Published in final edited form as:

Psychol Aging. 2009 March ; 24(1): 116–128. doi:10.1037/a0013421.

Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons

Dan Mungas¹, Bruce R. Reed^{1,2}, Sarah Tomaszewski Farias¹, and Charles DeCarli¹

¹Department of Neurology, School of Medicine, University of California, Davis

²Veterans Administration Northern California Health Care System

Abstract

This study examined how age and education influence the relationship between neuropsychological test scores and brain structure in demographically diverse older adults spanning the range from normal cognition to dementia. A sample of 351 African Americans, 410 Hispanics, and 458 Caucasians received neuropsychological testing; volumetric MRI measures of total brain, white matter hyperintensity, and hippocampus were available for 79 African Americans, 102 Hispanics, and 134 Caucasians. Latent variable modeling was used to examine effects of age, education, and brain volumes on test scores and determine how much variance brain volumes explained in unadjusted and age and education adjusted scores. Age adjustment resulted in weaker relationships of test scores with MRI variables and adjustment for ethnicity yielded stronger relationships. Education adjustment increased relationships with MRI in the combined sample and in Hispanics, made no difference in Caucasians, but decreased some associations in African Americans. Results suggest that demographic adjustment is beneficial when demographic variables are strongly related to test scores independent of measures of brain structure, but adjustment has negative consequences when effects of demographic characteristics are mediated by brain structure.

Keywords

Neuropsychological tests; ethnicity; brain structure; age; education; structural MRI

There is controversy about whether scores of cognitive and neuropsychological tests used in clinical identification of cognitive impairment in older persons should be adjusted for effects of demographic variables including age and education. It is a well-established practice in clinical psychology and neuropsychology to adjust scores for age, typically by using age-group specific norms. This practice likely derives from the field of intellectual assessment of children and adults. Age-group norms have obvious relevance for children, where normal development results in substantial increases in cognitive abilities from infancy to adulthood. Age adjustment throughout adulthood, in contrast, raises different issues. Developmental changes across the adult lifespan are small in comparison with developmental trajectories from early childhood to early adulthood, and are not due to brain maturation, but rather, result from learning and from acquired brain injury. With respect to education, norms of neuropsychological tests are sometimes adjusted for effects of education, but most often are not. This, for example, has been the approach used for the Wechsler scales (The Psychological Corporation, 1997) for more

than 50 years. However, the need for education adjustment has been reevaluated in light of more recent studies showing prominent education effects in ethnically diverse samples. A number of studies have shown that using norms developed for non-minority populations to interpret test scores of minorities leads to a large number of false-positive errors; cognitive impairment is erroneously identified in a relatively large percent of cognitively normal minorities (Fillenbaum *et al.*, 1990; Gasquoine, 1999; Manly *et al.*, 1998; Ramírez *et al.*, 2001; Stern *et al.*, 1992). Education is an important factor that accounts for differences in test scores between minorities and non-minorities, and so adjusting for education might improve clinical accuracy for detecting cognitive impairment in older ethnic minority individuals by improving diagnostic specificity.

The controversy about demographic adjustment of neuropsychological tests most frequently has been debated in the context of race or ethnic group-specific norms. Proponents have argued that group-specific norms reduce assessment bias, and specifically, help to equate test sensitivity and specificity across racial and ethnic groups (Manly & Echemendia, 2007; Mungas *et al.*, 1996). It is well known that sensitivity and specificity are inversely related, that sensitivity is increased at the expense of decreased specificity, and vice versa. Thus, it is not surprising that use of group-specific norms tends to increase specificity for minority groups at the expense of sensitivity, but increases sensitivity for Caucasians at the expense of specificity (Mungas *et al.*, 1996). The argument against group-specific norming is, in part, that test scores capture real differences in ability and that “norming away” such differences makes it difficult to determine and remedy the underlying causes of such differences such as discrepancies in prenatal care, educational opportunities, etc. (Brandt, 2007). Put slightly differently, this view is that group-specific norms effectively reduce variance in neuropsychological test scores that is actually biologically relevant, e.g. associated with differences in brain structure and function. Thus, adjusting for group effects might actually decrease sensitivity to the brain changes that are the focus of the neuropsychological assessment.

A previous study from our group (Mungas *et al.*, 1996) examined effects of age and education adjustment of the Mini-Mental State Examination (Folstein *et al.*, 1975) and concluded that adjusted scores showed less diagnostic bias across Hispanics and Caucasians. A subsequent reexamination of the results of this study (Kraemer *et al.*, 1998) concluded that overall diagnostic accuracy was slightly better with raw test scores, and these authors argued that adjustment of scores was not advisable because it decreased validity. Both of these studies examined simultaneous adjustment for both age and education.

Age effects on cognitive test scores are well known although their basis is not entirely understood. While it is clear that age related diseases like Alzheimer's disease (AD) cause cognitive decline, it is less clear whether age-associated changes in the absence of overt disease are intrinsic to aging itself or rather, are an effect of sub-clinical accumulations of pathologic changes (Keller, 2006). Education also has ubiquitous effects on test performance, and this is particularly important in multicultural applications of cognitive tests where there is great heterogeneity of education (Gasquoine, 1999; Mungas *et al.*, 1996; Mungas *et al.*, 2005b; Shadlen *et al.*, 2001; Stricks *et al.*, 1998). Besides its basic effects on development of cognitive skills and acquisition of knowledge, education might have long term effects on brain development and structure that impart cognitive reserve, that is, a protective effect that moderates direct effects of neuropathology on cognition (Richards & Dreary, 2005; Roe *et al.*, 2007). Advanced age and low education are risk factors for diseases like Alzheimer's disease that cause cognitive impairment (Caamano-Isorna *et al.*, 2006), so it is possible that adjustment for these variables effectively removes variance in cognitive outcomes that is associated with disease, and this could decrease sensitivity to disease effects. However, age adjustment continues to be routine and education based norms have increasingly been

incorporated into interpretation of test scores (Fillenbaum *et al.*, 2001; Gontkovsky *et al.*, 2002; Lucas *et al.*, 2005).

Previous research in this area generally has used clinical diagnosis as the external validity standard, and the primary outcomes of interest have been effects of demographic adjustment on diagnostic sensitivity and specificity. Different cut scores change the relative balance of sensitivity and specificity, and consequently, the appropriateness of demographic adjustment by changing cut scores is primarily a matter of the specific goals for the assessment, specifically, the relative importance of sensitivity versus specificity.

However, adjustment of test scores may also be done using regression equations that quantify the effects of age, gender, ethnicity, etc. (e.g. (Fillenbaum *et al.*, 2001)) and statistically remove those effects from the score. This approach has the potential to alter the fundamental relationship of scores to outcomes, and if so, this becomes an important factor in evaluating whether it is beneficial to adjust test scores. If adjustment weakens the relationship of the test variable with the outcome, then sensitivity will necessarily be lower for a given level of specificity, but if adjustment improves validity the opposite will hold. The effects of demographic adjustment on this type of external validity of neuropsychological test scores have not been well studied.

The purpose of this study was to examine relative validity of age adjusted, education adjusted, and unadjusted neuropsychological test scores in a demographically diverse sample. Unlike previous studies, we used quantitative measures of brain volumes derived from structural MRI as the standard for evaluating effects of demographic adjustment of test scores. MRI has important advantages as an external standard over clinical diagnosis. While clinical diagnosis relates to biological changes, it nevertheless relies on formal or informal assessment of cognitive functioning that might be subject to the same biases as cognitive test scores, and so, is not optimal as a validation criterion. Structural MRI is a biological marker that is strongly associated with pathological features related to degenerative and vascular diseases that can serve as a proxy for common brain diseases of aging. While there is not yet a sufficient empirical base to conclude that MRI changes have the same meaning in different racial or ethnic groups, derivation of MRI measures is blind to demographic characteristics, and in this regard MRI provides unbiased measurement of brain structure.

Further, MRI measures of brain volume are associated with cognitive function, both in normal aging and across the full spectrum of cognition in aging. Both whole brain and hippocampal volume correlate robustly with cognitive dysfunction in older adults (Golomb *et al.*, 1994; Jack *et al.*, 1992; Kramer *et al.*, 2004; Mungas *et al.*, 2001), and also predict longitudinal cognitive decline in MCI and AD (Jack *et al.*, 1999; Mungas *et al.*, 2005a). The two most common causes of cognitive impairment and dementia in later life are AD and cerebrovascular disease (CVD). Whole brain and hippocampal atrophy have proven to be particularly useful in AD. Hippocampal atrophy predicts AD at autopsy (although it is not entirely specific (Jack *et al.*, 2002)) and correlates with the severity of AD pathology (Gosche *et al.*, 2002; Jack *et al.*, 2002). MRI has been used to quantify infarcts and white matter lesions and MRI evidence of these lesions is considered diagnostic of CVD. CVD is also associated with cortical atrophy (Du *et al.*, 2005; Fein *et al.*, 2000).

Whatever brain atrophy represents, it is clearly associated with major diseases of aging and diminished cognitive function, which along with unbiased measurement, makes it an excellent external standard for evaluating validity of neuropsychological tests. In this study, we directly evaluated how adjustment of test scores affects relationships with structural MRI measures. In addition, we directly examined whether adjustment for age and education enhances or detracts

from test validity within a large and demographically diverse sample of older persons, and also within more homogenous subgroups.

Methods

Participants

Participants were 1219 persons recruited by the UC Davis Alzheimer's Disease Center under protocols designed to increase representation of ethnic minorities and maximize heterogeneity of cognitive functioning. There were 458 Caucasians, 410 Hispanics, and 351 African Americans; 295 Hispanics were tested in Spanish, and all others were tested in English. A community screening program designed to identify and recruit individuals with cognitive functioning representative of the community dwelling population identified 1066 individuals (340 Caucasians, 388 Hispanics, 338 African Americans). The remaining 153 were initially seen for clinical evaluation at a university memory/dementia clinic and referred for research. An additional 67 individuals who were not Hispanic, Caucasian, or African Americans were part of this overall sample but were excluded from analyses for this study.

All community recruits were 60 years of age or older. Clinical patients under 60 were included if they were being evaluated for cognitive impairment associated with diseases of aging. Inclusion criteria included ability to speak English or Spanish. All participants signed informed consent under protocols approved by institutional review boards at UC Davis, the Veterans Administration Northern California Health Care System, and San Joaquin General Hospital in Stockton, California.

A sub sample was referred for clinical evaluation and a research MRI on the basis of Spanish and English Neuropsychological Assessment Scales (SENAS) measures of episodic memory, semantic memory, attention span, visual spatial abilities, and verbal abstraction. A 25% random sample of those with normal cognition were invited to participate in clinical evaluation and MRI, and all with memory or non-memory cognitive impairment were invited to participate. Exclusion criteria for selection in this stage included unstable major medical illness, major primary psychiatric disorder (history of schizophrenia, bipolar disorder, or recurrent major depression), and substance abuse or dependence in the last five years.

The sub sample in this study that received MRI included 315 individuals (134 Caucasians, 102 Hispanics, 79 African Americans). These participants received a multidisciplinary clinical evaluation that included a detailed medical history, physical exam, and neurological exam. A bilingual physician examined Spanish-speaking patients. A family member or other informant with close contact with the participant was interviewed to obtain information about level of independent functioning. Diagnosis of cognitive syndrome (Normal, Mild Cognitive Impairment (MCI), Demented) and, in the instance of dementia, underlying etiology was made according to standardized criteria and methods. Clinical diagnosis was not a primary variable in analyses for this study and was used primarily to describe the clinical characteristics of this sample. Clinical status of the full sample was estimated as follows. Sampling percentages were used as weights to relate the sub sample that had received clinical evaluation back to the overall sample and to estimate the prevalence of specific diagnostic categories in the whole sample. Estimated prevalence by diagnosis was: cognitively normal – 64.4%, MCI – 25.3%, and demented – 10.2%. Percentages within ethnic subgroups were similar, especially for MCI, but African Americans were less likely to be demented (5.8%) and more likely to be normal (69.5%) and Hispanics were more likely to be demented (14.2%) and less likely to be normal (61.9%).

Neuropsychological Measures

The Spanish and English Neuropsychological Assessment Scales (SENAS (Mungas & Reed, 2000; Mungas *et al.*, 2004; Mungas et al., 2005b; Mungas *et al.*, 2005c)) were used to measure four specific domains of cognitive functioning. The SENAS is the result of an extensive development process that has used item response theory methodology to create English and Spanish language measures of cognitive domains that are relevant to the neuropsychological assessment of older patients. English and Spanish language versions of scales are psychometrically matched, but in addition, measures of specific cognitive domains are psychometrically matched within English and Spanish versions (Mungas et al., 2004). This study used a subset of SENAS tests that were averaged within domains to create composite measures: Object Naming and Picture Association (Semantic Memory), Word List Learning I and Word List Learning II (Episodic Memory), Pattern Recognition and Spatial Localization (Spatial). An Executive Function composite was created using a set of fluency and working memory measures that have been recently developed using the same methods as the original SENAS scales. Previous confirmatory factor analyses showed that these measures define the dimensions corresponding to the composite measures in this study (Mungas et al., 2004; Mungas et al., 2005c).

Quantitative MRI Measures

Brain imaging was obtained at the University of California at Davis MRI Research Center on a 1.5T GE Signa Horizon LX Echospeed system or the Veterans Administration at Martinez on a 1.5 T Marconi system. Comparable imaging parameters were used at each site as follows:

1. Axial spin echo, T2 weighted double echo image with TE1 equal to 20 ms, TE2 equal to 90 ms, TR equal to 2420 ms, with a field of view of 24 cm and a slice thickness of 3 mm.
2. Coronal 3D spoiled gradient recalled echo (IR-prepped SPGR) acquisition, T1 weighted image with TR equal to 9.1 ms, a flip angle of 15 degrees, a field of view 24 cm, and a slice thickness of 1.5 mm.
3. Axial high resolution FLAIR image with a TE1 of 120 ms, a TR of 9000 ms, a TI of 2200 ms, and 24 cm field of view with a slice thickness of 3 mm.

Analysis of brain and white matter hyperintensity (WMH) volumes was based on a Fluid Attenuated Inversion Recovery (FLAIR) sequence designed to enhance WMH segmentation (Jack *et al.*, 2001). Images were orientated parallel to a hypothetical line connecting the Anterior Commissure (AP) and Posterior Commissure (PC). WMH segmentation was performed in a two-step process (DeCarli *et al.*, 1992; DeCarli *et al.*, 1999). In brief, non-brain elements were manually removed from the image by operator guided tracing of the dura matter within the cranial vault including the middle cranial fossa, but excluding the posterior fossa and cerebellum. The resulting measure of the cranial vault was defined as the total cranial volume (TCV) and was used to correct for differences in head size.

The first step in image segmentation required the identification of brain matter. Image intensity nonuniformities (DeCarli *et al.*, 1996) were then removed from the image and the resulting corrected image was modeled as a mixture of two Gaussian probability functions with the segmentation threshold determined at the minimum probability between these two distributions (DeCarli et al., 1992). Once brain matter segmentation was achieved, a single Gaussian distribution was fitted to the image data and a segmentation threshold for WMH was a priori determined at 3.5 SDs in pixel intensity above the mean of the fitted distribution of brain parenchyma. Morphometric erosion of two exterior image pixels was also applied to the brain matter image before modeling to remove the effects of partial volume CSF pixels on WMH

determination. Intra- and inter-rater reliability for these methods are high (DeCarli *et al.*, 2005).

Boundaries for the hippocampus were manually traced from the coronal 3D-T1 weighted images after reorientation along the axis of the left hippocampus. While the borders were traced on the coronal slices, corresponding sagittal and axial views were simultaneously presented to the operator in separate viewing windows in order to verify hippocampal boundaries. The rostral end of the hippocampus was identified using the sagittal view to distinguish between amygdala and the head of the hippocampus. The axial view was used as a separate check. In anterior sections, the superior boundary of the hippocampus was the amygdala. In sections in which the uncus lies ventral to caudal amygdala, the uncus was included in the hippocampus. In more posterior sections that do not contain amygdala, the hippocampal (choroid) fissure and the superior portion of the inferior horn of the lateral ventricle formed the superior boundary. The fimbria were excluded from the superior boundary of the hippocampus. The inferior boundary of the hippocampus was the white matter of the parahippocampal gyrus. The lateral boundary was the inferior (temporal) horn of the lateral ventricle, taking care in posterior sections to exclude the tail of the caudate nucleus. The posterior boundary of the hippocampus was the first slice in which the fornices were completely distinct from any gray/white matter of the thalamus. Intra-rater reliability determined for both right and left hippocampus using this method is quite good with intraclass correlations (ICCs) of .98 for right hippocampus and .96 for left hippocampus.

A sample of 11 individuals received scans on both the Martinez Veterans Administration and UC Davis Research scanners for quality control purposes, and volumetric measures used in this study were independently derived and compared. ICCs comparing volumes from the two scanners were .98 for TCV, .94 for total brain matter, .92 for total hippocampus, and .91 for WMH.

WMH was log transformed to achieve a normal distribution. Brain matter and hippocampus volumes were normalized to TCV prior to analyses to control for differences in size of the intracranial vault. These two variables were each linearly regressed on TCV using only normal cases to minimize contamination of results by disease related atrophy. The resulting regression equations were used to calculate predicted brain and hippocampal volumes for each individual, and these predicted values were subtracted from the actual observed values. This resulted in normalized measures: Brain Matter and Hippocampus. Thus, for example, Brain Matter was the difference (in cubic centimeters) between the measured total brain matter volume and the total brain matter volume predicted by TCV in cognitively normal individuals. Log(WMH) (White Matter Hyperintensity in subsequent results) was not normalized because, conceptually, a normal value is zero, and empirically, log(WMH) was not related to TCV.

Data Analysis

Latent variable modeling was performed using the Mplus application (Muthén & Muthén, 2004, 2006). The four neuropsychological test scores (Semantic Memory, Episodic Memory, Executive Function, Spatial) were the primary dependent variables, the normalized MRI measures (Brain Matter, Hippocampus, White Matter Hyperintensity) were the primary independent variables, and age and education were also included as independent variables in selected analyses. MRI data was available for about 26% of the total sample and was missing by design for the remainder. A maximum likelihood estimator was used with the missing data option of Mplus. Mplus uses full information maximum likelihood estimation with missing data, and this method efficiently uses all available data to estimate model parameters. Missing values, MRI variables primarily in this study, are not directly imputed. Rather, the observed means, variances, and covariances, which can be based on different numbers of cases for different variables, are used for estimating model parameters. This approach provides unbiased

estimation of model parameters under conditions of missing at random, and is preferably to the more traditional deletion of cases with any missing data (listwise deletion), where biased estimation is of greater concern (Schafer & Graham, 2002). The condition of missing at random is satisfied when missingness of data elements (in this case, MRI values in non-selected individuals) is not a function of missing data elements or can be explained by observed variables (Little & Rubin, 2002). Since selection for MRI was determined by neuropsychological variables that were included in analytic models and by random selection of neuropsychologically normal cases, missing at random is a reasonable assumption for this dataset.

The overall analytic model for one neuropsychological test score, Episodic Memory, is presented in Figure 1. All four neuropsychological test scores were incorporated in the same analysis; Figure 1 shows the model for one exemplar to enhance clarity of presentation. This model includes a measurement component that defines latent variables and a structural component that examines relationships among the latent variables, and observed Age and Education. For the measurement component, latent variables first were created that corresponded to the residuals of the four neuropsychological variables when regressed on Age or Education. For example, Episodic Memory was regressed on Education and a latent variable (episodic memory) defined by a single indicator, observed Episodic Memory, was created to capture the residual variance from that regression. The residual variances of the observed test scores were fixed at .15 times the sample variance for the indicator variable. The value of .15 corresponds to measurement error, and was selected based on previous results that demonstrated reliability of SENAS scales of approximately .85 across a broad range of ability (Mungas et al., 2004). MRI variables also were modeled as latent variables with single indicators (Brain Matter volume (BM) for the bm latent variable, etc.), and variances were fixed at .10 times the sample variances to correspond to conservatively estimated reliability of .90. For the structural component, the relationships between the latent MRI variables (bm, hc, wmh) and the residual cognitive scores (episodic memory in Figure 1) after accounting for Age or Education were of primary interest. The relationship of Age or Education with the latent MRI variables was also of interest.

Separate models were estimated to examine effects of adjustment for Age and Education. The process for Age was as follows. First, Age effects on neuropsychological test scores were freely estimated as were MRI latent variable effects on residual test scores, but Age effects on MRI latent variables were constrained to 0. This corresponds to removing age effects from the test scores, but not from the MRI variables, which is what would happen in a clinical situation where test scores are adjusted for Age. The R^2 values describing the amount of variance in residual tests scores explained jointly by the three MRI variables provided indices of the strength of association of Age adjusted test scores with brain structure. This model was then re-estimated constraining the effect of Age on observed test scores to 0, and R^2 values from this analysis were used as indices of effect sizes for unadjusted scores. The same process was followed to obtain Education adjusted scores. Age and Education adjustment effects were evaluated in the full sample, and in each of the ethnic subgroups.

A comprehensive model that simultaneously incorporated effects of Age, Education, and ethnicity was estimated to clarify relationships in the data that help to explain results of Age and Education adjustment. This was a multiple group analysis that allowed for independent estimation of model parameters in the three ethnic subgroups. All four test scores were incorporated in this model. Age and Education effects on observed test scores were freely estimated. Age and Education effects on MRI variables were initially constrained to 0. Regression coefficients of test scores on Age and Education and of residual test scores on latent MRI variables were initially constrained to be equal across ethnic groups, and the correlation of Education with Age was constrained to be 0 in all three ethnic groups. Intercepts were freely

estimated for ethnic groups. Modification indices (Muthén & Muthén, 2004, 2006) were examined to identify constrained effects (Age and Education effects on MRI variables constrained to 0, correlation of Age and Education constrained to 0, regression coefficients constrained to be equal across ethnic subgroups) that would improve model fit if freely estimated. These effects were then freely estimated to arrive at a best fitting model estimating the pathways depicted in Figure 1 in the three ethnic subgroups.

Results

Sample characteristics are presented in Table 1. Mean education markedly differed across ethnic groups in both the full and MRI subsamples. Hispanics had less than 8 years of education on average, compared with 13 for African Americans and 14 for Caucasians. Standard deviations were also nearly two times larger for Hispanics than African Americans or Caucasians reflecting broader variability of education in Hispanics. Mean age was slightly lower in Hispanics than African Americans (2 to 3 years on average) and Caucasians (3 to 4 years). Percentages of males were somewhat higher in the Caucasians than in the African Americans and Hispanics.

Table 2 shows average neuropsychological test scores and MRI values for the three ethnic subgroups. Hispanics had the lowest test scores for all four variables. Scores of African Americans were intermediate to those of Hispanics and Caucasians for all but Episodic Memory, where African American and Caucasian means were essentially the same. Hispanics had the largest average, normalized Brain Matter volumes. Mean Brain Matter for Hispanics was significantly greater than for African Americans, and African Americans had significantly greater mean volumes than Caucasians. Hispanics still had larger Brain Matter after controlling for age as well as clinical diagnosis (normal versus MCI versus demented). Mean Hippocampus volume did not significantly differ across groups ($p=.16$). Table 2 shows raw white matter hyperintensity volumes so that results are directly interpretable in the same units as the other MRI volumes. Hispanics on average had significantly smaller mean White Matter Hyperintensity (log transformed) in comparison with African Americans and Caucasians, which did not differ. Ethnic group differences in White Matter Hyperintensity volumes were no longer significant after controlling for age.

Tables 3 and 4 present the critical results. Table 3 compares the percents of variance that MRI variables explained for unadjusted, Age adjusted, and Education adjusted scores in the combined sample of African Americans, Caucasians, and Hispanics and in specific subgroups. In the combined sample, MRI variables consistently explained the most variance in Education adjusted cognitive scores, ranging from 1.2 to 3.3 times that in unadjusted scores. MRI variables accounted for less variance in the Age adjusted score than in the unadjusted score for all but semantic memory where the difference was negligible. For episodic memory, the MRI effect on the Age adjusted score was less than half that for the unadjusted score, and for executive function and spatial, Age adjusted scores had negligible relationships with MRI variables.

Table 3 also presents percent of variance in different scores explained jointly by MRI variables within specific ethnicity subgroups. Several important patterns emerge. First, R^2 values for unadjusted scores were consistently higher in all three ethnic subgroups than in the combined sample. Second, Age adjusted scores were least strongly related to MRI variables in all subgroups. Third, effects of Education adjustment differed across subgroups. Education adjustment yielded stronger relationships with MRI in Hispanics, made no clear difference in Caucasians, and in African Americans, Education adjustment made no difference for two cognitive scores (semantic memory and spatial), but decreased relationships with MRI for episodic memory and executive function. Thus, Education adjustment yielded the strongest

relationships with MRI for Hispanics, unadjusted and Education adjusted scores were equivalent for Caucasians, and unadjusted scores were equivalent to or better than Education adjusted scores for African Americans. This pattern of results is illustrated in Figure 2, which presents R^2 values for episodic memory within the combined sample and the ethnic subgroups.

Table 4 shows effects of individual MRI variables on each cognitive variable for unadjusted, Age adjusted, and Education adjusted scores, using the combined sample of African Americans, Hispanics, and Caucasians. Where significant relationships were found, MRI effects on Education adjusted scores were generally equal to or larger than effects on unadjusted scores, which were equal to or larger than effects on Age adjusted scores.

A final multiple group model that included both Age and Education effects was estimated to clarify patterns of relationships among all of the variables of interest in this study. All four cognitive test scores were simultaneously included in this model. The initial model constrained Age and Education correlations with MRI to zero in all groups, specified no correlation between Age and Education in all groups, and constrained regression coefficients to be equal in the three groups. Modification indices were used to identify and correct sources of model misfit. This process resulted in freely estimating Age correlations with MRI variables, with coefficients that were equal across groups, and freely estimating correlations of Age and Education in African Americans. In addition, model fit was improved by freely estimating the regression coefficient of residual spatial on latent wmh in African Americans, by freely estimating effects of Education on observed Semantic Memory in the three groups, freely estimating Age effects on Executive Function in all three groups, and freely estimating the effect of Age on Spatial in African Americans. These modifications resulted in a well fitting model ($\chi^2 = 67.3$, $df = 51$, $p = .06$, $CFI = .994$, $TLI = .989$, $RMSEA = .028$, $SRMR = .050$).

The final, best fitting, multiple group model identified only a few parameters that significantly differed across the ethnic groups. Table 5 shows relationships of MRI and cognitive variables in the three ethnic groups as estimated in the final multiple group model. The only significant group difference was that spatial was much more strongly related to wmh in African Americans. Table 6 shows relationships from the final multiple group model of observed cognitive scores with Age and Education. Simple correlations of Age and Education with cognitive scores are included for comparison purposes. Semantic Memory was significantly related to Education in all three groups but was most strongly related in Hispanics and least strongly related in African Americans. The Semantic Memory-Education relationship for Hispanics was significantly greater than that for African Americans ($p < .001$), but other group comparisons were not significant. The relationship of age with Executive Function was significant but weak in Hispanics and was non-significant for African Americans and Caucasians. Spatial was weakly related to age in African Americans, though in an unexpected, positive direction. Otherwise, relationships of cognitive variables to Age, Education, and MRI variables did not significantly differ across groups.

Figure 3 presents standardized path coefficients (solid lines) from the final model for Executive Function and Brain Matter. For purposes of presentation, within group correlations of age with education were freely estimated for Figure 3, though they were not significantly different from zero in the final model. Simple unadjusted associations of Age and Education with observed Executive Function (dashed lines) are also included for comparison purposes. The top frame presents these results for African Americans, the middle frame for Hispanics, and the bottom for Caucasians. The pattern of results was similar for other test scores and MRI variables. It is particularly noteworthy that simple Age effects on test scores (dashed lines) were substantially stronger than Age effects independent of MRI variables (solid lines). For African Americans, a model adjusted coefficient of -0.03 was found in comparison with an unadjusted value of -0.26; for Hispanics coefficients were -0.17 versus -0.41, and for Caucasians, values were -0.10

versus -0.36. Age was very strongly related to all three MRI variables (-0.55, -0.60, -0.53) and this pattern of results suggests that much of the Age effects on test scores were mediated by Age associated differences in MRI variables. For Education, effects adjusted for MRI variables and simple bivariate effects were the same for Caucasians (.29 versus .28) and Hispanics (.46 versus .44), but adjusted effects were weaker for African Americans (.32 versus .39). This means that Education explained about 15% of the variance in Executive Function in African Americans when it was a lone predictor, but explained about 10% of the variance independent of other variables in the full model. Education was not directly related to MRI in African Americans, but was indirectly related via mediating effects of Age. These results suggest that Age adjustment had a negative effect on validity because it removed variance from the test score that is associated with brain structure. Education adjustment in Caucasians and Hispanics removed variance from test scores that is unrelated to brain structure, and so did not adversely affect relationships with MRI variables. In African Americans, Education adjustment likely had a negative effect because of the association of Education and Age; adjusting for Education also adjusted for Age.

Finally, we repeated analyses excluding cases that were scanned on the Martinez Veterans Administration scanner. This was done to assure that scanner differences were not contributing to spurious findings. The sample for these analyses included about two thirds of the MRI cases (n=216) and all of the cases who did not receive MRI, resulting in a total sample of 1120. The pattern of results was substantively the same as in the full sample.

Discussion

There were three major findings from this study: 1) Age adjusted test scores were not as strongly related to MRI variables as unadjusted scores, and this was true in all ethnic subgroups and in the combined sample. 2) MRI effects on test scores were stronger in the individual ethnic subgroups than in the combined sample. 3) Effects of education adjustment were variable; education adjusted scores were more strongly related to MRI in the combined and Hispanic samples, were about equal to unadjusted scores in Caucasians, and were inferior to or equal to unadjusted scores in African Americans. Thus, these results suggest that adjusting test scores for age has negative effects on validity, adjusting for ethnic group has positive effects, and adjusting for education has more variable and complicated effects.

At a descriptive level, the reason for adverse effects of adjusting for age can be discerned in Figure 3. Age was significantly related to test scores, and adjusting for age essentially removed age related variance from the test scores. But age was also strongly related to all three MRI variables, and much of the effects of age on test scores were mediated via the MRI variables. Consequently, removing age effects from test scores also removed variance that is associated with age related changes in brain structure.

Conceptually, the likely explanation for these results is that age is strongly associated with pathology and so removing age variance also removes effects of pathology on cognitive test scores. This is probably because Alzheimer's disease, common in old age, has a prodrome of many years before it is clinically identified. Significant cognitive changes are present years before diagnosis (Amieva *et al.*, 2005; Elias *et al.*, 2000), and changes in brain function appear decades before symptom onset (Reiman *et al.*, 2005), both findings presumably due to developing Alzheimer neuropathology. Similar processes likely occur with other diseases of aging. Because AD and other diseases become geometrically more common with age, the correlation between age and cognitive test scores, which is the basis of age corrections, is partly due to an underlying effect of incipient disease on cognitive function. Thus, to correct for "age" is to correct for both age and pathology, and this reduces validity for detecting pathology.

Age could also have effects on test scores due to cohort effects. That is, people born between 1915 and 1925 might have had very different experiences than those born between 1935 and 1945, and consequently, age effects on test scores could reflect different experiences along with age associated disease effects. This study directly examined the effects of age on test scores independent of MRI variables, and results showed that these effects were significant only for semantic memory in all three groups and executive function in Hispanics. The weak, positive correlation of age with spatial in African Americans may well reflect sampling error. These results show that brain structure explains most of the age related variance in test scores. The significant relationships of age with semantic memory independent of brain structure could represent cohort effects, though other explanations are possible, and the associations are weak, regardless of the source.

Education consistently was strongly related to test scores, but in contrast to age, education effects were independent of MRI variables in all but the African American subgroup. Consequently, education adjustment removed a substantial component of variability from test scores that was not related to brain structure, which in effect increased the salience of the variance component that was associated with brain structure. That is, the MRI effect is the signal of interest, and education contributed noise that obscured this signal. An argument against adjusting for education is that education might have effects on test scores that are mediated via disease. This study directly tested the extent to which education effects were mediated by one indicator of disease, brain structure, and failed to support this hypothesis, except in African Americans. In the African American group, education was related to brain structure via a mediating effect of age. In this case, age and education are positively correlated, perhaps reflecting rising educational opportunities for African Americans across the middle decades of the twentieth century. This secular trend confounds the relationships between age, education, and pathology; adjusting for education also adjusts, to some degree, for age; age and pathology are linked and consequently, education adjustment had a negative effect on validity in this group, most likely for the same reasons that age adjustment had a negative effect. Another consideration is that the Caucasian and African American sub samples were much more homogenous with respect to education, and education effects on all four cognitive variables were larger in Hispanics and the combined sample than in the non-Hispanic groups.

It is noteworthy that MRI effects were stronger in all ethnic subgroups than in the full sample. Sizeable ethnic differences in neuropsychological test scores are well documented, even in cognitively normal individuals (Manly et al., 1998; Mungas et al., 2005b). While there were ethnic group differences in average test scores in this study, relationships of test scores to MRI variables, age, and education were remarkably similar across the three ethnic groups. The explanation for why adjustment for ethnicity improves validity most likely is the same as for education. That is, there is substantial ethnicity variance in test scores that is independent of brain structure. Removing this noise from the overall test score essentially makes the signal related to brain structure more salient.

Associations of MRI measures with cognitive test scores were strikingly similar across groups. Episodic memory was independently related to all three MRI variables, executive function was related to brain matter and white matter hyperintensity, and spatial ability was related to brain matter. These relationships conform to a-priori expectations and did not differ across groups. Only one MRI-cognition relationship differed across groups; spatial ability was related to white matter hyperintensity only in African Americans. This finding requires replication and further study to clarify its scientific significance.

Semantic memory had the weakest associations with MRI variables in all individual groups and in the combined sample. Semantic memory is a component of crystallized intelligence, and as such, represents the accumulation of knowledge over a lifetime. Consequently, this

ability may be relatively resistant to impairment of brain structure and function, particularly in more mildly impaired individuals.

Three methodological issues merit further discussion. First, this study included participants with a broad range of cognitive function as the result of an explicit attempt to include individuals who would meet clinical criteria for normal cognition, mild cognitive impairment, and dementia. Our approach also explicitly included the full sample in analyses, and did not evaluate relationships within groups defined by degree of cognitive impairment. Brain diseases and injuries causing cognitive impairment are continuously variable, as are the cognitive effects of these diseases. Diagnoses of normal cognition, MCI, and dementia are labels that arbitrarily divide these continuous dimensions to help with communication of complex clinical information. Separate analyses within diagnostic subgroups would be problematic for several reasons. Methodologically, this strategy restricts variability and decreases sample size and statistical power, both of which obscure important effects. But there is a more compelling substantive problem with subgroup analyses. The relationship of neuropsychological test results to brain structure across the full practical range of both is of clinical importance, and this cannot be effectively studied within individual subgroups that are based on arbitrary divisions of that range. This is exemplified by studies that show small correlations of hippocampal volume with memory in normals (Van Petten, 2004) in contrast to striking correlations in clinically heterogeneous samples (Grundman *et al.*, 2003; Mungas *et al.*, 2001; Petersen *et al.*, 2000). A strength of this study was the broad variability of all of the variables of interest, demographic, cognitive, and MRI.

The second methodological consideration is that there were some differences across ethnic subgroups, for example, in normalized brain matter and in prevalence of clinical diagnoses. This raises a possibility that group difference in important background variables could result in spurious relationships among the primary variables of this study. The multi-group design is a major strength of this study. Even though there were differences across the ethnic subgroups, relationships among various variables were highly similar within the different groups. In particular, relationships of age and education to MRI variables were highly similar across groups, and the relationships of MRI variables to test scores were the same with only one minor difference. These similarities across very disparate groups help to establish the generality of results. The ability to evaluate different patterns of results in the different groups was an important feature that helped to clarify both that education adjustment had different effects in the different groups, but in addition, to identify confounding effects of age on education adjustment in the African American group.

Third, MRI measures were not as strongly related to cognitive test scores as was clinical diagnosis of normal versus MCI versus demented. Specifically, clinical diagnosis alone accounted for 21% of the variance in semantic memory, 46% in episodic memory, 30% of executive function, and 21% of spatial ability (results not shown). These results suggest that the MRI variables did not capture all of the clinically relevant information inherent in this sample. Additional measures of brain structure and function might enhance the clinical power of brain based biological measures while maintaining the advantages related to reduced measurement bias. An additional consideration is that correlations between cognitive and MRI measures might have been attenuated to some extent by a predominantly normal to mildly impaired sample. This is supported by a recent meta analysis of studies of the association between hippocampal volume and memory in cognitively normal individuals that showed positive but very small correlations (Van Petten, 2004). The sample in the present study was not entirely normal and correlations with hippocampal volume were stronger than in the Van Petten study, but range of biological variability is an important consideration when selecting brain based measures as validity standards.

The issue of demographic adjustment of neuropsychological test scores was recently revisited by Brandt (Brandt, 2007) in his presidential address to the International Neuropsychological Society. He directly addressed the issue of race/ethnic group norms and adjustments, and argued that this might have unintended adverse effects because valid group differences in ability that might result from underlying disease and brain injury could be removed from test scores, making them less sensitive to disease effects. For example, African American and Caucasians differ in prevalence of hypertension, and applying race specific norms to cognitive test scores might obscure hypertension effects that contribute to group differences in cognitive function. Manly and Echemendia (Manly & Echemendia, 2007) contrasted research on hypertension and cognition, and argued that there is a compelling body of evidence that specific scores of standard hypertension measures have the same meaning in different racial groups, but similar equivalence of cognitive test scores has not been established for neuropsychological tests. They concluded that the appropriateness of race/ethnicity adjustment is governed by a complex set of circumstances including the specific purpose of the test and effects of variables that are confounded with race or ethnicity, and emphasized the importance of research to clarify those variables and their effects.

This study used methods that can be useful to empirically test some of the issues raised by Brandt and Manly and Echemendia. Specifically, it presents a model for disentangling demographic effects on cognition that are mediated by brain variables from those that are unrelated to disease and brain variables relevant to neuropsychological assessment. The implications of these results for the question of whether or not it is best to adjust neuropsychological test scores for age and education can be understood from the perspective of test validity. Validity is purpose specific; no test is simply “valid”, but rather, a test is valid for a specific purpose. It must be emphasized that this study pertains to a small part of the age range, the 60+ end of the continuum, and results are relevant to assessment questions specific to this group. For this age range, a common and important use of neuropsychological tests is to detect the effects of brain pathology and diseases of aging. Results of this study indicate that age adjustment actually weakens validity with respect to one major purpose for neuropsychological assessment of older persons.

In contrast, the results of this study suggest that, in the context of detecting effects of brain pathology, it is beneficial to adjust for variables that affect test performance but are not related to disease and brain structure. Ethnicity fulfilled this criterion, as did education in some contexts. A further implication of the results is that adjustment for demographic variables is more important where there is considerable heterogeneity of the demographic variable of interest and where those variables account for a larger amount of variance in the test score. For example, education adjustment was most beneficial in the whole sample and the Hispanic sub group, both of which had considerable variability of education, but not in the more homogeneous Caucasian and African American sub groups.

An important limitation of this study is that these conclusions are based on a specific sample and specific neuropsychological tests and MRI variables. This is an area of research that has major practical implications, so there is a need to evaluate the generality of these results to other samples and assessment methods. In addition, there is a compelling body of literature to suggest that quality of education experience is an important variable above and beyond years of education (Manly *et al.*, 2004; Manly *et al.*, 1999; Manly *et al.*, 2002). This is not a variable that we examined, and it makes the point that there are additional demographic variables that might prove important when seeking to understand the complex relationships of demographic variables, disease processes, brain pathology, and cognition.

A practical implication of this study is that age adjustment in the context of neuropsychological assessment of older individuals is not advisable because it decreased sensitivity to brain

structure differences in all groups. When a sufficient sample size or adequate normative data is available for specific ethnic groups, adjustment of scores for ethnicity effects is advisable with respect to increasing sensitivity to brain differences. When adjustment for ethnicity is not practical due to insufficient sample size in a given study or lack of norms, education might serve as a useful proxy variable, but only if there is not a relationship of education to age in the population of interest. Finally, education adjustment does not appear to be beneficial within homogenous groups of Caucasian and African American older persons, but is beneficial within Hispanics and, by extension, any group in which there is a wide range of educational exposure.

Acknowledgments

This work was supported in part by grants AG10220, AG10129, and AG021028 from the National Institute on Aging, Bethesda, MD, and by the California Department of Public Health Alzheimer's Disease Program, contracts 06-55311, and 06-55312.

References

- Amieva H, Jacqmin-Gadda H, Orgogozo JM, Le Carret N, Helmer C, Letenneur L, et al. The 9 year cognitive decline before dementia of the alzheimer type: A prospective population-based study. *Brain* 2005;128(Pt 5):1093–1101. [PubMed: 15774508]
- Brandt J. 2005 ins presidential address: Neuropsychological crimes and misdemeanors. *Clin Neuropsychol* 2007;21(4):553–568. [PubMed: 17613978]
- Caamano-Isorna F, Corral M, Montes-Martinez A, Takkouche B. Education and dementia: A meta-analytic study. *Neuroepidemiology* 2006;26(4):226–232. [PubMed: 16707907]
- DeCarli C, Maisog J, Murphy DG, Teichberg D, Rapoport SI, Horwitz B. Method for quantification of brain, ventricular, and subarachnoid csf volumes from mr images. *J Comput Assist Tomogr* 1992;16(2):274–284. [PubMed: 1545026]
- DeCarli C, Massaro J, Harvey D, Hald J, Tullberg M, Au R, et al. Measures of brain morphology and infarction in the framingham heart study: Establishing what is normal. *Neurobiol Aging* 2005;26(4):491–510. [PubMed: 15653178]
- DeCarli C, Miller BL, Swan GE, Reed T, Wolf PA, Garner J, et al. Predictors of brain morphology for the men of the nhlbi twin study. *Stroke* 1999;30(3):529–536. [PubMed: 10066847]
- DeCarli C, Murphy DG, Teichberg D, Campbell G, Sobering GS. Local histogram correction of mri spatially dependent image pixel intensity nonuniformity. *J Magn Reson Imaging* 1996;6(3):519–528. [PubMed: 8724419]
- Du AT, Schuff N, Chao LL, Kornak J, Ezekiel F, Jagust WJ, et al. White matter lesions are associated with cortical atrophy more than entorhinal and hippocampal atrophy. *Neurobiol Aging* 2005;26(4):553–559. [PubMed: 15653183]
- Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of alzheimer disease: A 22-year prospective study of the framingham cohort. *Arch Neurol* 2000;57(6):808–813. [PubMed: 10867777]
- Fein G, Di Sclafani V, Tanabe J, Cardenas V, Weiner MW, Jagust WJ, et al. Hippocampal and cortical atrophy predict dementia in subcortical ischemic vascular disease. *Neurology* 2000;55(11):1626–1635. [PubMed: 11113215]
- Fillenbaum G, 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(4):502–509. [PubMed: 11396552]
- Fillenbaum G, Heyman A, Williams K, Prosnitz B, Burchett B. Sensitivity and specificity of standardized screens of cognitive impairment and dementia among elderly blacks and white community residents. *Journal of Clinical Epidemiology* 1990;43:651–660. [PubMed: 2370572]
- Folstein M, Folstein S, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;12:189–198. [PubMed: 1202204]

- Gasquoine PG. Variables moderating cultural and ethnic differences in neuropsychological assessment: The case of hispanic americans. *Clinical Neuropsychology* 1999;13(3):376–383.
- Golomb J, Kluger A, de Leon MJ, Ferris SH, Convit A, Mittelman MS, et al. Hippocampal formation size in normal human aging: A correlate of delayed secondary memory performance. *Learn Mem* 1994;1(1):45–54. [PubMed: 10467585]
- Gontkovsky ST, Mold JW, Beatty WW. Age and educational influences on rbans index scores in a nondemented geriatric sample. *Clin Neuropsychol* 2002;16(3):258–263. [PubMed: 12607139]
- Gosche KM, Mortimer JA, Smith CD, Markesbery WR, Snowdon DA. Hippocampal volume as an index of alzheimer neuropathology: Findings from the nun study. *Neurology* 2002;58(10):1476–1482. [PubMed: 12034782]
- Grundman M, Jack CR Jr, Petersen RC, Kim HT, Taylor C, Datvian M, et al. Hippocampal volume is associated with memory but not nonmemory cognitive performance in patients with mild cognitive impairment. *J Mol Neurosci* 2003;20(3):241–248. [PubMed: 14501003]
- Jack CR Jr, Dickson DW, Parisi JE, Xu YC, Cha RH, O'Brien PC, et al. Antemortem mri findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology* 2002;58(5):750–757. [PubMed: 11889239]
- Jack CR Jr, O'Brien PC, Rettman DW, Shiung MM, Xu Y, Muthupillai R, et al. Flair histogram segmentation for measurement of leukoaraiosis volume. *J Magn Reson Imaging* 2001;14(6):668–676. [PubMed: 11747022]
- Jack CR Jr, Petersen RC, O'Brien PC, Tangalos EG. Mr-based hippocampal volumetry in the diagnosis of alzheimer's disease. *Neurology* 1992;42(1):183–188. [PubMed: 1734300]
- Jack CR Jr, Petersen RC, Xu YC, O'Brien PC, Smith GE, Ivnik RJ, et al. Prediction of ad with mri-based hippocampal volume in mild cognitive impairment. *Neurology* 1999;52(7):1397–1403. [PubMed: 10227624]
- Keller JN. Age-related neuropathology, cognitive decline, and alzheimer's disease. *Ageing Res Rev* 2006;5(1):1–13. [PubMed: 16084778]
- Kraemer HC, Moritz DJ, Yesavage J. Adjusting mini-mental state examination scores for age and educational level to screen for dementia: Correcting bias or reducing validity? *International Psychogeriatrics* 1998;10(1):43–51. [PubMed: 9629523]
- Kramer JH, Schuff N, Reed BR, Mungas D, Du AT, Rosen HJ, et al. Hippocampal volume and retention in alzheimer's disease. *J Int Neuropsychol Soc* 2004;10(4):639–643. [PubMed: 15327742]
- Little, RJ.; Rubin, DB. *Statistical analysis with missing data*. Second edition. John Wiley and Sons; New York: 2002.
- Lucas JA, Ivnik RJ, Smith GE, Ferman TJ, Willis FB, Petersen RC, et al. Mayo's older african americans normative studies: Norms for boston naming test, controlled oral word association, category fluency, animal naming, token test, wrat-3 reading, trail making test, stroop test, and judgment of line orientation. *Clin Neuropsychol* 2005;19(2):243–269. [PubMed: 16019707]
- Manly JJ, Byrd DA, Touradji P, Stern Y. Acculturation, reading level, and neuropsychological test performance among african american elders. *Appl Neuropsychol* 2004;11(1):37–46. [PubMed: 15471745]
- Manly JJ, Echemendia RJ. Race-specific norms: Using the model of hypertension to understand issues of race, culture, and education in neuropsychology. *Arch Clin Neuropsychol* 2007;22(3):319–325. [PubMed: 17350797]
- Manly JJ, Jacobs DM, Sano M, Bell K, Merchant CA, Small SA, et al. Cognitive test performance among nondemented elderly african americans and whites. *Neurology* 1998;50(5):1238–1245. [PubMed: 9595969]
- Manly JJ, Jacobs DM, Sano M, Bell K, Merchant CA, Small SA, et al. Effect of literacy on neuropsychological test performance in nondemented, education-matched elders. *Journal of the International Neuropsychological Society* 1999;5(3):191–202. [PubMed: 10217919]
- Manly JJ, Jacobs DM, Touradji P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between african american and white elders. *Journal of the International Neuropsychological Society* 2002;8(3):341–348. [PubMed: 11939693]
- Mungas D, Harvey D, Reed BR, Jagust WJ, DeCarli C, Beckett L, et al. Longitudinal volumetric mri change and rate of cognitive decline. *Neurology* 2005a;65(4):565–571. [PubMed: 16116117]

- Mungas D, Jagust WJ, Reed BR, Kramer JH, Weiner MW, Schuff N, et al. Mri predictors of cognition in subcortical ischemic vascular disease and alzheimer's disease. *Neurology* 2001;57(12):2229–2235. [PubMed: 11756602]
- Mungas D, Marshall SC, Weldon M, Haan M, Reed B. Age and education correction of mini mental state examination for english and spanish speaking elderly. *Neurology* 1996;46:700–706. [PubMed: 8618670]
- Mungas D, Reed BR. Application of item response theory for development of a global functioning measure of dementia with linear measurement properties. *Statistics in Medicine* 2000;19:1631–1644. [PubMed: 10844724]
- Mungas D, Reed BR, Crane PK, Haan MN, González H. Spanish and english neuropsychological assessment scales (senas): Further development and psychometric characteristics. *Psychological Assessment* 2004;16(4):347–359. [PubMed: 15584794]
- Mungas D, Reed BR, Haan MN, Gonzalez H. Spanish and english neuropsychological assessment scales: Relationship to demographics, language, cognition, and independent function. *Neuropsychology* 2005b;19(4):466–475. [PubMed: 16060821]
- Mungas D, Reed BR, Tomaszewski Farias S, DeCarli C. Criterion-referenced validity of a neuropsychological test battery: Equivalent performance in elderly hispanics and non-hispanic whites. *Journal of the International Neuropsychological Society* 2005c;11:620–630. [PubMed: 16212690]
- Muthén, LK.; Muthén, BO. Mplus user's guide. Third addition. Muthén & Muthén; Los Angeles, CA: 2004.
- Muthén, LK.; Muthén, BO. Mplus. Statistical analysis with latent variables. User's guide. Muthén & Muthén; Los Angeles: 2006.
- Petersen RC, Jack CR Jr. Xu YC, Waring SC, O'Brien PC, Smith GE, et al. Memory and mri-based hippocampal volumes in aging and ad. *Neurology* 2000;54(3):581–587. [PubMed: 10680786]
- Ramírez M, Teresi JE, Silver S, Holmes D, Gurland B, Lantigua R. Cognitive assessment among minority elderly: Possible test bias. *Journal of Mental Health and Aging* 2001;7:91–118.
- Reiman EM, Chen K, Alexander GE, Caselli RJ, Bandy D, Osborne D, et al. Correlations between apolipoprotein e epsilon4 gene dose and brain-imaging measurements of regional hypometabolism. *Proc Natl Acad Sci U S A* 2005;102(23):8299–8302. [PubMed: 15932949]
- Richards M, Dreary IJ. A life course approach to cognitive reserve: A model for cognitive aging and development. *Annals of Neurology* 2005;58:617–622. [PubMed: 16178025]
- Roe CM, Xiong C, Miller JP, Morris JC. Education and alzheimer disease without dementia: Support for the cognitive reserve hypothesis. *Neurology* 2007;68(3):223–228. [PubMed: 17224578]
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychol Methods* 2002;7(2):147–177. [PubMed: 12090408]
- Shadlen MF, Larson EB, Gibbons LE, Rice MM, McCormick WC, Bowen J, et al. Ethnicity and cognitive performance among older african americans, japanese americans, and caucasians: The role of education. *Journal of the American Geriatrics Society* 2001;49:1371–1378. [PubMed: 11890499]
- Stern Y, Andrews H, Pittman J, Sano M, Tatemichi T, Lantigua R, et al. Diagnosis of dementia in a heterogeneous population: Development of a neuropsychological paradigm-based diagnosis of dementia and quantified correction for the effects of education. *Archives of Neurology* 1992;49:453–460. [PubMed: 1580806]
- 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. *Journal of the International Neuropsychological Society* 1998;4(4):311–318. [PubMed: 9656604]
- The Psychological Corporation. Wais-iii-wms-iii technical manual. The Psychology Corporation; San Antonio: 1997.
- Van Petten C. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: Review and meta-analysis. *Neuropsychologia* 2004;42(10):1394–1413. [PubMed: 15193947]

Figure 1.

Analytic model for data analysis. Observed variables are represented by rectangles and labels are capitalized, latent variables are represented by ovals and labels are in lowercase letters. BM = Brain Matter volume, HC = Hippocampus volume, WMH = White Matter Hyperintensity volume. Latent cognitive variables (episodic memory in this model) represent the residual observed Episodic Memory variance after accounting for effects of Age or Education. Latent cognitive and MRI variables assume and are adjusted for by estimates of measurement error.

Figure 2.

Variance in episodic memory explained by structural MRI variables (R^2) for unadjusted cognitive test scores (■), age-adjusted scores (□), and education-adjusted scores (■) in the combined sample and within ethnic subgroups.

Figure 3. Relationships among age, education, brain matter volume, and the executive function composite score. Numerical values are standardized coefficients from the final multiple group model including both education and age as independent variables. Solid lines represent independent relationships adjusted by other variables in the figure. Dashed lines for age and education indicate the simple effects of these variables on Executive Function, unadjusted by other variables in the model.

Table 1

Demographic characteristics of full sample and MRI sub sample.

Ethnicity		N	Gender N (%) Male N (%) Female	Education (years) M (SD)	Age (years) M (SD)
African American	Full	351	118 (33.6) M 233 (66.4) F	13.1 (2.8)	72.5 (7.5)
	MRI	79	28 (35.4) M 51 (64.6) F	13.1 (3.2)	74.1 (7.1)
Hispanic	Full	410	139 (33.9) M 271 (66.1) F	6.5 (4.9)	70.4 (7.4)
	MRI	102	32 (31.3) M 70 (68.6) F	7.7 (5.6)	72.2 (7.4)
Caucasian	Full	458	187 (40.8) M 271 (59.2) F	14.3 (3.2)	73.8 (7.6)
	MRI	134	61 (45.5) M 73 (54.5) F	14.3 (3.5)	75.6 (7.4)
All	Full	1219	444 (36.4) M 775 (63.6) F	11.3 (5.1)	72.3 (7.6)
	MRI	315	121 (38.4) M 194 (61.6) F	11.9 (5.1)	74.1 (7.5)

Table 2

Mean (standard deviation) neuropsychological test scores and MRI variables within ethnic groups. Neuropsychological test scores are presented in standard score units with a mean of 0.0 and standard deviation of 1.0 in a demographically diverse normative sample. Brain Matter and Hippocampal volumes are expressed in cubic centimeter units and represent the difference between the observed brain volume components and the volume predicted by total intracranial volume in normals. White Matter Hyperintensity volumes are raw (not log transformed) values, and are also expressed in cubic centimeter units.

Variable	African American	Hispanic	Caucasian
Semantic Memory	0.03 (0.66)	-0.80 (0.88)	0.64 (0.74)
Episodic Memory	0.00 (0.80)	-0.55 (0.89)	-0.03 (1.04)
Executive Function	-0.15 (0.59)	-0.59 (0.69)	0.14 (0.71)
Spatial	-0.15 (0.73)	-0.68 (0.93)	0.31 (0.76)
Brain Matter Volume	-13.81 (55.07)	11.43 (48.37)	-32.30 (53.58)
Hippocampal Volume	-0.28 (0.58)	-0.11 (0.58)	-0.22 (0.67)
White Matter Hyperintensity Volume	8.72 (8.62)	6.59 (7.77)	11.35 (15.43)

Table 3

Variance in test scores accounted for by MRI variables within ethnic subgroups and in the combined sample. Tabled values are R^2 values from regressions of latent variables representing unadjusted and adjusted test score dependent variables on jointly entered MRI variables. The highest value in each row is bolded.

Test	Ethnic Group	Type of Adjustment		
		Age Adjusted	Unadjusted	Education Adjusted
Semantic Memory	African American	0.128	0.255	0.258
	Hispanic	0.000	0.039	0.083
	Caucasian	0.096	0.174	0.208
	Combined	0.027	0.020	0.065
Episodic Memory	African American	0.041	0.166	0.100
	Hispanic	0.128	0.269	0.357
	Caucasian	0.154	0.275	0.276
	Combined	0.063	0.140	0.218
Executive Function	African American	0.122	0.280	0.213
	Hispanic	0.124	0.330	0.460
	Caucasian	0.037	0.127	0.145
	Combined	0.007	0.048	0.155
Spatial	African American	0.321	0.460	0.454
	Hispanic	0.024	0.085	0.165
	Caucasian	0.041	0.115	0.121
	Combined	0.005	0.020	0.102

Table 4

Relationship of component MRI variables to specific test scores. Tabled results are standardized regression coefficients from analyses in which latent variables representing unadjusted and Age and Education adjusted test scores were dependent variables and MRI latent variables were joint independent variables. Results are from the combined sample. The highest value in each row is bolded.

Test	MRI Variable	Type of Adjustment		
		Age Adjusted	Unadjusted	Education Adjusted
Semantic Memory	bm	-0.132	-0.059	0.116
	hc	0.139	0.143	0.134
	wmh	0.030	-0.016	-0.112
Episodic Memory	bm	0.033	0.170**	0.271****
	hc	0.232***	0.235****	0.224***
	wmh	-0.025	-0.116	-0.165**
Executive Function	bm	0.000	0.120	0.285****
	hc	0.060	0.069	0.033
	wmh	-0.045	-0.123	-0.210***
Spatial	bm	-0.027	0.054	0.224***
	hc	0.046	0.053	0.013
	wmh	-0.041	-0.093	-0.188**

bm = brain matter volume, hc = hippocampus volume, wmh = white matter hyperintensity volume

* p< .05

** p<.01

*** p<.001

**** p<.0001

Table 5

Relationship of test scores with MRI latent variable within different ethnic groups. Tabled values are standardized coefficients from the final, best fitting multiple group model that identified and freely estimated effects of age, education, and MRI variables that differed across groups. wmh was more strongly associated with Spatial in African Americans (bold font); all other MRI effects on test scores were found to be equal across groups in the best fitting model.

		bm	hc	wmh
Semantic Memory	African American	0.14	0.14	-0.12
	Hispanic	0.10	0.12	-0.09
	Caucasian	0.12	0.14	-0.11
Episodic Memory	African American	0.29***	0.26****	-0.22**
	Hispanic	0.24***	0.25****	-0.18**
	Caucasian	0.22***	0.22****	-0.17**
Executive Function	African American	0.36****	0.01	-0.28***
	Hispanic	0.30****	0.01	-0.24***
	Caucasian	0.29****	0.01	-0.23***
Spatial	African American	0.31***	0.02	-0.57****
	Hispanic	0.24***	0.01	-0.10
	Caucasian	0.30***	0.02	-0.14

bm = brain matter volume, hc = hippocampus volume, wmh = white matter hyperintensity volume

* p< .05

** p<.01

*** p<.001

**** p<.0001

Table 6

Relationship of MRI Adjusted and Unadjusted test scores with age and education. MRI Adjusted values are standardized coefficients from the final, best fitting, multiple group model that identified and freely estimated effects of age, education, and MRI variables that differed across groups. Values in bold were freely estimated in the respective groups. MRI adjusted values reflect the correlations of age and education with test scores independent of the effects of MRI variables. Unadjusted values are simple correlations of age and education with test scores, and are unadjusted for MRI effects. Unadjusted correlations did not differ across groups.

		Age		Education	
		Unadjusted	MRI Adjusted	Unadjusted	MRI Adjusted
Semantic Memory	African American	-0.32 ****	-0.19 ***	0.33 ****	0.29 ****
	Hispanic	-0.24 ****	-0.14 ***	0.59 ****	0.58 ****
	Caucasian	-0.26 ****	-0.17 ***	0.36 ****	0.39 ****
Episodic Memory	African American	-0.34 ****	-0.11	0.30 ****	0.20 ****
	Hispanic	-0.35 ****	-0.10	0.29 ****	0.31 ****
	Caucasian	-0.31 ****	-0.09	0.12 **	0.17 ****
Executive Function	African American	-0.26 ****	-0.03	0.39 ****	0.32 ****
	Hispanic	-0.40 ****	-0.17 **	0.44 ****	0.46 ****
	Caucasian	-0.30 ****	-0.10	0.28 ****	0.29 ****
Spatial	African American	-0.17 ****	0.15 *	0.41 ****	0.37 ****
	Hispanic	-0.27 ****	-0.08	0.52 ****	0.49 ****
	Caucasian	-0.24 ****	-0.09	0.29 ****	0.37 ****

* p<.05

** p<.01

*** p<.001

**** p<.0001