cognitive impairment on affected individuals, their families, and society. The population of older adults in the United States is becoming increasingly diverse with respect to race, ethnicity, and educational background (United States Census Bureau, 2012). Prior work has shown that race is associated with level of cognitive performance, although such associations are inconclusive: low test scores at one occasion may be a function of other factors such as early life exposures (e.g., education) (Glymour et al., 2005; Reardon & Galindo, 2007). Cross-sectional studies of older adults have consistently shown lower test scores in black and Hispanic older adults in comparison with non-Hispanic whites (Alley et al., 2007; Atkinson et al., 2005; Early et al., 2013; Herzog & Wallace, 1997; Karlamangla et al., 2009; Leveille et al., 1998; Masel & Peek, 2009; Rodgers et al., 2003; Schwartz et al., 2004; Sisco et al., 2014; Sloan & Wang, 2005; Wolinsky et al., 2011; Zsembik & Peek, 2001). Lower education levels similarly have been associated with lower test scores (Albert et al., 1995; Alley et al., 2007; Atkinson et al., 2005; Early et al., 2013; Manly et al., 2004; Mungas et al., 2000, 2005; Schiave, 1996; Schneider et al., 2012).

Independent of its relationship to level, race and education could also moderate longitudinal cognitive decline. Whether differences in longitudinal cognitive change parallel cross-sectional differences is an important question because, although cross-sectional differences and longitudinal change may be indicative of both brain disease as well as early and enduring social inequities, associations involving the former are often confounded by cohort differences (Schiave, 2005, 2008; Schiave et al., 2005; Willis, 1989) and other factors (Glymour et al., 2005). Several potential reasons might explain why race or education may contribute to cognitive decline. Among them, minorities and persons with lower education are at elevated risk for medical comorbidities such as hypertension and diabetes that may contribute to accelerated cognitive decline (Carter et al., 1996; Duggan et al., 2014; Zhao et

al., 2014). Further, social contexts including geographic exposures, socioeconomic position, and group identity may influence cognitive level over time (Glymour & Manly, 2008).

Longitudinal studies of the association of cognitive decline with race/ethnicity or education have contributed contradictory results (Alley et al., 2007). Many longitudinal studies have shown no association of race/ethnicity (Carvalho et al., 2014; Early et al., 2013; Masel & Peek, 2009; Schneider et al., 2012; Sisco et al., 2014; Wilson et al., 2010) or education (Early et al., 2013; Glymour et al., 2005; Johansson et al., 2004; Schneider et al., 2012; Tucker-Drob et al., 2009; Wilson et al., 2009; Zahodne et al., 2011) with rate of cognitive decline. Some have shown faster decline in racial and ethnic minorities (Albert et al., 1995; Wolinsky et al., 2011), while other studies have shown faster decline in non-Hispanic whites (Barnes et al., 2005; Sloan et al., 2005). Similarly, some have reported accelerated decline in persons with less education (Bosma et al., 2003; Butler et al., 1996; Lyketsos et al., 1999; Sachs-Ericsson et al., 2005), while other studies report the opposite (Johansson et al., 2004; Scarmeas et al., 2006).

Clarifying whether race/ethnicity and education are related to cognitive decline, and if these findings are consistently found in studies of different samples utilizing different methods, is important for understanding causes of cognitive decline in the increasingly diverse older population. Cognitive decline in older persons is the cardinal indicator of progressive brain diseases of aging. If cognitive change varies by race/ethnicity and education, this would support the hypothesis that health disparities associated with minority status and lower education account for differences in underlying brain diseases that ultimately explain differences in cognitive performance. Alternately, cross-sectional effects on cognition in the context of no association with longitudinal decline would suggest baseline differences might be due to various life experience and opportunity differences but are not related to disparities in diseases of aging that cause cognitive decline.

Inconsistencies across studies in associations of race/ethnicity and education with cognitive decline could be due to methodological factors such as differing sampling strategies across studies, regional variability among subgroups enrolled in specific studies, and use of different cognitive tests that vary in their sensitivity to cognitive decline. Large scale studies using representative samples from different geographic regions and assessing cognition using the same tests would be especially relevant to clarifying how race/ethnicity and education affect cognitive decline. However, there are few nationally representative studies with longitudinal cognitive assessment that have sampled large numbers of older adults. Two such studies are the Health and Retirement Study (HRS; Juster & Suzman, 1995) and the Midlife in the US study (Lachman & Firth, 2004). Although these studies provide important resources for answering question about demographic influences on late life cognition, their measurements of cognition are relatively limited. In the absence of nationally representative studies with comprehensive cognitive assessment, an alternate approach is to combine data from other studies. Common findings across diverse samples likely represent robust determinants of successful or unsuccessful cognitive aging. Divergent findings are also informative because they might suggest more nuanced differences attributable to sampling and other methodological differences between studies. However,

comparable measurement of cognition is an important prerequisite for this strategy to be effective in meaningfully comparing results across studies.

A method for harmonizing, or mapping cognitive tests from different studies onto representations of general and specific cognitive domains, would represent a major advance towards addressing this problem. Merging cognitive test results from multiple studies that use different tests presents a major methodological barrier to determining whether race/ethnicity or education predicts cognitive change (Bauer et al., 2010; Curran et al., 2010). Both Alley and colleagues (2007) and Manly and colleagues (1998) found that the association of education with cognitive level and change differed depending on the test. Even if the same cognitive abilities are measured in different studies (e.g., episodic memory, attention), specific tests of these abilities may have different reliabilities and floor/ceiling effects. These differing psychometric properties make it difficult to distinguish true differences from biases introduced by psychometric artifacts.

The goal of the current study was to implement a method for harmonizing cognitive tests across studies conducted in different regions of the US to better understand the effects of race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic) and education on cognitive trajectories. We combined longitudinal data from 5,803 older adults collected across four studies. We derived precise measurements of overall or general cognitive performance, memory, and executive function from differing but overlapping sets of cognitive tests using methods based in item response theory (IRT; Embretson & Reise, 2000) and confirmatory factor analysis (CFA; Brown, 2006). The harmonized dataset allowed us to examine similarities and differences in the effects of race/ethnicity and education on cognitive trajectories across the four samples. We then used latent growth models to evaluate trajectories of cognitive performance. Because prior research has consistently shown cross-sectional differences but has been inconsistent regarding associations of race and education with cognitive change in older adulthood, we hypothesized that cross-sectional effects of race and education would be robust across studies. However, differences among studies in the rate of cognitive decline would be relatively large, that is, different patterns of longitudinal change would be observed. If correct, this would be consistent with the notion that longitudinal findings are more sensitive to sample and methodology differences which might point the way toward a better understanding of differential influences on cross-sectional scores and cognitive decline.

Methods

Participants

We used data from four longitudinal studies of older adults with a combined sample size of 5,803 persons and 22,510 study visits. Studies included the Washington Heights, Hamilton Heights, Inwood Columbia Aging Project (WHICAP) (Tang et al., 2001); Spanish and English Neuropsychological Assessment Scales (SENAS) study (Mungas et al., 2004); Duke Memory Health and Aging (MHA) study (Carvahlo et al., 2014); and Neurocognitive Outcomes of Depression in the Elderly (NCODE) (Steffens et al., 2004). Longitudinal cognitive data are available in all studies. WHICAP is a study of community-living older adults representatively sampled from northern Manhattan. Participants, identified using three

contiguous US consensus tracts in Northern Manhattan, New York, and in Washington/Hamilton Heights and Inwood neighborhoods, were non-Hispanic white, non-Hispanic black, or Hispanic older adults of mostly Caribbean origin. SENAS is a study of cognitive function in an ethnically and educationally diverse sample of older adults living in northern CA. The study enrolled white, black, and Mexican Hispanic older adults. The Duke MHA study is a longitudinal study of normal aging, mild cognitive disorders, and Alzheimer's disease in individuals 55 years of age or older in Durham, NC and surrounding areas. Participants were recruited from both clinical and community samples through the Bryan Alzheimer's Disease Research Center. NCODE is a study of depression in older adults from clinics in Durham, NC that also included non-depressed control participants. Duke MHA and NCODE recruited mostly non-Hispanic white and black older adults.

To avoid reporting results on an internally scaled, nongeneralizable metric, we externally scaled the harmonized summary measures to a nationally representative metric by taking advantage of overlapping cognitive tests across the different studies and the neuropsychological battery used by the Aging, Demographics, and Memory Substudy (ADAMS) using a published procedure (Gross et al., 2014a, 2014b). ADAMS is a nationally representative study of dementia that recruited 856 adults over age 70 living in the United States interviewed in 2002-2004 (Langa et al., 2005). Its parent study, the Health and Retirement Study (HRS), is an ongoing longitudinal survey of community-living retired persons (Juster & Suzman, 1995).

Variables

Neuropsychological test batteries—The ADAMS neuropsychological battery included 10 tests described here. Neuropsychological tests from other studies are listed in Table 1. Detailed descriptions are available in citations provided. The Trail-Making Test is a two-part timed test of processing speed (Part A) and processing speed and task-switching ability (Part B) that requires one to connect an ordered series of numbers and an alternating ordered series of letters and numbers, respectively (Reitan, 1958). Digit Span Forward and Backward are measures of auditory attention and short-term memory store (forward), with an additional manipulation requirement (backward) (Wechsler, 1987). Semantic fluency measures language ability via speeded retrieval of words by requiring participants to name as many items as possible in one minute from a preselected category (e.g., animals) (Benton et al., 1976). Letter/phonemic fluency measures language ability and executive function; administration is similar to that for semantic fluency but involves speeded retrieval of words that begin with specific letters. The Symbol Digit Modalities Test is a timed test of processing speed and attention that requires respondents to match numbers to a series of symbols as fast as possible (Smith, 1973). A 10-noun word recall task from the Consortium to Establish a Registry for Alzheimer's Disease battery was used to measure memory (Morris et al., 1989). We used the sum of three immediate recall trials, and the delayed recall trial from this test. The 15-item Boston Naming Test is a language task in which participants name objects represented in a series of drawings (Williams et al., 1989). The Logical Memory test is a test of memory for short stories; we used the immediate and delayed total recall for stories I and II. The Mini-mental state examination (MMSE) is a test of global mental status that consists of items that assess registration and recall, visuospatial

ability, orientation to time and place, attention and calculation, complex commands, and language (Folstein et al., 1975).

Race—In each study, participants self-reported their race as white, black or African American, or other categories. Participants were also asked whether they self-identified as Hispanic. We included persons who reported being non-Hispanic white, non-Hispanic black, or Hispanic. We excluded persons of other racial/ethnic groups including American Indian/Eskimo/Aleutian islanders, Asian and Pacific islanders, and others (N=77).

Education—Participants in each study reported the number of years spent in formal education. We classified years of education into levels of 12 years or less (high school), 13 to 16 years (college), and more than 16 years (graduate).

Descriptive and adjustment variables—Variables that were available in all four studies included sex, baseline age, years of education, and clinical diagnostic status (dementia, no dementia). We also adjusted for depressive symptoms in NCODE.

Analysis plan

The integrative data analysis was conducted in three steps. First, to describe demographic characteristics of the individual and pooled study samples, we calculated descriptive statistics including means and proportions. Second, we derived factor scores for general or overall cognitive performance, memory, and executive function across study samples and cognitive tests. The approach is consistent with the moderated non-linear factor model approach (Bauer & Hussong, 2009; Curran et al., 2014). Third, we examined effects of race and ethnicity on cognitive trajectories across datasets using parallel process latent growth models.

Derivation of cognitive factors—Each study administered different but overlapping sets of cognitive tests. We scaled estimates of neuropsychological abilities for overall cognitive performance, memory, and executive function to a nationally representative metric using data from ADAMS HRS using previously published procedures (Gross et al., 2014a, 2014b). In a pre-statistical harmonization, we first identified tests that were common and unique to each dataset and assigned tests to cognitive domains based on their content (Griffith et al., 2013). The overall cognitive performance factor represents a summary of correlated test results from all tests, and thus is a psychometrically justifiable and meaningful way of summarizing test performance. Cognitive tests were selected for each domain based on expert opinion of a group of clinical psychologists with expertise in aging (AMB, JJM, DMM, HRR, BS). We also conducted parallel analyses with scree plots of tests in each domain to confirm the indicators were sufficiently unidimensional (Hayton et al., 2004) (see Supplemental Figures 13-16). Cognitive test scores were discretized into nine or ten categories to facilitate estimation of models and to avoid modeling of skewed data. Four of the cognitive test scores in the ADAMS HRS battery were from the Logical Memory test, so in the model for overall cognitive performance, we added a specific factor (Supplemental Figure 1) to allow residual correlations between the items to ensure the general factor was not overly influenced by variability in Logical Memory (McDonald et al., 1999). We call the

latent trait measured by all the items overall cognitive performance, consistent with prior work (Gross et al., 2014a, 2014b; Jones et al., 2010; Sisco et al., 2014).

In the second phase of harmonization, we used two-parameter logistic graded response IRT models (Samejima, 1969) for each cognitive domain to obtain item parameters for cognitive tests in the ADAMS HRS battery. Separate models were estimated for overall cognitive performance, memory, and executive function (Supplemental Figure 1); individual tests are listed in Table 2. All cognitive test scores were indicators for the overall cognitive factor, memory tests were used for the memory factor, and executive function tests were used for the executive function factor. IRT facilitates item and test equating, linking, and evaluation of item bias through item parameters that describe location and discrimination (Jöreskog & Moustaki, 2001; Lord, 1953, 1980; McArdle et al., 2009; McHorney, 2003). Parameters in ADAMS were estimated using a maximum likelihood estimator with robust standard error estimation (MLR) using Mplus (version 7.11, Muthen & Muthen, Los Angeles CA, 1998-2008) software and made use of available complex sampling weights. Latent factors were fixed to a standardized metric in the nationally representative sample to identify the model (mean 0, variance 1).

In a third harmonization phase, in a pooled dataset with all the studies and time points, we derived plausible values of factor scores representing overall cognitive performance, memory, and executive function with a two-parameter graded response IRT model for each cognitive trait. We ensured linkage to the nationally representative metric by constraining item parameters belonging to common tests in the pooled dataset (e.g., SENAS, WHICAP, Duke MHA, NCODE) and ADAMS HRS to their values estimated in ADAMS from phase 2. Tests in common across datasets anchor the metric in each dataset to that in ADAMS and to one another. Latent variable means and variances were freely estimated in the pooled IRT analysis. The models for each cognitive domain used all longitudinal data and incorporated a clustering term for participants to correct standard errors for within-person correlations; these models are likely to provide factor scores equivalent to those from models that estimate a latent variable for each time point separately. Estimated factor scores from models using all items, memory items, and executive function items were the factor scores representing overall cognition, memory, and executive function, respectively. The factor scores were averages from 30 draws from the Bayesian posterior distribution of plausible values (Asparouhov & Muthén, 2010). These scores were scaled to have a mean of 50 and standard deviation of 10 in the ADAMS reference sample, which is representative of community-living older adults in the US aged 70 and older.

Cognitive test scores were discretized to place all tests on a similar scale (Jones et al., 2010). Cutpoints are shown in Table 3 of the Supplemental Materials. Percentile cutoffs for tests administered in ADAMS are based on weighted data. Percentiles for tests not included in ADAMS were estimated within the contributing study. Thresholds for the study-specific percentile cuts are anchored to the normative ADAMS sample in step 3 by fixing thresholds of tests from ADAMS to the ADAMS metric.

In the final phase of harmonization, we examined the quality of the link between each study and the national metric using (1) simulation and (2) testing for differential item functioning

by study. The simulation, described in Supplemental Materials 1, revealed minimal evidence of bias by dataset across the range of cognitive scores.

Differential item functioning (DIF): We used pooled data to examine and adjust for biases in individual items attributable to data source. Item parameters should have the same relation with the underlying trait (e.g., item discrimination) and location along the latent trait across datasets. We used Bayesian alignment analysis to identify and correct measurement bias as a function of study source (Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2014). Alignment analysis is a type of multiple-group confirmatory factor analysis, available in Mplus version 7.1, in which item loadings and thresholds are systematically tested for differences across multiple groups (e.g., study). The ultimate goal of the method is to facilitate comparisons of a factor across groups while accommodating DIF, and we used it to inform measurement non-invariance in our final models. In the alignment approach, a configural model with invariant parameters across study is estimated (e.g., all items are assumed to be DIF-free). Next, the alignment procedure systematically tests each factor loading and threshold to determine which is different across which dataset; the algorithm seeks the best-fitting final solution that minimizes the number of non-invariant parameters. We conducted parameter testing based on the posterior distribution of the test statistic by specifying a Bayesian estimator (Muthén & Asparouhov, 2014). Alignment analysis is an efficient alternative to Multiple Indicator Multiple Causes models (Jöreskog & Goldberger, 1975), which requires a potentially lengthy sequence of multiple tests guided by model modification indices (Asparouhov & Muthén, 2014). Any DIF, indicated by an association between study assignment and a particular test after controlling for the overall level of cognitive performance and adjustment variables, suggests measurement bias such that the item does not measure the latent cognitive ability in the same way in the different studies. DIF can be accounted for by treating the item as if it were different items across studies, which also allows for estimation of different item parameters. When the predictor is study membership, differential item functioning suggests administration and/or scoring differences that we did not detect in the prestatistical harmonization step, or possibly DIF with respect to some demographic characteristic that differs across studies; these analyses indicate there is a problem but do not diagnose what the problem is. After accounting for DIF, any differences observed in the rate of cognitive decline should be attributable to race or to education, and not study-specific measurement differences (Crane, Narasimhalu, Gibbons et al., 2008; Pedraza & Mungas, 2008).

Models of cognitive performance by race/ethnicity and educational attainment

—In a parallel process latent growth model, growth processes for each cognitive outcome (overall or general cognitive performance, memory, executive function) were characterized by latent variables for an initial intercept, a linear slope or rate of decline, and quadratic growth (Curran & Hussong, 2009; McArdle & Bell, 2000; Muthen, 1997; Muthen & Curran, 1997; Stull, 2008). The timescale of interest was age, which we centered at 75 years to facilitate interpretation of the random intercept term as mean cognitive performance at age 75. Three models were estimated. First, a “null” model was estimated with regressions of the intercept and slope on indicators for years of education, race/ethnicity, sex, dementia status, and in NCODE, depression status based on clinical diagnosis. Second, we added indicators

for each study and interaction terms between study and race/ethnicity to accommodate between-study differences in race/ethnicity effects on the level and change in each cognitive outcome. Third, we repeated the procedure for years of education in place of race/ethnicity. To evaluate fit of models to the data, we used graphical tools to examine residuals and detect outliers (Supplemental Figures 5-6), and calculated the empirical r^2 as the squared correlation between observed and model-estimated cognitive outcomes (Singer & Willet, 2003). In a planned sensitivity analysis, we re-ran all models using a timescale of time since study recruitment; no conclusions differed when we used these models. In a further sensitivity analysis, we evaluated the possibility that correlates of normative cognitive decline and dementia-related decline are different by repeating the entire analysis excluding records of people with dementia; overall conclusions did not differ.

Handling of missing data—In integrative longitudinal data analysis, missing data arise in two ways, by design and by loss to follow-up, and were handled at different steps of the analysis. Missing data due to loss to follow-up was handled during the estimation of parallel process latent growth models, which assume observations are missing at random conditional on variables in the model. This assumption is the most reasonable assumption; violations of it are unlikely to affect the interpretation of our key conclusions regarding whether demographic factors or study membership affects cognitive decline. Missing data for specific cognitive tests due to study design are assumed to be missing at random conditional on variables in the measurement model, and handled using maximum likelihood methods during estimation of the model (Little et al., 2012). The maximum likelihood estimator with robust standard errors in Mplus uses all available data to estimate parameters that can heuristically be thought of as resulting from the average of an infinite number of multiple imputations given the observed data. This is a reasonable approach because missingness is chiefly due to study membership, not to characteristics of people. Thus, to the extent that selection into one study versus another in our collection is random, missingness on a cognitive test because it was not given in a study is not a formidable problem.

Results

Descriptive characteristics

The mean age in the pooled sample was 76 years (range 45, 103 years) (Table 2). Most participants were female (67%), and the mean years of education was 11 (range 0, 22 years). The pooled sample was 38% non-Hispanic white, 30% non-Hispanic black, and 32% Hispanic. Hispanic participants came from the WHICAP study in New York, which included mostly Caribbean Hispanic older adults, and the SENAS study in northern CA, which includes mostly Mexican Hispanics. The median amount of follow-up time in the sample was 4.0 years (range 0, 23 years). The number of study visits ranged from 1 to 36 with a median of 3 in the overall sample. Study-specific mean levels of cognitive performance in each study except WHICAP were above the average of the nationally representative ADAMS HRS, probably reflecting a combination of volunteer recruitment bias and younger age. Mean years of education was highest in the Duke MHA and NCODE datasets. Mean overall cognitive performance was lowest in WHICAP and highest in NCODE.

Derivation and calibration of the cognitive factors

We identified 48 indicators representing 27 cognitive tests administered as part of the neuropsychological batteries across the studies (Table 1). Most indicators (96%) loaded highly (standardized loadings above 0.5) onto the domain they represented (see Table 2 for overall cognitive factor loadings). Of the 48 cognitive indicators, 21 (44%) were in common across two or more studies. Twelve cognitive indicators (Trail Making parts A and B, semantic and phonemic fluency, digit span forwards and backwards, Boston Naming, word recall, constructional praxis, MMSE, and Logical Memory) were administered in at least four of the studies. Two of 22 cognitive indicators in WHICAP were in common with ADAMS HRS. Twelve of 22 cognitive indicators in SENAS were in common with ADAMS HRS. Twelve of 17 cognitive indicators in Duke MHA were in common with ADAMS HRS. Sixteen of 18 cognitive indicators in NCODE were in common with ADAMS HRS. As few as one anchor item has been used to calibrate two scales (Jones & Fonda, 2004), though a simulation study on the topic suggested at least five anchor tests are optimal to produce an accurate linking (Wang et al., 2004).

Using available cognitive test data, we derived summary factors for overall cognitive performance, memory, and executive function across studies on the same metric based on the mean of 30 draws from the Bayesian posterior distribution of plausible scores. Each factor score has a mean of 50 and standard deviation (SD) of 10 based on the nationally representative sample of older adults in ADAMS HRS, (e.g., a 1 point difference corresponds to a 0.1 SD unit difference). Higher scores indicate less impairment. The factors for overall cognitive performance, memory, and executive function were approximately normally distributed in each study (Figure 1). Reliability (internal consistency) of the factors, based on test information functions derived from the standard error of measurement, for overall cognitive performance and memory were above 0.90 between scores of 40 and 60 in each study, which captured most of the range in each sample (Figure 2). The executive function factor was less reliable, reflecting that different tests of executive function measure related but more heterogeneous cognitive processes (e.g., planning, set-shifting) compared, for example, to tests of memory.

Alignment analyses of items in common across studies revealed two items that were not psychometrically comparable across study. Semantic and phonemic fluency demonstrated DIF in item difficulty (i.e., thresholds) in WHICAP compared to the other studies, such that they were less difficult than in other studies. We allowed item discrimination and difficulty parameters for these items to vary across dataset when we obtained factor scores, which ensures that DIF does not endanger the overall harmonization.

The mean annual change in overall or general cognitive performance was -0.030 SD units. To confirm that model-estimated change in the overall cognitive performance factor reflected change in observed measures, we estimated random effects models for each individual cognitive test in the sample. The mean annual rate of decline was -0.039 SD units across all the measures, some of which overlapped across study while some did not.

Differences in cognitive performance by race/ethnicity and study

We estimated change processes for overall cognitive performance, memory, and executive function by race/ethnicity and study. Model-estimated means and 95% confidence intervals (CI) are plotted in Figures 3 (mean level) and 4 (annual change) by race/ethnicity and study. Supplemental Table 1 provides numeric results corresponding to those in the Figures. Supplemental Figures 7-9 show trajectory plots by race and study. Model fits were excellent for overall cognitive performance (empirical $r^2 = 0.89$), memory (empirical $r^2 = 0.78$), and executive function (empirical $r^2 = 0.85$), and diagnostic analyses of residuals suggested adequate model fit (Supplemental Figure 5).

Levels of cognition by race and study

The estimated level of cognitive performance (i.e., cross-sectional performance) at age 75 for each race/ethnicity group is presented by study in Figure 3. In general, for levels of overall cognitive performance, memory, and executive function, differences by race within study tended to be larger than between-study differences. For all cognitive outcomes in each study, non-Hispanic white participants outperformed non-Hispanic blacks by about 0.5 SD units. For the two studies that included Hispanics, non-Hispanic whites outperformed Hispanics by 0.9 SD units in WHICAP and 0.5 SD units in SENAS. There were some exceptions; the difference between non-Hispanic white and non-Hispanic black participants in SENAS was 0.2 SD units for overall cognitive performance and negligible for memory.

Comparing race groups across studies, mean overall cognitive performance and executive function for non-Hispanic white participants differed by less than 0.3 SD units. Memory was more variable between studies among non-Hispanic whites (up to 0.9 SD units different between WHICAP and SENAS).

In contrast, mean overall cognitive performance, memory, and executive function differed only by 0.2 to 0.3 SD units among non-Hispanic blacks across the studies. Among Hispanics in WHICAP compared to those in SENAS, mean overall cognitive performance, memory, and executive function differed by between 0.2, 0.6, and 0.3 SD units, respectively. Hispanics in WHICAP performed higher than Hispanics in SENAS for memory.

Changes in cognition by race and study. The estimated annual rate of cognitive change over time by cognitive domain, study, and race/ethnicity is shown in Figure 4. In contrast to the pattern for level of cognitive performance, race group differences in rate of cognitive decline varied across studies, but even when significant, were small. Significant race effects were found in WHICAP and SENAS, but non-Hispanic whites declined more in both studies. NCODE showed no substantial decline on any cognitive measure, while the other three studies showed average annual decline for all groups of about 0.04 to 0.10 SD/year. The largest difference across race groups for the three cognitive outcomes was less than 0.02 SD/year for WHICAP and 0.05 SD/year for SENAS (Memory). These results show greater differences across studies than do findings related to baseline scores, confirming our hypothesis that longitudinal change would be more affected than cross-sectional scores by study specific characteristics. However, ever where ethnic group differences were found, the

pattern did not match that for baseline scores where non-Hispanic whites had the most positive outcomes.

Differences in cognitive performance by education and study. Educational attainment was associated with the level of cognitive performance in predictable ways. The models fit well to the data, as indicated by empirical r^2 statistics for overall cognitive performance (empirical $r^2 = 0.89$), memory (empirical $r^2 = 0.78$), and executive function (empirical $r^2 = 0.85$). Diagnostic analyses of residuals suggested adequate model fit (Supplemental Figure 6). Supplemental Figures 10-12 show trajectory plots by years of education and study.

Levels of cognition by education and study. In general, for levels of overall cognitive performance, memory, and executive function, differences by education within study tended to be larger than between-study differences. Overall cognitive performance, memory, and executive function were higher in those with more education and these effects were similar across studies. Participants with a high school education performed between 0.3 SD units and 0.7 SD units worse on memory and overall cognitive performance than participants with a college education and between 0.5 SD units and 1.0 SD units worse than participants with a graduate education. Disparities by years of education in executive function performance were greater (between 0.3 and 0.8 SD units between high school and college, and between 0.6 and 1.2 SD units between high school and graduate education).

With respect to between-study differences by years of education, among participants with a high school education, mean overall cognitive performance, memory, and executive function differed by as much as 0.4, 0.6, and 0.5 SD units across study, respectively. Among participants with a college or a graduate education, mean overall cognitive performance, memory, and executive function differed by as much as 0.3, 1.0, and 0.3 SD units across study, respectively.

Changes in cognition by education and study. As with race/study trends, trends in slopes by education and study were opposite from trends in intercept by education and study: between-study differences in education effects were larger than education differences within each study. Among all participants, mean annual rate of decline in overall cognitive performance, memory, and executive function differed by as much as 0.06, 0.09, and 0.03 SD/year across studies.

With respect to education differences in cognitive change within each study, overall cognitive performance, memory, and executive function differed by just 0.01 SD/year between education groupings in WHICAP, 0.04 SD/year between education groupings in SENAS, 0.02 SD/year between education groupings in Duke MHA, and 0.01 SD/year between education groupings in NCODE.

Discussion

We used integrative longitudinal data analysis to examine differences in cognitive trajectories by race and education among older adults across four longitudinal cohort studies. Non-Hispanic whites consistently had higher baseline scores than blacks in all four studies, and even larger differences were observed in comparison with Hispanics in the two studies

that included Hispanics. These differences were robust and from 0.5 SD to more than 1.0 SD. In contrast, racial/ethnic differences in rate of change in cognition were small, and when present, favored blacks and Hispanics. That is, these minority groups declined less on average than non-Hispanic whites. The largest average differences in annual decline across ethnic groups were about 0.05 SD per year, which translates to a 10-year difference in overall decline of 0.5 SD. For education, results similarly showed large education related differences in baseline scores and much smaller differences in rate of change. There was an overall trend for higher education to be associated with slightly more rapid cognitive decline.

Taken together, results of this study show robust effects of ethnicity and education on baseline test performance with better performance for non-Hispanic whites and those with more education. In contrast, differences in rate of change were limited in magnitude, and minorities and those with lower education declined at slightly slower rates. These results do not support the hypothesis that racial/ethnic and education-level differences in cross-sectional cognitive test scores are mirrored in longitudinal rates of decline.

Several reasons have been posited to explain cross-sectional differences by race/ethnicity and education in cognitive test scores (Albert 1995; Jones, 2003; Zsembik & Peek, 2001). Differential functioning of test items, or biases, exist in certain tests (Jones, 2003; Pedraza & Mungas, 2008). Black and Hispanic older adults on average have lower levels of education, different occupational attainment, and live in lower socioeconomic status (SES) neighborhoods (Braveman et al., 2005; Williams, 1999; Williams & Collins, 1995). Level of educational attainment is strongly associated with cross-sectional cognitive test scores, explaining up to 50% of variance in some studies (Albert et al., 1995; Alley et al., 2007; Atkinson et al., 2005; Early et al., 2013; Manly et al., 2004; Mungas et al., 2000, 2005; Schiave, 1996). Additionally, quality of education has been identified as an important variable that explains substantial variance in cross-sectional test results after accounting for years of education (Crowe et al., 2013; Johnson et al., 2006; Manly et al., 2002; Sachs-Ericsson et al., 2005; Sisco et al., 2014). Some older adults, particularly in southern U.S. states, may have attended race-segregated schools, lacked access to higher quality of schooling, and had less encouragement to stay in school (Glymour et al., 2008; Glymour & Manly, 2008). Such early life experiences have been shown to have strong effects on cognitive test scores in later life and appear to explain much of observed race/ethnicity differences (Jones, 2003; Sisco et al., 2014).

Understanding why some people show decline cognitively in late life at a rapid rate while others maintain cognitive health presents a major challenge that has important potential prevention and public policy implications. Results of this study suggest that race/ethnicity and education do not have strong effects on cognitive decline, which would imply that variables that are strongly linked to race/ethnicity and education also are not major determinants of cognitive decline. It seems likely that diseases of aging such as Alzheimer's disease, cerebrovascular disease, and other progressive dementing illnesses are strong determinants of cognitive decline. Studies to examine similarities and differences of effects of these diseases in different race/ethnic and education groups will be important for clarifying the seeming paradox that minority status and lower education, which are

associated with broad health disparities that could result in brain injury and cognitive decline, were not associated with faster cognitive decline in four different studies of different populations that used different methods. The recent availability of brain imaging and other in-vivo biomarkers makes it possible to directly measure biological effects of these diseases and to evaluate their differential effects in diverse older persons.

Harmonized analyses of merged samples provide a unique opportunity to identify broad and generalizable contributions of race/ethnicity and education to late life cognition apart from study-specific biases. The present study makes a unique contribution to research on late life cognition in diverse populations by integrating results from multiple diverse studies to address questions about how ethnicity and education influence late-life cognitive trajectories. The methods used to harmonize measures across studies represent a distinct methodological advance for research in this area, and specifically, make it possible to merge results from different studies to address important questions. Consistent findings across studies of different samples are vital to science because they tend to highlight results that are broadly generalizable. Results of this study were consistent in showing robust effects of race/ethnicity and education on baseline scores but more variable effects on rate of change. The variability across studies likely represents systematic differences in samples and methods.

Our study has several strengths. First, the integration of studies with diverse measures, research designs, and demographic characteristics gives our integrative data analysis a novel perspective from which to generalize findings across sample, age, and demographic characteristics (Curran & Hussong, 2009; Shadish et al., 1991). As we move into a new era where datasets from research projects are publically available and funding for new studies is limited, this ability to harmonize results across studies will greatly contribute to the ability to address important research hypothesis and efficiently utilize the studies that have contributed to this overall body of data. Second, latent factors were constructed to take full advantage of all available cognitive data and maximize precision. Our analytic approach is robust to the inclusion of additional studies; because we externally scaled the factors to a national reference group, item parameters are not overly influenced by studies with larger sample sizes.

An important limitation of the present study is that we did not adjust for all possible confounders of the relationship between race and cognitive level and decline, such as vascular risk factors and educational quality (Manly et al., 2003; Sisco et al., 2014), because not every study measured relevant variables in the same way. It is conceivable that some variables are stronger confounders in certain studies than others, but we are unable to assess this hypothesis. A second limitation is that our calibration procedure was designed to derive scores across studies that are comparable to a nationally representative sample. This approach does not make any given study nationally representative by itself and does not standardize rates of change across studies. Third, although IRT provides a roadmap for linking a harmonized measure within and across studies, harmonized measures within a specific study are limited by the psychometric characteristics of the donor items; if a study chose poor donor tests, harmonization will not help. The factor analytic framework we used addresses multidimensionality, while more traditional approaches such as standardizing and

averaging tests do not. In a similar vein, we mapped certain cognitive indicators to factors based on clinical judgment. Although sufficiently unidimensional, the overall cognitive measure and the executive functioning measures are probably not completely unidimensional. This is a psychometric concern that we partially addressed by including a specific factor for items from the Logical Memory test. A fourth limitation is that race and education (e.g., group) differences in cognitive performance in this integrative data analysis could be biased due to imbalances in measurement precision of the factors across study. Such a bias would usually make it harder to detect group differences within a given study, compared to between-study differences, which might be driving our findings with respect to cognitive change. However, this potential limitation is based on the presumption that cognitive tests, which overlapped appreciably across dataset, are differentially sensitive to cognitive change in one study more so than another. A fifth study limitation is that we examined differential item functioning with respect to study membership, but did not evaluate item-level differences with respect to other variables such as age, sex, years of education, or race. Although prior research has suggested differential item functioning based on these variables for some cognitive tests used here, a key concern in the integrative data analysis was whether tests were comparable across dataset but comprehensive evaluation of item bias was not a primary focus of the present study. A final limitation is that the samples used are different with respect to recruitment factors, lapses between study visits, and age ranges. These differences may have contributed to findings we attributed to between-study variability. For example, higher cross-sectional performance in memory among Hispanics in WHICAP compared to Hispanics in SENAS might be attributable to regional differences in the Hispanic population in the western U.S. (SENAS) vs. the Northeastern U.S. (WHICAP), or a function of different recruitment methods (Table 2). This limitation is also a strength in terms of external validity; our pooled sample was large and represents a level of diversity in background variables, including different studies with different study procedures, that is more extensive than individual studies in the integrative data analysis.

We used cognitive tests to measure longitudinal cognitive performance, which presumably reflects disease-related changes but which is also influenced by other factors including intrinsic motivation and environmental characteristics. Importantly, longitudinal designs allow us to separate disease from life experience. Still, cognition is a complex phenotype determined by many characteristics. This complexity is partly reflected in how we named the overall or general cognitive *performance* factor. The broad range of tests making up the summary factor measure a person's performance, not necessarily their *functioning* in the environment. There could be bias in how we measured functioning, or participants may have some cognitive or brain reserve that makes their "true" level of functioning different from their performance (e.g., Stern, 2012). Our goal was to summarize cognitive performance across a broad range of tests and remove group-level biases. We claim the overall cognitive performance factor sufficiently summarizes an individual's performance on the tests, not that it reflects an individual's level of functioning as a single hierarchical ability.

Why conduct harmonization in the first place? The substantive message from this study is that differences due to race/ethnicity and education in baseline scores are not mirrored in differences in longitudinal change. Combining scores on harmonized metrics allowed us to

examine similarities and differences across studies. We could have done the same thing with coordinated, or parallel but separate, analyses within specific studies. Results would not be on the same metric, but the general pattern of results of scientific interest most likely would have been preserved, albeit with caveats about the differences in metrics addressed in our approach by the simulation study (Supplemental Materials 1). Harmonizing would have greater impact when individual studies are not large enough to address the questions of interest, although such a scenario also would present greater risk because unknown biases might be present, prohibiting us from verifying that results obtained by combining studies match those obtained within studies. An ideal harmonization approach, which we conducted, entails carefully choosing donor items to build in high-quality measurement at the item level, selecting samples appropriate to the research question being addressed, and identifying potential biases within samples that either should be removed prior to harmonization or addressed in analytic models. The alternatives to harmonizing include: a) not combining results from different studies, which can have scientific costs, b) parallel analysis that in many cases cannot be done due to sample size issues, and c) standardizing and averaging together individual tests, which has the same disadvantages of our approach to harmonization and none of the advantages (Gross et al., 2014b).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was possible through extensive existing networks of investigators from multiple studies across multiple institutions that were willing to make their data available, for which we are grateful.

Financial Disclosure: This work was supported by National Institutes of Health grants R03 AG045494 (PI: Gross) and R13 AG030995 (PI: Mungas). Dr. MacKay-Brandt was supported by a National Institutes of Mental Health post-doctoral fellowship (T32MH20004). Dr. Gibbons was supported by P50 AG05136 (PI: Raskind). Dr. Potter was supported by National Institutes of Health Grant K23 MH087741. Data provided by the Joseph and Kathleen Alzheimer's Disease Research Center at Duke was supported by a grant from the National Institute on Aging, P30 AG028377. We thank Dr. Kathleen Welsh-Bohmer for her consent to use the data from Duke data. Data provided by the NCODE study was supported by National Institutes of Health Grants R01 MH054846 and P50 MH060451. We thank Dr. David Steffens for his consent to use NCODE data. The contents do not necessarily represent views of the funding entities. Funders had no deciding roles in the design and conduct of the study.

References

- Albert MS, Jones K, Savage CR, Berkman L, Seeman T, Blazer D, Rowe JW. Predictors of cognitive change in older persons: MacArthur studies of successful aging. *Psychol Aging*. 1995; 10(4):578–89. [PubMed: 8749585]
- Alley D, Suthers K, Crimmins E. Education and Cognitive Decline in Older Americans: Results From the AHEAD Sample. *Res Aging*. 2007; 29(1):73–94. [PubMed: 19830260]
- Asparouhov, T.; Muthén, B. Plausible values for latent variables using Mplus. Technical Report. 2010. Accessed July 10, 2014 from <http://www.statmodel.com/download/Plausible.pdf>
- Asparouhov T, Muthén B. Multiple-group factor analysis alignment. *Struct. Equ. Modeling*. 2014; 21:1–14.
- Atkinson HH, Cesari M, Kritchevsky SB, Penninx BW, Fried LP, Guralnik JM, Williamson JD. Predictors of combined cognitive and physical decline. *J Am Geriatr Soc*. 2005; 53(7):1197–202. [PubMed: 16108938]

- Barnes LL, Wilson RS, Li Y, Aggarwal NT, Gilley DW, McCann JJ, Evans DA. Racial differences in the progression of cognitive decline in Alzheimer disease. *Am J Geriatr Psychiatry*. 2005; 13(11): 959–67. [PubMed: 16286439]
- Bauer DJ, Hussong AM. Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychol Methods*. 2009; 14(2):101–25. [PubMed: 19485624]
- Benton, A.; Hamsher, K. *Multilingual Aphasia Examination*. University of Iowa; Iowa City, IA: 1976.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 327(8476):307–10. [PubMed: 2868172]
- Bosma H, van Boxtel MP, Ponds RW, Houx PJ, Burdorf A, Jolles J. Mental work demands protect against cognitive impairment: MAAS prospective cohort study. *Experimental Aging Research*. 2003; 29:33–45. [PubMed: 12735080]
- Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, Posner S. Socioeconomic status in health research: one size does not fit all. *JAMA*. 2005; 294(22):2879–88. [PubMed: 16352796]
- Brown, TA. *Confirmatory factor analysis for applied research*. Guilford Press; New York, NY: 2006.
- Butler SM, Ashford JW, Snowdon DA. Age, education, and changes in the Mini-Mental State Exam scores of older women: Findings from the Nun Study. *J Am Geriatr Soc*. 1996; 44:675–681. [PubMed: 8642159]
- Carter JS, Pugh JA, Monterrosa A. Non-insulin-dependent diabetes mellitus in minorities in the United States. *Ann Intern Med*. 1996; 125(3):221–232. [PubMed: 8686981]
- Carvalho JO, Tommet D, Crane PK, Thomas M, Claxton A, Habeck C, Manly JL, Romero H. Deconstructing racial differences: The effects of quality of education and cerebrovascular risk factors. *Journal of Gerontology, Series B: Psychological Sciences*. 2014
- Crane PK, Narasimhalu K, Gibbons LE, Mungas DM, Haneuse S, Larson EB, et al. Item response theory facilitated calibrating cognitive tests and reduced bias in estimated rates of decline. *Journal of Clinical Epidemiology*. 2008; 61:1018–1027. [PubMed: 18455909]
- Crowe M, Clay OJ, Martin RC, Howard VJ, Wadley VG, Sawyer P, Allman RM. Indicators of childhood quality of education in relation to cognitive function in older adulthood. *J Gerontol A Biol Sci Med Sci*. 2013; 68(2):198–204. [PubMed: 22546959]
- Curran PJ, Hussong AM. *Integrative Data Analysis: The Simultaneous Analysis of Multiple Data Sets*. *Psychological Methods*. 2009; 14:81–100. [PubMed: 19485623]
- Curran PJ, McGinley JS, Bauer DJ, Hussong AM, Burns A, Chassin L, Sher K, Zucker R. A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research*. 2014; 49:214–231. [PubMed: 25960575]
- Duggan C, Carosso E, Mariscal N, Islas I, Ibarra G, Holte S, Thompson B. Diabetes prevention in Hispanics: report from a randomized controlled trial. *Prev Chronic Dis*. 2014; 11:E28. [PubMed: 24576395]
- Early DR, Widaman KF, Harvey D, Beckett L, Park LQ, Farias ST, Reed BR, Decarli C, Mungas D. Demographic predictors of cognitive change in ethnically diverse older persons. *Psychol Aging*. 2013; 28(3):633–45. [PubMed: 23437898]
- Embretson, SE.; Reise, SP. *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers; Mahwah, NJ: 2000.
- Espino DV, Lichtenstein MJ, Palmer RF, Hazuda HP. Ethnic differences in mini-mental state examination (MMSE) scores: Where you live makes a difference. *J Am Geriatr Soc*. 2001; 9(5): 538–548. [PubMed: 11380745]
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12(3):189–98. [PubMed: 1202204]
- Gibbons LE, McCurry S, Rhoads K, Masaki K, White L, Borenstein AR, Crane PK. Japanese-English language equivalence of the Cognitive Abilities Screening Instrument among Japanese-Americans. *Int Psychogeriatr*. 2009; 21(1):129–37. [PubMed: 18947456]
- Glymour MM, Manly JJ. Lifecourse Social Conditions and Racial and Ethnic Patterns of Cognitive Aging. *Neuropsychol Rev*. 2008; 18:223–254. [PubMed: 18815889]

- Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *American Journal of Epidemiology*. 2005; 162:267-278.
- Glymour MM, Kawachi I, Jencks C, Berkman L. Does childhood schooling affect old age memory and cognitive function? Using state schooling laws as natural experiments. *Journal of Epidemiology and Community Health*. 2008; 62:532-537. [PubMed: 18477752]
- Griffith, L.; van den Heuvel, E.; Fortier, I.; Hofer, S.; Raina, P.; Sohel, N.; Belleville, S. Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis [Internet]. AHRQ Methods for Effective Health Care; Rockville (MD): 2013. Agency for Healthcare Research and Quality (US); 2013 Mar. Report No.: 13-EHC040-EF
- Gross AL, Jones RN, Fong TG, Tommet D, Inouye SK. Calibration and validation of an innovative approach for estimating general cognitive performance. *Neuroepidemiology*. 2014a; 42:144-153. [PubMed: 24481241]
- Gross AL, Sherva R, Mukherjee S, Newhouse S, Kauwe JSK, Munsie LM, Crane PK, Alzheimer's Disease Neuroimaging Initiative, GENAROAD Consortium, and AD Genetics Consortium. Calibrating longitudinal cognition in Alzheimer's disease across diverse test batteries and datasets. *Neuroepidemiology*. 2014b; 43(3-4):194-205. [PubMed: 25402421]
- Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organ Res Methods*. 2004; 7:191.
- Herzog AR, Wallace RB. Measures of cognitive functioning in the AHEAD Study. *J Gerontol B Psychol Sci Soc Sci*. 1997; 52B:37-48. Special Issue. [PubMed: 9215356]
- Johansson B, Berg S, Hofer SM, Allaire JC, Maldonado-Molina MM, Piccinin AM, Pedersen NL, McClearn GE. Change in cognitive capabilities in the oldest old: The effects of proximity to death in genetically related individuals over a 6-year period. *Psychol Aging*. 2004; 19:145-156. [PubMed: 15065938]
- Johnson A, Flicker LJ, Lichtenberg PA. Reading ability mediates the relationship between education and executive function tasks. *J Int Neuropsychol Soc*. 2006; 12:64-71. [PubMed: 16433945]
- Jones RN, Fonda SJ. Use of an IRT-based latent variable model to link different forms of the CES-D from the Health and Retirement Study. *Soc Psychiatry Psychiatr Epidemiol*. 2004; 39(10):828-35. [PubMed: 15669664]
- Jones RN. Racial bias in the assessment of cognitive functioning of older adults. *Aging and Mental Health*. 2003; 7(2):83-102. [PubMed: 12745387]
- Jones RN, Gallo JJ. Education and sex differences in the mini-mental state examination: effects of differential item functioning. *J Gerontol B Psychol Sci Soc Sci*. 2002; 57(6):548-58.
- Jones RN, Rudolph JL, Inouye SK, Yang FM, Fong TG, Milberg WP, et al. Development of a unidimensional composite measure of neuropsychological functioning in older cardiac surgery patients with good measurement precision. *Journal of Clinical and Experimental Neuropsychology*. 2010; 32:1041-1049. [PubMed: 20446144]
- Jöreskog K, Goldberger AS. Estimation of a Model with a Multiple Indicators and Multiple Causes of a Single Latent Variable. *JASA*. 1975; 70:631-639.
- Jöreskog KG, Moustaki I. Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*. 2001; 36(3):347-387.
- Juster FT, Suzman R. An overview of the Health and Retirement Study. *Journal of Human Resources*. 1995; 30(Suppl.):7-56.
- Karlamangla AS, Miller-Martinez D, Aneshensel CS, Seeman TE, Wight RG, Chodosh J. Trajectories of cognitive function in late life in the United States: Demographic and socioeconomic predictors. *American Journal of Epidemiology*. 2009; 170:331-342.
- Lachman, ME.; Firth, KM. The adaptive value of feeling in control during midlife. In: Brim, OG.; Ryff, CD.; Kessler, R., editors. *How healthy are we? A national study of wellbeing at midlife*. University of Chicago Press; Chicago: 2004. p. 320-349.
- Langa KM, Plassman BL, Wallace RB, Herzog AR, Heeringa SG, Willis RJ. The Aging, Demographics, and Memory Study: study design and methods. *Neuroepidemiology*. 2005; 25:181-91. [PubMed: 16103729]

- LaVeist T, Pollack K, Thorpe R Jr, Fesahazion R, Gaskin D. Place, not race: disparities dissipate in southwest Baltimore when blacks and whites live under similar conditions. *Health Aff (Millwood)*. 2011; 30(10):1880–7. [PubMed: 21976330]
- Leveille SG, Guralnik JM, Ferrucci L, Corti MC, Kasper J, Fried LP. Black/White differences in the relationship between MMSE scores and disability: The Women's Health and Aging Study. *J Gerontol B Psychol Sci Soc Sci*. 1998; 53B:P201–P208. [PubMed: 9602835]
- Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Stern H. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012; 367(14):1355–60. [PubMed: 23034025]
- Lord FM. The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*. 1953; 13(4):517–549.
- Lord, FM. Applications of item response theory to practical testing problems. Erlbaum; Mahwah, NJ: 1980.
- Lyketsos CG, Chen LS, Anthony JC. Cognitive decline in adulthood: An 11.5-year follow-up of the Baltimore Epidemiologic Catchment Area Study. *The American Journal of Psychiatry*. 1999; 156:58–65. [PubMed: 9892298]
- 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 non-demented elderly African Americans and whites. *Neurology*. 1998; 50:1238–1245. [PubMed: 9595969]
- Manly JJ, Touradji P, Tang MX, Stern Y. Literacy and memory decline among ethnically diverse elders. *J Clin Exp Neuropsychol*. 2003; 25(5):680–90. [PubMed: 12815505]
- Masel MC, Peek MK. Ethnic differences in cognitive function over time. *Annals of Epidemiology*. 2009; 19:778–783. [PubMed: 19656690]
- McArdle, JJ.; Bell, RQ. Recent trends in modeling longitudinal data by latent growth curve methods. In: Little, TD.; Schnabel, KU.; Baumert, J., editors. *Modeling longitudinal and multiplegroup data: Practical issues, applied approaches, and scientific examples*. Lawrence Erlbaum; Mahwah, NJ: 2000. p. 69-108.
- McArdle JJ, Grimm KJ, Hamagami F, Bowles RP, Meredith W. Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*. 2009; 14:126–149. [PubMed: 19485625]
- McDonald, R. Test theory: A unified treatment. Erlbaum Associates; Mahwah, NJ: 1999.
- McHorney CA. Ten recommendations for advancing patient-centered outcomes measurement for older persons. *Annals of Internal Medicine*. 2003; 139:403–409. [PubMed: 12965966]
- Mislevy RJ. Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*. 1986; 11(1):3–31.
- Morales LS, Flowers C, Gutierrez P, Kleinman M, Teresi JA. Item and scale differential functioning of the Mini-Mental State Exam assessed using the Differential Item and Test Functioning (DFIT) Framework. *Med Care*. 2006; 44(11 Suppl 3):S143–51. [PubMed: 17060821]
- Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C, CERAD investigators. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. 1989; 39:1159–1165. [PubMed: 2771064]
- Mungas D, Reed BR, Haan MN, González H. Spanish and English Neuropsychological Assessment Scales: Relationship to demographics, language, cognition, and independent function. *Neuropsychology*. 2005; 19:466–475.
- Mungas D, Reed BR, Marshall SC, González HM. Development of psychometrically matched English and Spanish neuropsychological tests for older persons. *Neuropsychology*. 2000; 14:209–223.
- Mungas D, Reed BR, Crane PK, Haan MN, González H. Spanish and English Neuropsychological Assessment Scales (SENAS): Further development and psychometric characteristics. *Psychol Assess*. 2004; 16(4):347–59. [PubMed: 15584794]
- Muthen, B. Latent variable modeling with longitudinal and multilevel data. In: Raftery, A., editor. *Sociological Methodology*. Blackwell Publishers; Boston: 1997. p. 453-480.

- Muthen BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*. 1997; 2:371–402.
- Muthen, LK.; Muthen, BO. *Mplus user's guide: Seventh Edition*. Muthen & Muthen; Los Angeles, CA: 1998–2008.
- Muthén B, Asparouhov T. IRT studies of many groups: the alignment method. *Front Psychol*. 2014; 5:978. [PubMed: 25309470]
- Obisesan TO, Vargas CM, Gillum RF. Geographic variation in stroke risk in the United States. Region, urbanization, and hypertension in the Third National Health and Nutrition Examination Survey. *Stroke*. 2000; 31(1):19–25. [PubMed: 10625710]
- Orlando Edelen MO, Thissen D, Teresi JA, Kleinman M, Ocepek-Welikson K. Identification of differential item functioning using item response theory and the likelihood-based model comparison approach. Application to the Mini-Mental State Examination. *Med Care*. 2006; 44(11 Suppl 3):S134–42. [PubMed: 17060820]
- Pedraza O, Mungas D. Measurement in cross-cultural neuropsychology. *Neuropsychol Rev*. 2008; 18(3):184–93. [PubMed: 18814034]
- Rabbitt P, Diggle P, Holland F, McInnes L. Practice and drop-out effects during a 17-year longitudinal study of cognitive aging. *J Gerontol B Psychol Sci Soc Sci*. 2004; (2):84–97.
- Reardon SF, Galindo C. Patterns of Hispanic Students' Math Skill Proficiency in the Early Elementary Grades. *Journal of Latinos and Education*. 2007; 6(3):229–252.
- Reitan R. Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*. 1958; 8:271–276.
- Rodgers WL, Ofstedal MB, Herzog AR. Trends in scores on tests of cognitive ability in the elderly U.S. population, 1993–2000. *J Gerontol B Psychol Sci Soc Sci*. 2003; 58(6):S338–46. [PubMed: 14614121]
- Sachs-Ericsson N, Blazer DG. Racial differences in cognitive decline in a sample of community-dwelling older adults: the mediating role of education and literacy. *Am J Geriatr Psychiatry*. 2005; 13(11):968–75. [PubMed: 16286440]
- Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*. 1969 No. 17.
- Scarmeas N, Albert SM, Manly JJ, Stern Y. Education and rates of cognitive decline in incident Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2006; 77(3):308–16. [PubMed: 16484637]
- Schaie, KW. *Intellectual Development in Adulthood: The Seattle Longitudinal Study*. Cambridge University Press; New York: 1996.
- Schaie, KW. *Developmental Influences on Adult Intellectual Development: The Seattle Longitudinal Study*. Oxford University Press; New York: 2005.
- Schaie, KW. Historical processes and patterns of cognitive aging. In: Hofer, SM.; Alwin, DF., editors. *Handbook on Cognitive Aging: Interdisciplinary Perspective*. Sage; Thousand Oaks, CA: 2008. p. 368–383.
- Schaie KW, Willis SL, Pennak S. A historical framework for cohort differences in intelligence. *Research in Human Development*. 2005; 2:43–67. [PubMed: 16858496]
- Schneider AL, Sharrett AR, Patel MD, Alonso A, Coresh J, Mosley T, Selnes O, Selvin E, Gottesman RF. Education and cognitive change over 15 years: the atherosclerosis risk in communities study. *J Am Geriatr Soc*. 2012; 60(10):1847–53. [PubMed: 23013064]
- Schwartz BS, Glass TA, Bolla KI, Stewart WF, Glass G, Rasmussen M, Bressler J, Shi W, Bandeen-Roche K. Disparities in cognitive functioning by race/ethnicity in the Baltimore Memory Study. *Environ Health Perspect*. 2004; 112(3):314–20. [PubMed: 14998746]
- Shadish, WR.; Cook, TD.; Leviton, LC. *Foundations of program evaluation: Theories of practice*. Sage; Newbury Park, CA: 1991.
- Shih RA, Ghosh-Dastidar B, Margolis KL, Slaughter ME, Jewell A, Bird CE, Eibner C, Denburg NL, Ockene J, Messina CR, Espeland MA. Neighborhood socioeconomic status and cognitive function in women. *Am J Public Health*. 2011; 101(9):1721–8. [PubMed: 21778482]
- Singer, JD.; Willet, J. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press; New York, NY: 2003.

- Sisco S, Gross AL, Shih RA, Sachs BC, Glymour MM, Bangen KJ, Benitez A, Skinner J, Schneider BC, Manly JJ. 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*. 2014 Epub.
- Sloan FA, Wang J. Disparities among older adults in measures of cognitive function by race or ethnicity. *J Gerontol B Psychol Sci Soc Sci*. 2005; 60B:P242–P250. [PubMed: 16131618]
- Smith, A. Symbol Digits Modalities Test. Western Psychological Services; Los Angeles: 1973.
- Steffens DC, Welsh-Bohmer KA, Burke JR, Plassman BL, Beyer JL, Gersing KR, Potter GG. Methodology and preliminary results from the neurocognitive outcomes of depression in the elderly study. *J Geriatr Psychiatry Neurol*. 2004; 17(4):202–211. [PubMed: 15533991]
- Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol*. 2012; 11(11):1006–12. [PubMed: 23079557]
- Stull DE. Analyzing growth and change: Latent variable growth curve modeling with an application to clinical trials. *Quality of Life Research*. 2008; 17:47–59. [PubMed: 18080215]
- Takane Y, de Leeuw J. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*. 1987; 52:393–408.
- Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, Mayeux R. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. *Neurology*. 2001; 56(1): 49–56. [PubMed: 11148235]
- Tucker-Drob EM, Johnson KE, Jones RN. The cognitive reserve hypothesis: A longitudinal examination of age-associated declines in reasoning and processing speed. *Developmental Psychology*. 2009; 45:431–446.
- United States Census Bureau. U.S. Census Bureau Projections Show a Slower Growing, Older, More Diverse Nation a Half Century from Now. 2012. Accessed July 10, 2014 from <http://www.census.gov/newsroom/releases/archives/population/cb12-243.html>
- van Exel E, Gussekloo J, de Craen AJM, Bootsma-van der Wiel A, Houx P, Knook DL, Westendorp RGJ. Cognitive function in the oldest old: Women perform better than men. *Journal of Neurology, Neurosurgery, and Psychiatry*. 2001; 71:29–32.
- Wang W-C. Effects of anchor item methods on the detection of differential item functioning within the family of Rasch models. *Journal of Experimental Education*. 2004; 72(3):221–261.
- Wechsler, D. Wechsler Memory Scale-Revised. Psychological Corporation; San Antonio, Texas: 1987.
- Williams DR, Collins C. US Socioeconomic and Racial Differences in Health: Patterns and Explanations. *Annual Review of Sociology*. 1995; 21:349–386.
- Williams DR. Race, socioeconomic status, and health. The added effects of racism and discrimination. *Ann N Y Acad Sci*. 1999; 896:173–88. [PubMed: 10681897]
- Williams BW, Mack W, Henderson VW. Boston naming test in Alzheimer's disease. *Neuropsychologia*. 1989; 27(8):1073–1079. [PubMed: 2797414]
- Willis, SL. Cohort differences in cognitive aging: a sample case. In: Schaie, KW.; Schooler, C., editors. *Social Structure and Aging: The Psychological Processes*. Erlbaum; New York: 1989. p. 95-112.
- Wilson RS, Aggarwal NT, Barnes LL, Mendes de Leon CF, Hebert LE, Evans DA. Cognitive decline in incident Alzheimer disease in a community population. *Neurology*. 2010; 74:951:955.
- Wilson RS, Hebert LE, Scherr PA, Barnes LL, Mendes d.L. Evans DA. Educational attainment and cognitive decline in old age. *Neurology*. 2009; 72:460:465.
- Wolinsky FD, Bentler SE, Hockenberry J, Jones MP, Weigel PA, Kaskie B, Wallace RB. A prospective cohort study of long-term cognitive changes in older Medicare beneficiaries. *BMC Public Health*. 2011; 11:710. [PubMed: 21933430]
- Zahodne LB, Glymour MM, Sparks C, Bontempo D, Dixon RA, MacDonald SWS, Manly JJ. Education does not slow cognitive decline with aging: 12-year evidence from the Victoria longitudinal study. *JINS*. 2011; 17:1039:1046.
- Zhao B, Jose PO, Pu J, Chung S, Ancheta IB, Fortmann SP, Palaniappan LP. Racial/Ethnic Differences in Hypertension Prevalence, Treatment, and Control for Outpatients in Northern California 2010-2012. *Am J Hypertens*. 2014 [Epub ahead of print].

Zsembik BA, Peek MK. Race differences in cognitive functioning among older adults. *J Gerontol B Psychol Sci Soc Sci.* 2001; 56B:S266–S274. [PubMed: 11522808]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

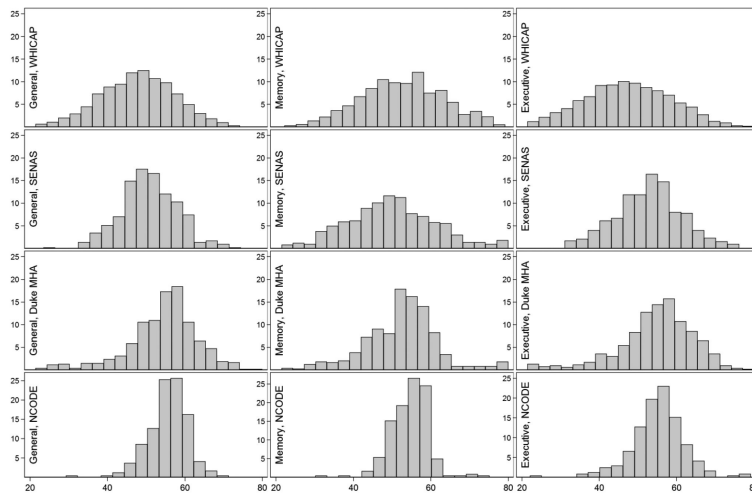


Figure 1. Histograms of general and domain-specific cognitive performance by study (N=5,803)
 Legend. Distributions of general and domain-specific cognitive performance by study. The Y-axes are percentages in each sample and the X-axes are level of cognitive performance. WHICAP = Washington Heights-Inwood Columbia Aging Project; SENAS = Spanish and English Neuropsychological Assessment Scales; Duke MHA = Duke Mental Health and Aging Study; NCODE = Neurocognitive Outcomes of Depression in the Elderly.

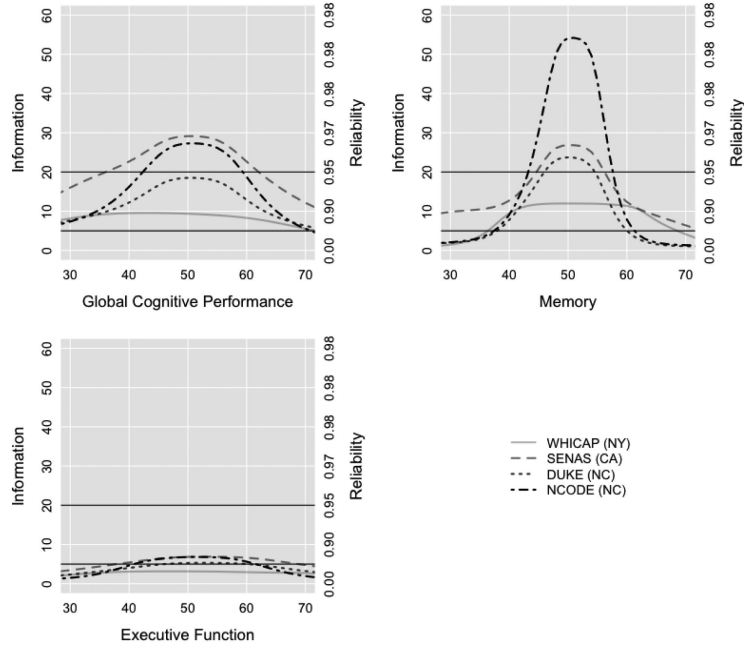


Figure 2. Test information plots for general cognitive performance, memory, and executive function by study (N=5,803)

Legend. Dataset-specific information of the general cognitive performance (top left), memory (top right), and executive function (bottom left) factors are plotted over the range of cognitive ability. Test information is calculated as the inverse of the square of the standard error of measurement. A dotted vertical line denotes a score of 50, which represents the mean cognitive performance in the Aging, Demographics, and Memory Substudy Health and Retirement Study (ADAMS HRS) normative sample. The bell-shaped distributions are consistent with scores optimized for the study of between-persons differences and longitudinal change across a wide range of cognitive performance. The horizontal line at a reliability of 0.80 indicates acceptable reliability for between-persons differences and the one denoting a reliability of 0.95 represents acceptable reliability for empirical analyses of within-person change. $Reliability = 1 - 1 / Information$, where information was calculated as the square root of the inverse of the standard error of measurement.

WHICAP = Washington Heights-Inwood Columbia Aging Project; SENAS = Spanish and English Neuropsychological Assessment Scales; Duke MHA = Duke Mental Health and Aging Study; NCODE = Neurocognitive Outcomes of Depression in the Elderly.

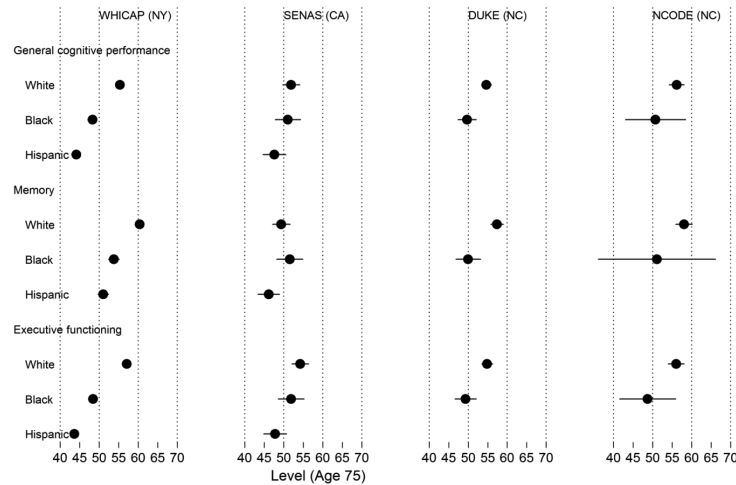


Figure 3. Model-estimated mean level of general cognitive performance, memory, and executive function by study and race at age 75 years (N=5,803)

Legend. Results from a parallel process latent growth curve model of general cognitive performance, memory, and executive function in the pooled sample, with indicators for race, study, and control variables (sex, clinical diagnostic status, and depressive symptoms in Neurocognitive Outcomes of Depression in the Elderly [NCODE]). Age from 75 years was the timescale of interest. This plot shows estimated mean levels at age 75 for each race and study group. The x-axis is on a T-score metric (mean 50, standard deviation 10) scaled to a national metric using the Aging, Demographics, and Memory Substudy Health and Retirement Study (ADAMS HRS) study. WHICAP = Washington Heights-Inwood Columbia Aging Project; SENAS = Spanish and English Neuropsychological Assessment Scales; Duke MHA = Duke Mental Health and Aging Study; NCODE = Neurocognitive Outcomes of Depression in the Elderly.

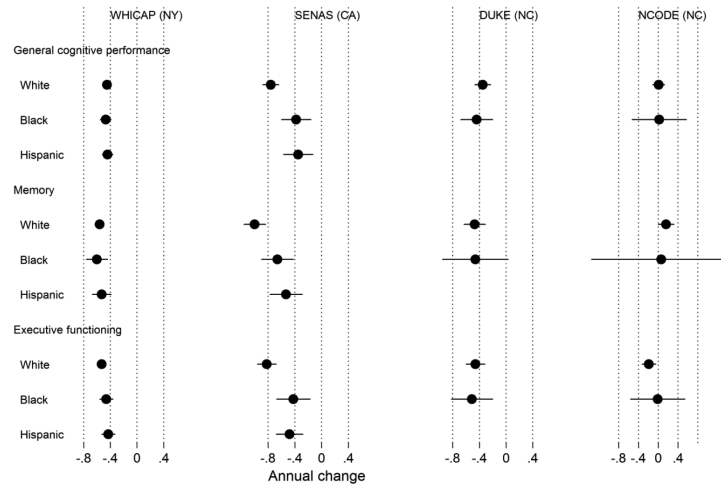


Figure 4. Model-estimated annual change in general cognitive performance, memory, and executive function by study and race (N=5,803)

Legend. Results from a parallel process latent growth curve model of general cognitive performance, memory, and executive function in the pooled sample, with indicators for race, study, and control variables (sex, clinical diagnostic status, depressive symptoms in Neurocognitive Outcomes of Depression in the Elderly [NCODE]). Age from 75 years was the timescale of interest. This plot shows estimated annual rates of change for each race and study group. The x-axis is on a T-score metric (mean 50, standard deviation 10) scaled to a national metric using the Aging, Demographics, and Memory Substudy Health and Retirement Study (ADAMS HRS) study. The model-estimated variance for annual changes in general cognitive performance, memory, and executive function was 0.029, 0.066, and 0.011, respectively. WHICAP = Washington Heights-Inwood Columbia Aging Project; SENAS = Spanish and English Neuropsychological Assessment Scales; Duke MHA = Duke Mental Health and Aging Study; NCODE = Neurocognitive Outcomes of Depression in the Elderly.

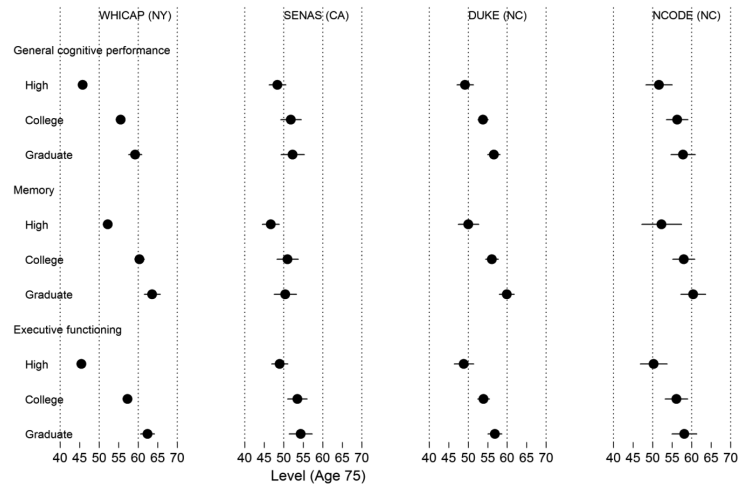


Figure 5. Model-estimated mean level of general cognitive performance, memory, and executive function by study and education level at age 75 years (N=5,803)

Legend. This plot shows estimated mean levels at age 75 for each education and study group. Results from a parallel process latent growth curve model of general cognitive performance, memory, and executive function in the pooled sample, with indicators for educational attainment, study, and control variables (sex, clinical diagnostic status, depressive symptoms in Neurocognitive Outcomes of Depression in the Elderly [NCODE]). Age from 75 years was the timescale of interest. The x-axis is on a T-score metric (mean 50, standard deviation 10) scaled to a national metric using the Aging, Demographics, and Memory Substudy Health and Retirement Study (ADAMS HRS) study. WHICAP = Washington Heights-Inwood Columbia Aging Project; SENAS = Spanish and English Neuropsychological Assessment Scales; Duke MHA = Duke Mental Health and Aging Study; NCODE = Neurocognitive Outcomes of Depression in the Elderly.

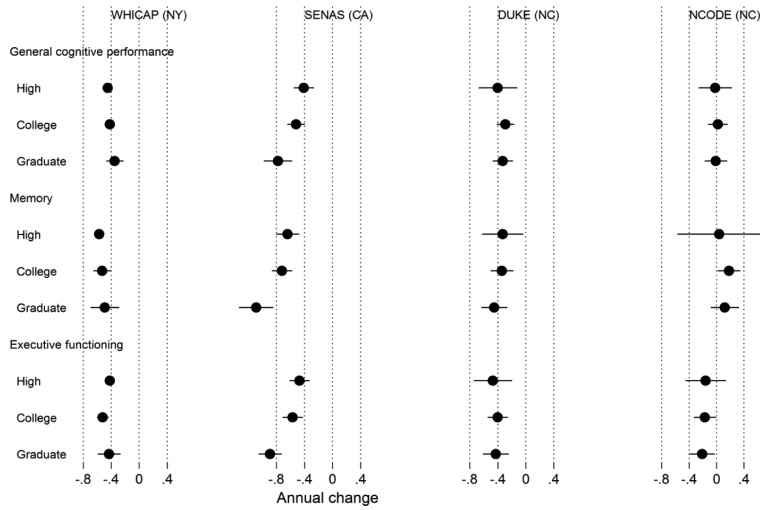


Figure 6. Model-estimated annual change in general cognitive performance, memory, and executive function by study and education level (N=5,803)

Legend. This plot shows estimated annual rates of change for each education and study group. Results from a parallel process latent growth curve model of general cognitive performance, memory, and executive function in the pooled sample, with indicators for level of education, study, and control variables (sex, clinical diagnostic status, depressive symptoms in Neurocognitive Outcomes of Depression in the Elderly [NCODE]). Age from 75 years was the timescale of interest. The x-axis is on a T-score metric (mean 50, standard deviation 10) scaled to a national metric using the Aging, Demographics, and Memory Substudy Health and Retirement Study (ADAMS HRS) study. The model-estimated variance for annual changes in general cognitive performance, memory, and executive function was 0.025, 0.064, and 0.009, respectively.

WHICAP = Washington Heights-Inwood Columbia Aging Project; SENAS = Spanish and English Neuropsychological Assessment Scales; Duke MHA = Duke Mental Health and Aging Study; NCODE = Neurocognitive Outcomes of Depression in the Elderly.

Table 1

Baseline demographic characteristics of the overall and study-specific samples (N=5,803)

	Pooled sample	Washington Heights-Inwood Columbia Aging Project (WHICAP)	Spanish and English Neuropsychological Assessment Scales (SENAS)	Duke Mental Health and Aging Study	Neurocognitive Outcomes of Depression in the Elderly (NCODE)
Sample size	5803	4115	525	578	585
Age, years, mean (SD)	76 (8)	77 (7)	74 (7)	73 (9)	72 (7)
Sex, Female, N, %	3889 (67)	2820 (69)	317 (60)	356 (62)	396 (68)
Race/ethnicity (N, %)					
Non-Hispanic White	2225 (38)	1014 (25)	235 (45)	460 (80)	516 (88)
Non-Hispanic Black	1715 (30)	1390 (34)	138 (26)	118 (20)	69 (12)
Hispanic	1863 (32)	1711 (42)	152 (29)	0 (0)	0 (0)
Education					
Years, mean (SD)	11 (5)	9 (5)	13 (5)	15 (3)	14 (3)
High school or less, (N, %)	3748 (65)	3206 (78)	243 (46)	112 (19)	187 (32)
College, (N, %)	1393 (24)	689 (17)	185 (35)	261 (45)	258 (44)
Graduate, (N, %)	662 (11)	220 (5)	97 (18)	205 (35)	140 (24)
Number of study visits, median (IQR)	3 (2, 5)	3 (2, 4)	4 (2, 6)	3 (2, 5)	7 (4, 12)
Follow-up time (years), median (IQR)	4 (1, 7)	4 (1, 8)	4 (2, 6)	3 (1, 4)	5 (2, 9)
Cognitive status, <i>n</i> (%)					
Cognitively normal	3363 (64)	2678 (65)	263 (51)	392 (68)	30 (42)
MCI	1020 (19)	731 (18)	178 (35)	86 (15)	25 (35)
Alzheimer's disease	896 (17)	706 (17)	73 (14)	100 (17)	17 (24)
Cognitive performance scores, mean (SD)					
General cognitive performance	48 (10)	47 (10)	51 (7)	54 (10)	56 (5)
Memory	53 (13)	53 (13)	50 (13)	56 (17)	55 (5)
Executive function	49 (12)	47 (12)	53 (8)	54 (11)	55 (7)

SD: standard deviation; MCI: Mild Cognitive Impairment.

Table 2

Neuropsychological tests from each study, included in general cognitive performance, memory, and executive functioning factors in the present integrative data analysis

Cognitive Domain	Neuropsychological Test	Variables from the tests	ADAMS HRS	Washington Heights-Inwood Columbia Aging Project (WHICAP)	Spanish and English Neuropsychological Assessment Scales (SENAS)	Duke Mental Health and Aging Study	Neurocognitive Outcomes of Depression in the Elderly (NCODE)
	Consortium to Establish a Registry for Alzheimer's Disease (CERAD) word list learning	Immediate and delayed recall	0.85		0.85	0.85	0.85
	Word recall test	Immediate and delayed recall			0.83		
	Wechsler Memory Scale, Logical Memory I & II	Story A, immediate recall	0.94		0.94	0.94	0.94
		Story A, delayed recall	0.97		0.97	0.97	0.97
		Story B, immediate recall	0.97				0.97
		Story B, delayed recall	0.98				0.98
Memory	Constructional Praxis	Total recall	0.75	0.75	0.75	0.75	0.75
		Delayed recall				0.73	0.73
		Recognition	0.74			0.74	0.74
Benton Visual Retention Test (BVRT)	Correct					0.88	0.88
		Matching		0.68			
		Recognition		0.69			
Bushke Selective Reminding Test	Immediate recall			0.88			
		Long-term recall		0.83			
		Delayed recall		0.80			
		Recognition		0.79			
Executive Functioning	Trail Making Test	Part A	0.84		0.84	0.84	0.84
		Part B	0.92		0.92	0.92	0.92
	Color Trail Making Test	Part 1		0.68			
		Part 2		0.66			
Digit Span Test	Forward and backwards		0.71		0.71	0.71	0.71
			0.80		0.80	0.80	0.80
	Digit Symbol Substitution, Wechsler Memory Scale-Revised	Number of number/symbol matches			0.78	0.78	
	Symbol-Digit Modalities Test	Number of	0.93				0.93

Cognitive Domain	Neuropsychological Test	Variables from the tests	ADAMS HRS	Washington Heights-Inwood Columbia Aging Project (WHICAP)	Spanish and English Neuropsychological Assessment Scales (SENAS)	Duke Mental Health and Aging Study	Neurocognitive Outcomes of Depression in the Elderly (NCODE)
		number/symbol matches					
	Controlled Oral Word Association Test	Sum of F,A,S words	0.81	0.75	0.81	0.81	0.81
		Sum of F,L words			0.70		
	Semantic fluency	Animals	0.80	0.76	0.80	0.80	0.80
		Vegetables			0.75	0.75	
		Fruits			0.78		
	Ascending digits	One list sorting		0.80			
		Two list sorting		0.76			
	Shape recognition	Time to complete		0.37			
	Consonant trigrams	Time to complete		0.34			
	WAIS Similarities	Number reported		0.75			
	WAIS Identities/Oddities	Number reported		0.61			
	Repetition	Maximum span		0.53			
	Comprehension	Total recall		0.67			
	Object naming	Number of objects named			0.82		
	Boston Naming Test	15-item	0.79	0.79	0.79		0.79
		30-item			0.73	0.73	
		60-item			0.81		
Other tests	Orientation	Total recall		0.68			
	Mini Mental Status Exam (MMSE)	Sum score	0.89		0.89	0.89	0.89
	Pattern recognition	Number of patterns			0.69		
	Rosen drawing test	Drawing score		0.55			
	Spatial Localization	Number of patterns			0.70		
	Total number of cognitive tests		10	15	14	10	10
	Total number of cognitive test indicators		16	22	22	17	18

Legend. Numbers in cells are standardized factor loadings, from a model for the overall cognitive performance factor, for the Legend. Numbers in cells are standardized factor loadings, from a model for the overall cognitive performance factor, for the cognitive test indicator representing the variable in the row for the dataset in the column.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript