



Published in final edited form as:

Neurobiol Aging. 2012 January ; 33(1): 83–95.e2. doi:10.1016/j.neurobiolaging.2010.01.021.

MRI predictors of cognitive change in a diverse and carefully characterized elderly population

Owen Carmichael¹, Dan Mungas¹, Laurel Beckett², Danielle Harvey², Sarah Tomaszewski Farias¹, Bruce Reed^{1,4}, John Olichney¹, Joshua Miller³, and Charles DeCarli¹

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

² Department of Public Health Sciences, School of Medicine, University of California, Davis

³ Department of Pathology, School of Medicine, University of California, Davis

⁴ Veterans Administration Northern California Health Care System

Abstract

Background—Trajectories of cognitive decline among elderly individuals are heterogeneous, and markers that have high reliability for predicting cognitive trajectories across a broad spectrum of the elderly population have yet to be identified.

Method—This study examined the utility of a variety of MRI-based brain measures, obtained at baseline, as predictors of subsequent declines in domain-specific measures of cognitive function in a cohort of 307 community-dwelling elderly individuals with varying degrees of cognitive impairment who were diverse across a number of relevant demographic variables and were evaluated yearly. Psychometrically matched measures of cognition were used to assess episodic memory, semantic memory, and executive function. Relationships between baseline MRI measures, including the volumes of the brain, hippocampus, and white matter hyperintensities (WMH), and cognitive trajectories were assessed in mixed effects regression models that modeled MRI effects on cognitive performance at baseline and rate of change as well as inter-individual variability in cognitive baseline and rate of change.

Results—Greater baseline brain volume predicted slower subsequent rate of decline in episodic memory and smaller WMH volume predicted slower subsequent rate of decline in executive function and semantic memory. Baseline hippocampal volume, while strongly related to baseline cognitive function, was not predictive of subsequent change in any of the cognitive domains.

Corresponding author: Owen Carmichael, Center For Neuroscience, 1544 Newton Court, Davis, CA, 95618, Telephone: 530-754-9657, Fax: 530-754-5036.

7 Disclosures

Owen Carmichael has no conflicts to report.

Dan Mungas has no conflicts to report.

Laurel Beckett has no conflicts to report.

Danielle Harvey has no conflicts to report.

Sarah Tomaszewski Farias has no conflicts to report.

Bruce Reed has no conflicts to report.

John Olichney has no conflicts to report.

Joshua Miller has no conflicts to report.

Charles DeCarli has no conflicts to report.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—Baseline measures of brain structure and tissue pathology predicted rate of cognitive decline in a diverse and carefully-characterized cohort, suggesting that they may provide summary measures of pre-existing neuropathological damage or the capacity of the brain to compensate for the impact of subsequent neuropathology on cognition. Conventional MRI measures may have utility for predicting cognitive outcomes in highly heterogeneous elderly populations.

2 Introduction

There are robust inter-individual differences in cognitive functioning throughout adulthood and this heterogeneity is compounded by differences in trajectories of cognitive change that emerge as people age. Longitudinal studies of older individuals reveal widely differing rates of cognitive decline, as well as many cases of stable function and even modest improvement (Albert, Jones et al. 1995) (Christensen, Mackinnon et al. 1999) (Colsher and Wallace 1991) (Rubin, Storandt et al. 1998) (Schaie 1988) (Zelinski, Gilewski et al. 1993; Wilson, Beckett et al. 2002). Variability in longitudinal trajectories translate into the increased variability of function over time that is one of the basic observations of cross sectional studies of cognitive aging (Christensen, Mackinnon et al. 1999). In the context of this broad heterogeneity, the prediction of late-life cognitive change has emerged as one of the central goals of cognitive aging research.

MRI is one biological measure of brain aging. Like cognition, MRI measures increase in variability with advancing age and inter-individual differences are associated with cognitive performance, even among the cognitively normal. This study focuses on the utility of structural neuroimaging-based measurements obtained at baseline to predict subsequent trajectories of cognitive function. The hypothesis underlying our approach, supported by convergent research from imaging, neuropathology, and epidemiology, is that much of the heterogeneity in late-life cognitive change is driven by associated heterogeneity in late-life neuropathological processes (Buckner 2004). Specifically, a variety of neurodegenerative processes may be incipiently present in the aging brain, even in supposedly cognitively normal individuals (Wilson, Beckett et al. 1999). These processes cause cell death on a macroscopic scale that is detectable through MRI. MRI measures, therefore, may provide indicators of prior neuronal loss attributable to these incipient neuropathological processes. In addition, regional brain atrophy may indicate that the brain lacks capacity to compensate for future pathological damage, leading to higher probability of poorer subsequent cognitive trajectories.

Prior MRI predictor studies have been limited in their ability to provide characterizations of gradual longitudinal cognitive change. In part, this is because in many studies, cognitive change is summarized in terms of longitudinal transitions between a limited number of clinical diagnostic categories representing normal cognitive function, mild cognitive impairment (MCI) and dementia (Chetelat and Baron 2003). However, numerous studies have suggested that neurodegenerative diseases of aging typically manifest themselves through gradual, progressive changes in brain structure and cognition, and that there is considerable variability in neuropathology and domain-specific cognitive function within each diagnostic group (Saito and Murayama 2007) (Saito and Murayama 2007) (Ganguli, Dodge et al. 2004) (Palmer, Wang et al. 2002; Manly, Tang et al. 2008) (Ritchie, Artero et al. 2001). Thus, boundaries for differentiating normal, MCI, and dementia are by nature arbitrary cut points in a continuum of pathology, and none of the three diagnoses strictly guarantees a particular pre-determined level of brain pathology or cognitive ability in a particular domain. This limits the interpretability and power of studies that relate MRI measures to subsequent clinical diagnosis.

Meanwhile, studies that used MRI to predict scores on dimensional cognitive test instruments have been limited by the measurement properties of those instruments (De Groot, De Leeuw et al. 2002). These instruments often fail to provide a linear relationship between cognitive ability and test score across a broad ability range (resulting in floor or ceiling effects), or do not provide comparable scores across multiple test languages. Additionally, neuropsychological batteries that test multiple cognitive domains (memory, executive function, *etc.*) have rarely demonstrated similar measurement properties across the different domains. Under these circumstances it is difficult to draw firm conclusions about the true nature of differential relationships between predictor variables and specific cognitive domains. Additionally, the generalizability of findings from prior MRI predictor studies has been limited by biases in cohort composition and frequency of cognitive evaluations. Many studies included a high proportion of individuals at risk for AD, including those with a family history of AD, subjective cognitive complaints, or clinically-diagnosed MCI; they also usually included a low proportion of minorities, individuals with low educational attainment, or individuals with high cardiovascular disease burden. This limits the applicability of the proposed MRI predictors across the broader elderly population. Meanwhile, prior epidemiological studies of aging that have assembled diverse, representative cohorts and collected a brain MRI generally have long intervals between cognitive evaluations that preclude acquisition of a sufficient number of measurements required for characterizing cognitive change (Carmichael, Kuller et al. 2007) (De Groot, De Leeuw et al. 2002). The long follow-up interval also precludes detection of cognitive changes that often occur between evaluations.

The aim of this study is to overcome these limitations by assessing MRI predictors of cognitive change in a highly diverse cohort of individuals whose cognitive function was assessed approximately yearly using sensitive, robust cognitive instruments (Mungas, Reed et al. 2004) (Mungas, Reed et al. 2005) (Mungas, Reed et al. 2000) (Mungas, Reed et al. 2005). To accomplish this, we recruited a study cohort with broad representation across three ethnic groups that is diverse in terms of socioeconomic status, education, cognitive function, cardiovascular health, early life experiences, and environmental exposures. The cognitive measures lack significant ceiling or floor effects, have linear measurement properties across a broad ability range, and provide comparable scores across speakers of Spanish and English. The diversity and careful characterization of this sample allows us to extend prior MRI predictor studies by determining whether a set of established MRI markers are clinically relevant for predicting decline in a broad subset of the elderly population.

3 Methods

3.A. Subjects

This study included 307 participants in an on-going longitudinal study of cognitive impairment in an educationally and ethnically diverse sample of older adults. These individuals were evaluated and followed within the research program of the University of California at Davis Alzheimer's Disease Center (UCD ADC). Participants were recruited into the study through two routes: 1) community outreach and 2) memory clinic referrals. Approximately 68% of participants were recruited through community based recruitment protocols designed to enhance the racial and ethnic diversity and the spectrum of cognitive dysfunction of the sample with an emphasis on normal cognition and MCI. Recruiters utilized various outreach methods such as soliciting in a community hospital lobby, a community survey, health fairs or word of mouth. The other 28% of the sample initially sought a clinical evaluation at the UCD ADC and subsequently were recruited for this study. These individuals predominantly had a clinical diagnosis of MCI. The overall sample included 99 African Americans, 79 Hispanics, and 129 Caucasians. Fifteen individuals from

other racial/ethnic groups, recruited from clinical evaluation, were removed due to the small size and heterogeneous nature of this group.

Regardless of recruitment source, inclusion criteria were age greater than 60 and ability to speak English or Spanish. Exclusion criteria 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. All participants signed informed consent, and all human subject involvement was overseen by institutional review boards at University of California at Davis, the Veterans Administration Northern California Health Care System and San Joaquin General Hospital in Stockton, California.

3.B. Cognitive assessment

The primary cognitive outcome measures in this study were from the Spanish and English Neuropsychological Assessment Scales (SENAS). The SENAS has undergone extensive development as a battery of cognitive tests relevant to diseases of aging (Mungas, Reed et al. 2004) (Mungas, Reed et al. 2005) (Mungas, Reed et al. 2000) (Mungas, Reed et al. 2005). Modern psychometric methods based on item response theory were used to create psychometrically matched measures across different scales and across English and Spanish versions. This study used a subset of SENAS tests to measure three cognitive domains: episodic memory, semantic memory, and executive function. The Episodic Memory measure was a composite score derived from a multi-trial word list learning test (Word List Learning 1, (Mungas, Reed et al. 2004)). Semantic Memory was a composite of highly correlated verbal (Object Naming) and nonverbal (Picture Association) tasks. Executive Function was a composite measure constructed from component tasks of Category Fluency, Phonemic (letter) Fluency, and Working Memory. Measure development and psychometric characteristics have been reported in previous publications (Mungas, Reed et al. 2004; Mungas, Reed et al. 2005; Crane, Narasimhalu et al. 2008). These measures do not have appreciable floor or ceiling effects for participants in this sample and have linear measurement properties across a broad ability range. They are near-normally distributed, which presents advantages for statistical analyses. Psychometrically matched measures of domain specific cognitive functions (i.e. measures with equivalent reliability and sensitivity) were used to facilitate unambiguous interpretation of any potential differential effects of the imaging variables on cognitive trajectories. The use of psychometrically matched measures in the current study allows us to draw more confident conclusions about the influence of brain imaging variables on domain-specific cognitive trajectories.

3.C. Imaging

Baseline MRI data was acquired on two 1.5T MRI scanners: a GE Signa machine located at UCD Medical Center (Sacramento, CA), and a Philips Eclipse machine located at the Veterans Administration Northern California Health Care System (Martinez, CA). High-resolution T1-weighted and fluid-attenuated inversion recovery (FLAIR) sequences required for measurement of MRI variables were acquired in each subject. The T1-weighted sequence was an axial-oblique 3D Fast Spoiled Gradient Recalled Echo (FSPGR) sequence with the following parameters: TE: 2.9 ms (min), TR: 9 ms (min), Flip angle: 15 deg, Slice thickness: 1.5 mm, Number of Slices: 128, FOV: 25 cm × 25 cm, Matrix: 256 × 56. The Axial-oblique 2D FLAIR Fast Spin Echo sequence had the following parameters: TE: 144 ms, TR: 11000 ms, TI: 2250 ms, Flip Angle: 90 deg, Slice thickness: 3 mm, FOV: 22 cm × 22 cm, Matrix: 256 (freq) × 92 (phase). Analogous sequences were installed on both the GE and Philips scanners.

After acquisition, all scans were digitally transmitted to the Imaging of Dementia and Aging laboratory (IDeA Lab) at UC Davis for analysis. Standard analyses of all MRIs included

measurement of total intracranial (TCV), cerebral brain (BV), hippocampal (HC) and WMH volumes. All images were analyzed by an operator blind to patient age, gender and diagnostic category to avoid introducing bias to the process. Quantification of BV, TCV, HC, and WMH were performed using the Quanta package of software routines, which was produced in-house.

TCV, BV, and WMH volumes were obtained from FLAIR according to a previously-reported analysis protocol (DeCarli, Fletcher et al. 2005). First, non-brain elements were manually removed from the image by operator guided tracing of the dura mater within the cranial vault including the middle cranial fossa, but excluding the posterior fossa and cerebellum. The volume of the traced region was defined as the TCV. Tissues outside the traced cranial vault were removed from the image. To identify brain matter, image intensity nonuniformities were removed from the image, and the corrected image was modeled as a mixture of two Gaussian probability functions corresponding to brain tissue and non-brain tissue; the segmentation threshold between brain and non-brain image intensities was located at the minimum probability between these two distributions (DeCarli, Maisog et al. 1992) (DeCarli, Murphy et al. 1996). The volume of voxels on the brain side of the threshold was defined as BV. Morphometric erosion of two exterior image pixels was then applied to the image of BV voxels to remove the effects of partially-volumed CSF pixels on WMH detection. A single gaussian distribution was then fitted to the intensity distribution of the remaining BV voxels, and all voxels with intensity greater than 3.5 standard deviations above the mean BV intensity were defined as WMH (DeCarli, Fletcher et al. 2005).

The HC was manually traced on T1-weighted scans to include the CA1 – CA4 fields, dentate gyrus, and the subicular complex using a protocol described previously (DeCarli, Reed et al. 2008). Briefly, all scans were resliced perpendicular to the long axis of the left HC and HC borders were manually traced on contiguous coronal slices in the anterior to posterior direction. The HP was bounded anteriorly by the amygdala, and tracing ended posteriorly at the first slice where the fornices were completely distinct from thalamic gray and white matter. The inferior boundary of the hippocampus was the white matter of the parahippocampal gyrus. The lateral boundary was the temporal horn of the lateral ventricle. The uncus was included in sections in which the uncus was ventral to caudal amygdala; the fimbria were excluded.

MRI-evident infarcts were identified by a trained neurologist through imaging characteristics and location using the Quanta software package. Quanta allows simultaneous views of T1-, PD-, and T2-weighted sequences, as well as a “pseudo-T1” image (PD subtracted from T2). Signal void, best seen on T2-weighted images, is interpreted to indicate a blood vessel. Only lesions >3 mm qualify for consideration as cerebral infarcts. Other necessary imaging characteristics include the following: 1) CSF density on the pseudo-T1 image and 2) distinct separation from the circle of Willis vessels if the infarct is in the basal ganglia region. The location and type of infarction (small vessel lacune or large vessel territory) was recorded for each infarct. Lacunar infarcts are defined as infarcts less than 1.5 cm in largest diameter and located in subcortical white matter or basal ganglia.

A rigorous protocol insured the validity of the MRI measures across differing scanners and analysts. Each analyst was required to produce ratings of training scans that agreed strongly with those of prior analysts as well as a neurologist skilled in neuroanatomy and neuroimaging; intra-class correlation coefficients (ICCs) between new analysts, all previously-trained analysts, and the neurologist were required to be above .95, .95, .95, and .9 for TCV, BV, WMH, and HC. A set of 7 cognitively-normal elderly individuals received one scan on the GE system and another scan on the Philips system within a 60-day interval; within-subject, between-scanner agreement in HC, TCV, and WMH was strong (ICCs = .

87, .96, and .89). In addition, 20 subjects received two scans each in a 2-week interval on the GE scanner, and BV and TCV were measured on all scans; within-subject agreement in the measures was strong (ICCs = .97 and .99).

BV was corrected for head size by linearly regressing BV against TCV, and replacing BV in the mixed effects modeling with the signed difference between the BV predicted based solely on TCV, and the true BV (Van Petten 2004). This residualized measurement represents the component of BV that is not already explained by TCV. HC was residualized in the same way. The residualized measurements had superior reliability to additive measurement noise than HC and BV divided by TCV (see Appendix). The residualized BV and HC measures were reasonably normally distributed on visual inspection, but because the distribution of WMH values was skewed we used log-transformed WMH in the mixed effects modeling.

3.D. Statistical modeling

We used mixed-effects repeated-measures regression models to assess relationships between baseline MRI measures and longitudinal changes over time in the episodic memory, semantic memory, and executive function measures (Pinheiro and Bates 2000). Analysis was restricted to individuals with cognitive measures at baseline and at least one follow-up time, and with MRI at baseline. The mean trajectories of cognitive measures over time were modeled as linear trends, with the predicted initial cognitive level (intercept) and the rate of change (slope) allowed to vary with baseline MRI measures and other covariates (fixed effects). In addition, our model allowed for people to have systematic differences, not accounted for by MRI or other predictors, in the starting level and rate of change of cognitive measures (random effects). Finally, the model assumed that the observed measures at an individual cognitive assessment time might differ from the person's general trajectory by a quantity drawn independently from a Gaussian distribution with zero mean and constant "noise" variance. Such mixed effects models are highly effective in assessing general trends over time and how they are modified by predictors, as well as characterizing between- and within-person variation from the general trend (Wilson, Beckett et al. 2002). Our estimation approach allowed for missing times of observation, unequal lengths and numbers of follow-ups, and possible correlation between the initial level of cognitive measurement for an individual and the rate of change over time. All models were fitted in using the *nlme* routines in R version 3.1 (Ihaka and Gentleman 1996). Model assumptions were validated graphically and by examining alternative models to test for violations such as non-linearity.

Each cognitive measure was modeled by a separate series of mixed-effects models in a sequential model-building approach. First, we built a reference model of the cognitive measure as a function of time, with random effects of subject on baseline and change rate, and fixed effects of ethnicity, years of educational attainment, and gender on baseline. Subject age was not entered as a fixed effect in these models due to earlier indications that age adjustment may attenuate relationships between MRI variables and cognitive outcomes in this cohort, a finding consistent with the hypothesis that age-related MRI change may represent a key explanation for decline in cognition in the elderly (Mungas, Reed et al. 2009). Ethnicity and education, on the other hand, were included as more direct ways to assess the influence of early childhood experiences and genetic traits on cognitive trajectory.

A set of "single-MRI" models was then constructed, each of which added fixed effects of a single MRI variable on baseline and change rate. Likelihood ratio tests assessed whether the explanatory power of each of the second models exceeded that of the reference model; for any MRI variable whose likelihood ratio was significant ($p < 0.05$), we used *F* tests on the single-MRI model marginal sums of squares to evaluate the significance of the fixed effects

of the MRI variable on baseline and change rate. Any fixed effect whose F test passed a liberal significance threshold ($p < 0.25$) was entered into a final “multiple-MRI” model to evaluate the independent fixed effects of all relevant MRI variables simultaneously. The incremental explanatory value added by the multiple-MRI model over the baseline model was evaluated by a likelihood ratio test. In addition, we used Cohen’s f^2 to quantify the effect size of adding the MRI variables to the reference model; in brief, this compares the variance explained (R^2) by the reference model fixed effects to the variance explained by the multiple-MRI model fixed effects. The significance of fixed effects in the multiple-MRI model was evaluated using F tests. In all models that included MRI variables, the duration of time between the baseline cognitive evaluation and the baseline MRI was entered as a fixed effect on cognitive baseline to guard against spurious effects caused by the time lag.

This study analyzed the subset of 307 individuals from the UCD ADC longitudinal cohort who received an MRI at baseline. We verified our assumption that the longitudinal cognitive characteristics of this sample reflect those of the full UCD ADC cohort of 369 individuals, which were published previously (Mungas, Beckett et al. 2009). In particular, we used the same mixed effects modeling approach as the prior analysis to relate baseline performance and rate of change in the three cognitive measures to baseline clinical diagnostic category (normal, MCI, and demented) and groups representing differing transitions in clinical diagnosis between first and final clinical evaluation (normal-to-normal, normal-to-MCI, etc.). We then qualitatively compared the model-estimated effects of clinical diagnosis on cognitive trajectories between the MRI subset and the full cohort.

Approximately 68% of the full UCD ADC cohort was recruited from the community, and the remainder were referred from memory clinics. A prior analysis suggested the clinic-recruited participants had a higher proportion of Caucasians, higher educational attainment, higher proportion of males, greater degrees of baseline functional impairment, and lower baseline BV and HC (Farias, Mungas et al. 2009). Therefore, we investigated whether recruitment source may have influenced the relationships between MRI variables, cognition, and demographic covariates that were ascertained in the MRI cohort. To do so, we re-estimated the same multiple-MRI models derived as described above on the subset of the MRI cohort that was recruited from the community, and qualitatively compared the model-derived relationships between MRI variables, cognitive measures, and demographic covariates that were derived from the full cohort, to those derived from the community-recruited subset.

4 Results

Cohort characteristics are listed in Table 1. The subject pool represents a range of educational attainment, substantial representation across all three ethnic groups, and broad heterogeneity in cognitive function measured through the dimensional measures as well as through clinical diagnosis. Brain MRI measures were also heterogeneous, with expected pathology-associated differences between individuals who were normal, MCI, and demented at baseline. Breakdowns of cognitive and MRI measures by ethnicity and diagnostic group are shown in Supplementary Tables 1 and 2. Example images showing the spectrum of raw BV, HC, and WMH values, before adjustment for cranial volume, are shown in Figure 3. Individuals were followed for an average of about 3 years, with evaluations occurring approximately annually. The average time lag between initial cognitive evaluation and MRI was approximately 3 months.

The longitudinal cognitive trajectories of this sample, as well as relationships between cognitive trajectories and clinical diagnosis, largely reflected those of the parent UCD ADC longitudinal cohort (Mungas, Beckett et al. 2009). In particular, all three baseline diagnostic

groups experienced statistically significant mean declines over time in episodic memory and executive function (Normal: $p < .001$ and $p = .049$; MCI: $p < .001$ and $p < .001$; Dementia: $p < .001$ and $p < .001$), but rate of change in semantic memory showed significant decline only among baseline MCI and baseline demented groups (Normal: $p = .48$; MCI: $p < .001$; Dementia: $p < .001$). The baseline normal group had significantly slower declines in executive function and semantic memory than the other groups ($p < .001$ for all comparisons), but other inter-group differences in decline rates were not significant (minimum p values for all other group comparisons in episodic, semantic, and executive were .29, .13, .62). In addition, robust differences between baseline diagnostic groups in baseline performance were evident across all cognitive domains (all p values for omnibus F tests were $< .001$). As in the prior analysis, the lack of differences in mean declines across baseline diagnostic groups appeared to be accounted for by the presence of subsequent decliners and subsequent non-decliners in the same baseline diagnostic group. Specifically, subjects transitioning from normal to MCI or MCI to dementia, along with subjects demented throughout followup, had significant declines in all cognitive measures (all $p < .001$), while subjects with a stable diagnosis or improvement from MCI to normal mostly exhibited change rates that were slightly positive or whose statistical significance was borderline (MCI to normal and MCI to MCI: $p = .34$, $p = .99$, $p = .60$ and $p = .06$, $p = .08$, $p < .001$ for episodic, semantic, executive; stable normal: $p < .001$ for a .04 s.d./year improvement in semantic, $p = .99$ for executive). However, in a departure from the prior analysis, episodic memory did decline significantly among subjects who were normal at baseline and throughout follow-up ($p < .001$). All comparisons between diagnostic groups indicated significantly greater rates of decline for groups with a clinical diagnosis of greater impairment (e.g., change per year in episodic, semantic, executive was smaller by .13, .11, and .12 s.d. per year in stable MCI compared to MCI declining to dementia; $p = .006$, $p = .001$, $p < .001$).

All single-MRI models added significant explanatory power over the baseline model according to likelihood ratio tests (all $p < .001$). However, in the single-MRI model for semantic memory, higher baseline WHM predicted greater decline but was not associated with baseline semantic memory using the $p < 0.25$ threshold ($p < .001$ and $p = .88$). Also, in the single-MRI model for semantic memory using BV as a predictor, lower baseline BV predicted worse baseline semantic memory but not change in semantic memory, at the $p < 0.25$ threshold ($p < .001$ and $p = .46$). All other MRI fixed effects in the single-MRI models passed the $p < 0.25$ threshold and therefore were included in the multiple-MRI models.

Multiple-MRI models for each cognitive outcome provided significant explanatory power over the corresponding reference models according to likelihood ratio tests (all $p < 0.001$). Adding the MRI variables had a moderate effect on variance explained by the models, with Cohen's f^2 values of 0.21, 0.18, and 0.12 for episodic memory, executive function, and semantic memory. The significance of fixed effects in the multiple-MRI models is summarized in Table 2. Greater BV at baseline was associated with greater baseline performance in all three domains as well as slower rate of subsequent decline in episodic memory. Greater HC at baseline was associated with greater episodic and semantic memory performance at baseline, but its associations with subsequent rate of change were not significant. Greater WMH at baseline was associated with poorer baseline performance in episodic memory and executive function, as well as faster subsequent declines in executive function and semantic memory. Each graph in Figures 1–3 illustrates how a typical cognitive trajectory is modified by varying the values of a single MRI variable of interest. Each line represents a model-predicted cognitive trajectory for a Hispanic female with mean educational attainment (12.6 years) and mean values of all other MRI variables. The three lines per graph represent predicted trajectories for subjects whose MRI variable of interest takes on the mean value and mean plus or minus 1 standard deviation. In addition, Figure 4 shows model-predicted episodic memory trajectories along with episodic memory

measurements for 81 randomly-selected subjects. The Figure confirms that use of linear trajectory models was appropriate for this data.

Because MRI-evident infarction is strongly associated with WMH and cognitive performance, we investigated whether the WMH-cognition relationships may have been driven by the presence of infarcts at baseline. To do so we re-estimated all multiple-MRI models including an additional binary variable indicating whether or not the individual had one or more infarcts evident on the baseline MRI. This variable did not change the results in any appreciable way. Specifically, single-MRI models for the infarct variable did not provide significantly improved goodness-of-fit over reference models for episodic and semantic memory by likelihood ratio tests ($p=.27, .29$). The likelihood ratio test for the single-MRI model for executive function was significant ($p=.035$); by F tests, baseline infarcts were significantly associated with baseline executive function but not longitudinal change ($p=.01, .90$). In the multiple-MRI model, the effect of infarcts on baseline executive function was not significant ($p=.17$), greater BV and WMH were associated with greater baseline executive function ($p<.001, p=.03$), and WMH was associated with slower subsequent declines in executive function ($p=.026$).

In addition, because MRI variables differed across baseline clinical diagnosis groups, we sought to assess whether the MRI variables acted purely as proxy measures of clinical diagnosis. To do so we re-estimated our models with baseline diagnosis as an additional predictor of baseline cognition and longitudinal change. While this addition attenuated the relationships between MRI variables and cognition, it did not substantially modify the direction or strength of most associations. Specifically, in the multiple-MRI model, greater HC, and smaller WMH, were still associated with baseline episodic memory, and greater BV was still associated with greater subsequent declines in episodic memory ($p=.08, .09, .007$). Greater baseline BV and smaller baseline WMH were still associated with baseline executive function, and greater baseline WMH was still associated with faster subsequent decline in executive function ($p<.001, p=.09, p=.06$). Greater baseline BV was still associated with greater baseline semantic memory, and greater baseline WMH was still associated with greater subsequent declines in semantic memory ($p=.04, .009$).

The pattern of MRI-cognition relationships was largely unchanged when the analysis was run on the subset of individuals who were recruited from community sources, although the reduction in number of subjects attenuated the strengths of the relationships. Specifically, greater baseline BV and HC were associated with greater baseline semantic memory, ($\beta=.17, .08; p=.006, .24$). Greater baseline WMH was associated with faster subsequent decreases in semantic memory ($\beta=-.04; p<.001$). Greater BV and HC, and smaller WMH, were associated with greater baseline episodic memory, and greater BV was associated with smaller subsequent decreases in episodic memory ($\beta=.17, .20, -.07, .03; p=.004, p<.001, p=.20, p=.09$). Greater BV and smaller WMH were associated with greater baseline executive function, and smaller WMH was associated with slower subsequent declines in executive function ($\beta=.16, -.04, -.03; p<.001, p=.34, p=.008$). This suggests that the MRI-cognition relationships are not strongly influenced by inter-individual differences in recruitment source.

Ethnicity, gender, and education significantly impacted baseline cognitive performance, and ethnicity was associated with rates of cognitive change. Specifically, baseline performance in all three domains was poorer among Hispanic subjects, and baseline executive function and semantic memory were greater among whites than both other ethnic groups ($p<.001, p=.002, p<.001$). Baseline episodic memory was poorer among males ($p<.001$), and greater educational attainment was associated with greater baseline performance in all three domains ($p=.001, p<.001, p<.001$ for episodic, executive, semantic). Rates of decline in

executive function and semantic memory were faster among whites, and rate of decline in episodic memory was slower among Hispanics ($p=.008$, $p=.03$, $p=.01$). Educational attainment was not significantly associated with rate of change in any domain ($p=.73$, $p=.92$, $p=.90$ for episodic, executive, semantic). In addition, when interaction effects of education and MRI variables on cognitive change were entered into the single-MRI and multiple-MRI models described above, none were significant, suggesting that higher-order education-MRI-cognition relationships did not account for the lack of effect of education on cognitive change (minimum $p=.36$). This general pattern of results did not change appreciably when the data set was reduced to include only the community-recruited subjects.

The inter-individual variability in cognitive baselines and change rates captured by random effects parameters was substantial. The standard deviations in baseline episodic memory, executive function, and semantic memory z-scores associated with inter-individual variability were .69, .52, and .63, representing 56, 57, and 63 percent of the cognitive score variance unexplained by the fixed effects. Standard deviations in change rate in the three domains were .10, .11, and .057 z-score units per year, representing 8, 11, and 6 percent of the cognitive score variance unexplained by the fixed effects.

5 Discussion

The key finding of this study is that in a broadly heterogeneous and carefully characterized elderly cohort, multiple MRI-based measures of brain structure were predictive of future performance across three highly informative measures of cognitive function. In particular, individuals with more brain tissue (BV) and less injured white matter (WMH) at baseline went on to have better subsequent performance in episodic memory, semantic memory, and executive function. These results did not appear to be driven by infarcts or recruitment source. The results suggest that brain MRI could provide useful information for prognosis of late-life cognitive outcomes across a broad spectrum of elderly individuals.

While hippocampal volume was strongly associated with baseline memory, it did not predict future cognitive performance. We feel that this is largely explained by the uniquely high level of heterogeneity in this cohort. That is, our cohort includes a subset of individuals whose relatively small baseline HC likely reflects the early effects of progressive neuropathology such as AD; by providing a marker of pre-existing neuropathological damage that is likely to worsen over time, diminished HC predicts future cognitive performance in these individuals. However, our cohort also includes individuals whose relatively small baseline HC likely reflects a wide variety of other influences, including genetics, early life experiences and environmental exposures, and overall physical health; in these individuals HC may in fact remain stable over time (Petit, Alfano et al. 1983; Dauncey and Bicknell 1999) (Sullivan, Pfefferbaum et al. 2001). In both groups, HC would be associated with concurrent cognitive function, as our associations between baseline HC and baseline cognition suggest; however, the diverse reasons for small baseline HC undercut its ability to predict future cognition. In contrast, prior studies that showed the predictive power of baseline HC generally included a high proportion of participants who were more likely to already exhibit early AD pathology at baseline, and for whom reduced HC would therefore more likely predict future hippocampal damage.

A similar argument would appear to apply to BV—specifically, that in our diverse cohort, small baseline BV may or may not have been caused by progressive neuropathology, and therefore may or may not be associated with subsequent rate of decline. However, we feel there are three possible reasons why brain volume was predictive. The first reason is that loss of BV may be the final common pathway of a wide variety of pathological processes, which are present to varying degrees across the cognitive spectrum of our cohort due to the

high level of cohort diversity (Wilson, Beckett et al. 1999). Because BV reflects the accumulated effects of all such processes, even among cognitively-normal participants, BV in this cohort possesses great variability that is primarily driven by progressive pathology. This would make BV a more specific measure of progressive pathological damage than HC and therefore a better predictor of cognitive change. The second reason is that brain volume may reflect the sheer bulk of brain tissue that is available for cognitive systems to recruit to compensate for future pathological damage, leading to improved outcomes on average for individuals with higher BV (Dickerson, Salat et al. 2005; Dai, Lopez et al. 2009) (Stern, Moeller et al. 2000) (Gould, Arroyo et al. 2006) (Cabeza, Anderson et al. 2002). The third reason is that because the hippocampus is complex, small, and partially bounded by other gray matter structures, HC may have greater measurement noise than BV or WMH, thus reducing the strength of its associations with cognitive change.

While prediction of future cognitive performance was the primary scientific goal of this study, cross-sectional relationships between MRI variables and cognitive performance validate our methodology by largely agreeing with a substantial body of prior cross-sectional imaging studies. Specifically, greater BV was associated with greater performance across all 3 domains, greater HC was associated with greater memory performance, and greater WMH was associated with poorer episodic memory and executive function.

The key strengths of this study are its heterogeneous, community-dwelling participants and careful, frequent measurement of cognition. The state-of-the-art cognitive instruments allowed sensitive and robust measurement of distinct aspects of longitudinal cognitive function across multiple languages and across a broad ability range; this is what enabled us to assess MRI predictors of cognition in a population that included significant numbers of Spanish-speaking and English-speaking participants and a broad range of baseline cognition and cognitive change. The diversity in cognitive characteristics across the sample reflects our intentional recruitment of individuals that were diverse in terms of socioeconomic status, cardiovascular health, and other factors that impact cognitive and brain health. Acquiring cognitive assessments frequently-- approximately yearly-- helped us to guard against interval censoring issues. These unique features of the study support the generalizability of our findings to a relatively broad subset of the elderly population at large.

The key limitation of this study was the lack of multiple MRI measurements per subject. As mentioned above, heterogeneity in neuropathology in this sample likely lead to heterogeneity in brain structure, which lead to high heterogeneity in cognitive trajectories. Imaging the brain at a single time point does not allow us to characterize longitudinal trajectories of brain structure, and therefore limits our ability to characterize longitudinal cognition. In particular, we were unable to detect individuals whose MRI-based brain measures were in the midst of deteriorating, and therefore would presumably exhibit poorer cognitive outcomes. This limitation of single-MRI studies has been exhibited repeatedly in the AD literature, where serial MRI measurements are superior to a single MRI for detecting and characterizing the disease (Jack, Shiung et al. 2005; Schott, Price et al. 2005). However, global MRI measures at a single time point did have utility for predicting future performance. Therefore, future studies should determine the value added by multiple MRIs over a single MRI for predicting subsequent cognitive trajectories in this heterogeneous sample.

Another limitation is our modeling of cognitive change as a linear trend over time. We used psychometrically-matched composite scales designed to have linear measurement characteristics that may have encouraged linearity in cognitive trajectories over time, and graphical plots such as Figure 4 suggested that linear trajectories provided adequate fits to our data. However, because many of the subjects had only two or three cognitive

evaluations, we likely lacked adequate statistical power to formally assess whether there were systematic trends toward non-linear cognitive changes at the population level. While the majority of prior work in this area involves linear models of cognitive change, recent evidence suggests that non-linear models may provide a more accurate characterization of individual cognitive trajectories (Johnson, Storandt et al. 2009).

Ethnicity effects in this study merit special consideration. The sample in this study was essentially a sample of convenience, and different ethnic groups were not equated in terms of amount of underlying brain pathology and cognitive trajectories. Caucasians on average had greater brain pathology as indicated by smaller brain and hippocampal volumes and faster cognitive decline. We controlled for ethnicity in analyses, which helps to limit confounding effects that might obscure relationships between MRI and cognition (Mungas, Reed et al. 2009). In addition, we were limited in our ability to assess the role of ethnicity as an independent predictor of cognitive trajectories because the clinic-recruited and community-recruited components of our sample differed significantly with respect to both ethnic makeup and MRI measures (Farias, Mungas et al. 2009). The consistency of findings between the full sample and the reduced sample of community recruits supports the validity of our conclusions, but future work should verify findings related to ethnicity on a larger cohort that lacks possible recruitment biases. Another important question that this study did not address is whether ethnicity modifies the relationship between MRI and cognition, that is, if the MRI-cognition relationships differ across ethnic groups. This would involve interaction effects and the sample size, while relatively large for a study comparing MRI and cognition, is not large enough to provide adequate statistical power to identify across-group differences in associations of MRI and cognition. Finding no significant difference would be especially problematic, because true lack of difference could not be distinguished from lack of statistical power to detect difference.

This study is part of an effort to understand the myriad biological processes that encourage or discourage cognitive decline in a diverse elderly population, and use that understanding to develop tools for diagnosis, prognosis, and treatment. The role of this study in that effort is two-fold: first, to suggest that methods that reduce brain atrophy or the formation of WMH could promote cognitive health, and second, to suggest that MRI measures of these brain properties could have prognostic value across a very heterogeneous sample. Future work should build on these findings by evaluating the value added to prognosis by measures that have greater explanatory value than our MRI measures, are easier to obtain, or provide complementary views of brain structure and function. In particular, recently-developed MRI measures of the spatial arrangement, shape, connectivity, and function of localized brain regions may assess the integrity of specific cognitive systems more precisely than BV and WMH, and therefore may have greater prognostic value for specific cognitive outcomes (DeCarli, Fletcher et al. 2005; Xie, Alcantara et al. 2009) (Greicius, Srivastava et al. 2004; Thompson, Hayashi et al. 2004). As mentioned above, multiple MRI per individual could provide measures of brain structural trajectories that help to indicate progressive neuropathological damage. A variety of molecular markers derived from blood or cerebrospinal fluid are potentially cheaper and easier to obtain than brain MRI, and provide microscopic measures of neuropathological activity that complement the macroscopic view provided by MRI; these markers may modify the brain-cognition relationships established here and therefore could add value for prognosis. Finally, given the genetic diversity present in a heterogeneous cohort, the contribution of genetic risk factors as determinants of brain structure and modifiers of brain-cognition relationships should be explored as well. Assessing the contribution of these additional measures to prognosis involves adding them to the foundational statistical models presented in this study, and formally testing whether they provide additional explanatory value.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute on Aging Grants AG10220, AG10129, AG 030514, AG031252, and AG021028 and by the California Department of Public Health Alzheimer's Disease Program Contracts 06-55311 and 06-55312. Dr. Miller receives research support from the NIH (R01CA120523-01A1 (Co-I), #1R21CA116409-01A2 (PI), #1R01HL083276-01A2 (Co-I), and #2R44DK077563-02 (Co-I)), the US Department of Defense [W81XWH-07-1-0650 (Co-PI)], and the American Cancer Society

8 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]
- Buckner RL. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. *Neuron*. 2004; 44(1):195–208. [PubMed: 15450170]
- Cabeza R, Anderson ND, Locantore JK, McIntosh AR. Aging gracefully: compensatory brain activity in high-performing older adults. *Neuroimage*. 2002; 17(3):1394–402. [PubMed: 12414279]
- Carmichael OT, Kuller LH, Lopez OL, Thompson PM, Dutton RA, Lu A, Lee SE, Lee JY, Aizenstein HJ, Meltzer CC, Liu Y, Toga AW, Becker JT. Ventricular volume and dementia progression in the Cardiovascular Health Study. *Neurobiol Aging*. 2007; 28(3):389–97. [PubMed: 16504345]
- Chetelat G, Baron JC. Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. *Neuroimage*. 2003; 18(2):525–41. [PubMed: 12595205]
- Christensen H, Mackinnon AJ, Korten AE, Jorm AF, Henderson AS, Jacomb P, Rodgers B. An analysis of diversity in the cognitive performance of elderly community dwellers: individual differences in change scores as a function of age. *Psychol Aging*. 1999; 14(3):365–79. [PubMed: 10509693]
- Colsher PL, Wallace RB. Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Ann Epidemiol*. 1991; 1(3):215–30. [PubMed: 1669503]
- Crane PK, Narasimhalu K, Gibbons LE, Pedraza O, Mehta KM, Tang Y, Manly JJ, Reed BR, Mungas DM. Composite scores for executive function items: demographic heterogeneity and relationships with quantitative magnetic resonance imaging. *J Int Neuropsychol Soc*. 2008; 14(5):746–59. [PubMed: 18764970]
- Dai W, Lopez OL, Carmichael OT, Becker JT, Kuller LH, Gach HM. Mild cognitive impairment and alzheimer disease: patterns of altered cerebral blood flow at MR imaging. *Radiology*. 2009; 250(3):856–66. [PubMed: 19164119]
- Dauncey MJ, Bicknell RJ. Nutrition and neurodevelopment: mechanisms of developmental dysfunction and disease in later life. *Nutr Res Rev*. 1999; 12(2):231–53. [PubMed: 19087453]
- De Groot JC, De Leeuw FE, Oudkerk M, Van Gijn J, Hofman A, Jolles J, Breteler MM. Periventricular cerebral white matter lesions predict rate of cognitive decline. *Ann Neurol*. 2002; 52(3):335–41. [PubMed: 12205646]
- DeCarli C, Fletcher E, Ramey V, Harvey D, Jagust WJ. Anatomical mapping of white matter hyperintensities (WMH): exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. *Stroke*. 2005; 36(1):50–5. [PubMed: 15576652]
- 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):274–84. [PubMed: 1545026]
- 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):519–28. [PubMed: 8724419]

- DeCarli C, Reed BR, Jagust W, Martinez O, Ortega M, Mungas D. Brain behavior relationships among African Americans, whites, and Hispanics. *Alzheimer Dis Assoc Disord*. 2008; 22(4):382–91. [PubMed: 19068502]
- Desgranges B, Baron JC, Eustache F. The functional neuroanatomy of episodic memory: the role of the frontal lobes, the hippocampal formation, and other areas. *Neuroimage*. 1998; 8(2):198–213. [PubMed: 9740762]
- Dickerson BC, Salat DH, Greve DN, Chua EF, Rand-Giovannetti E, Rentz DM, Bertram L, Mullin K, Tanzi RE, Blacker D, Albert MS, Sperling RA. Increased hippocampal activation in mild cognitive impairment compared to normal aging and AD. *Neurology*. 2005; 65(3):404–11. [PubMed: 16087905]
- Farias S, Mungas D, Reed B, Harvey D, DeCarli C. Progression of Mild Cognitive Impairment to Dementia in Clinic versus Community-Based Cohorts. *Archives of Neurology*. 2009 In Press.
- Ganguli M, Dodge HH, Shen C, DeKosky ST. Mild cognitive impairment, amnesic type: an epidemiologic study. *Neurology*. 2004; 63(1):115–21. [PubMed: 15249620]
- Gould RL, Arroyo B, Brown RG, Owen AM, Bullmore ET, Howard RJ. Brain mechanisms of successful compensation during learning in Alzheimer disease. *Neurology*. 2006; 67(6):1011–7. [PubMed: 17000970]
- Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc Natl Acad Sci USA*. 2004; 101(13):4637–42. [PubMed: 15070770]
- Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*. 1996; 5(3):299–314.
- Jack CR Jr, Shiung MM, Weigand SD, O'Brien PC, Gunter JL, Boeve BF, Knopman DS, Smith GE, Ivnik RJ, Tangalos EG, Petersen RC. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic MCI. *Neurology*. 2005; 65(8):1227–31. [PubMed: 16247049]
- Johnson DK, Storandt M, Morris JC, Galvin JE. Longitudinal study of the transition from healthy aging to Alzheimer disease. *Arch Neurol*. 2009; 66(10):1254–9. [PubMed: 19822781]
- Manly JJ, Tang MX, Schupf N, Stern Y, Vonsattel JP, Mayeux R. Frequency and course of mild cognitive impairment in a multiethnic community. *Ann Neurol*. 2008; 63(4):494–506. [PubMed: 18300306]
- Mungas D, Beckett L, Harvey D, Farias S, Reed B, Carmichael O, Olichney J, Miller J, DeCarli C. Heterogeneity of Cognitive Trajectories in Diverse Older Persons: What Does Diagnosis Tell Us? *Psychology And Aging*. 2009 Provisionally accepted.
- Mungas D, Reed BR, Crane PK, Haan MN, Gonzalez H. Spanish and English Neuropsychological Assessment Scales (SENAS): further development and psychometric characteristics. *Psychol Assess*. 2004; 16(4):347–59. [PubMed: 15584794]
- Mungas D, Reed BR, Farias ST, Decarli C. Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons. *Psychol Aging*. 2009; 24(1): 116–28. [PubMed: 19290743]
- Mungas D, Reed BR, Haan MN, Gonzalez H. Spanish and English neuropsychological assessment scales: relationship to demographics, language, cognition, and independent function. *Neuropsychology*. 2005; 19(4):466–75. [PubMed: 16060821]
- Mungas D, Reed BR, Marshall SC, Gonzalez HM. Development of psychometrically matched English and Spanish language neuropsychological tests for older persons. *Neuropsychology*. 2000; 14(2): 209–23. [PubMed: 10791861]
- 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. *J Int Neuropsychol Soc*. 2005; 11(5):620–30. [PubMed: 16212690]
- Palmer K, Wang HX, Backman L, Winblad B, Fratiglioni L. Differential evolution of cognitive impairment in nondemented older persons: results from the Kungsholmen Project. *Am J Psychiatry*. 2002; 159(3):436–42. [PubMed: 11870008]
- Petit TL, Alfano DP, LeBoutillier JC. Early lead exposure and the hippocampus: a review and recent advances. *Neurotoxicology*. 1983; 4(1):79–94. [PubMed: 6348600]

- Pinheiro, J.; Bates, D. *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer-Verlag; 2000.
- Ritchie K, Artero S, Touchon J. Classification criteria for mild cognitive impairment: a population-based validation study. *Neurology*. 2001; 56(1):37–42. [PubMed: 11148233]
- Rubin EH, Storandt M, Miller JP, Kinscherf DA, Grant EA, Morris JC, Berg L. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Arch Neurol*. 1998; 55(3): 395–401. [PubMed: 9520014]
- Saito Y, Murayama S. Neuropathology of mild cognitive impairment. *Neuropathology*. 2007; 27(6): 578–84. [PubMed: 18021380]
- Schaie KW. Variability in cognitive function in the elderly: implications for societal participation. *Basic Life Sci*. 1988; 43:191–211. [PubMed: 3284520]
- Schott JM, Price SL, Frost C, Whitwell JL, Rossor MN, Fox NC. Measuring atrophy in Alzheimer disease: a serial MRI study over 6 and 12 months. *Neurology*. 2005; 65(1):119–24. [PubMed: 16009896]
- Stern Y, Moeller JR, Anderson KE, Luber B, Zubin NR, DiMauro AA, Park A, Campbell CE, Marder K, Bell K, Van Heertum R, Sackeim HA. Different brain networks mediate task performance in normal aging and AD: defining compensation. *Neurology*. 2000; 55(9):1291–7. [PubMed: 11087770]
- Sullivan EV, Pfefferbaum A, Swan GE, Carmelli D. Heritability of hippocampal size in elderly twin men: equivalent influence from genes and environment. *Hippocampus*. 2001; 11(6):754–62. [PubMed: 11811670]
- Thompson PM, Hayashi KM, De Zubicaray GI, Janke AL, Rose SE, Semple J, Hong MS, Herman DH, Gravano D, Doddrell DM, Toga AW. Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage*. 2004; 22(4):1754–66. [PubMed: 15275931]
- Van Petten C. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. *Neuropsychologia*. 2004; 42(10):1394–413. [PubMed: 15193947]
- Wilson RS, Beckett LA, Barnes LL, Schneider JA, Bach J, Evans DA, Bennett DA. Individual differences in rates of change in cognitive abilities of older persons. *Psychol Aging*. 2002; 17(2): 179–93. [PubMed: 12061405]
- Wilson RS, Beckett LA, Bennett DA, Albert MS, Evans DA. Change in cognitive function in older persons from a community population: relation to age and Alzheimer disease. *Arch Neurol*. 1999; 56(10):1274–9. [PubMed: 10520945]
- Xie J, Alcantara D, Amenta N, Fletcher E, Martinez O, Persianinova M, DeCarli C, Carmichael O. Spatially localized hippocampal shape analysis in late-life cognitive decline. *Hippocampus*. 2009; 19(6):526–32. [PubMed: 19437501]
- Zelinski EM, Gilewski MJ, Schaie KW. Individual differences in cross-sectional and 3-year longitudinal memory performance across the adult life span. *Psychol Aging*. 1993; 8(2):176–86. [PubMed: 8323722]

1 Appendix: Reliability of residualized and normalized MRI variables

Our goal was to estimate the independent effects of BV, HC, and WMH on cognitive variables, but BV and HC were both strongly correlated with TCv. We considered two alternatives for accounting for this correlation in the statistical modeling: normalizing, in which BV and HC were divided by TCv; and residualizing, in which BV and HC were related to TCv in linear regression models, and BV and HC were replaced by the difference between their true values and the values predicted by linear regression based solely on TCv. Both approaches result in a derived MRI measurement (*e.g.*, normalized BV or residualized BV) based on a combination of two raw MRI measurements (BV and TCv). Each raw MRI measurement represents a true, underlying MRI quantity that has been corrupted by some level of measurement noise. We used a simulation study to assess the robustness of the derived measurements by determining how differing levels of additive noise in the raw measurements modified the reliability of the derived MRI measurements. To do so, we used

the MRI measures from this study as “true” scores and then added randomly-generated, normally-distributed artificial measurement noise to these scores to correspond to reliabilities of the raw measures of .95, .90, and .75. We then calculated normalized and residualized measures of BV and HC, and repeated this process over 100 independent simulations for each reliability level. “True” normalized and residualized measures were calculated and were correlated with derived measures with simulated measurement error. These correlation coefficients were used to estimate reliability of the derived BV and HC measures. The reliability of residualized measures was modestly but consistently higher than that of normalized measures. For BV, average reliability across simulations of residualized versus normalized was 0.84 as compared to 0.82 when raw score reliability was 0.95, it was 0.63 versus 0.59 when raw score reliability was 0.90, and it was 0.48 versus 0.43 when raw score reliability was 0.75. For HC, corresponding values were both 0.96 at reliability of 0.95, 0.87 versus 0.86 at reliability of 0.90, and 0.77 versus 0.74 at reliability of 0.75. For low levels of added noise, the difference in reliability between corresponding residualized and normalized measures was not great, but this difference increased with increasing levels of added noise. BV was more affected than HC because BV is much more highly correlated with ICV than is HC (0.86 versus 0.35). Differences between normalized and residualized measures were not substantial at realistic reliability levels, but normalized measures also are potentially problematic due to induced negative correlation between ICV and the derived normalized measure (Van Petten 2004). This problem does not apply to the residualized measures.

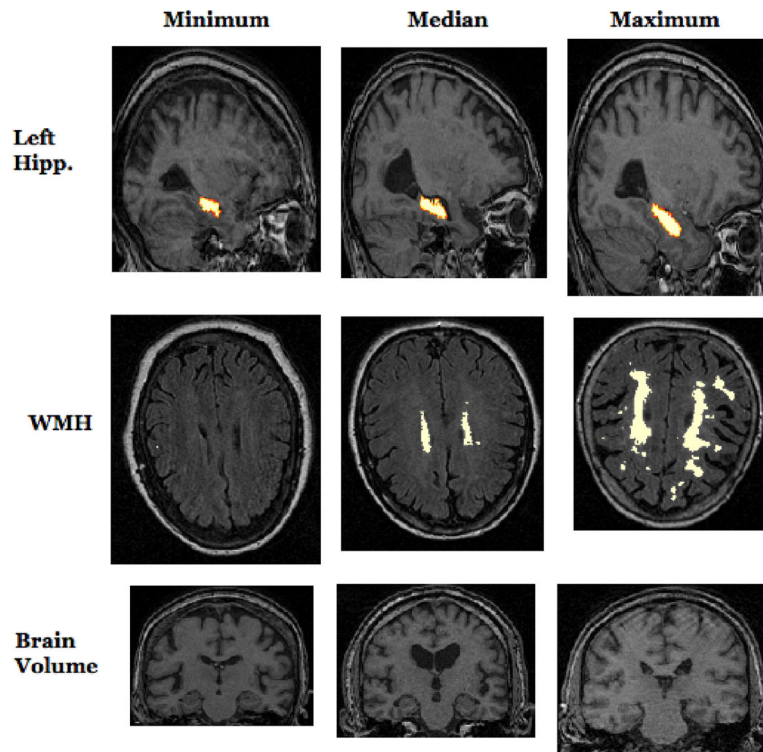


Figure 1. Example longitudinal trajectory of episodic memory, and the effect of brain volume in modifying that trajectory. Each line represents the estimated cognitive trajectory for a Hispanic female with mean educational attainment (12.6 years), mean WMH (9.10 cm³), and mean HC (3.62 cm³). The three lines per plot represent the estimated trajectories of an individual that exhibits these characteristics along with differing levels of BV: the population mean, the mean plus one standard deviation, and the mean minus one standard deviation. The episodic memory score is scaled as a z-score; the mean value over the entire population is 0 and a one-unit difference represents a difference of one standard deviation.

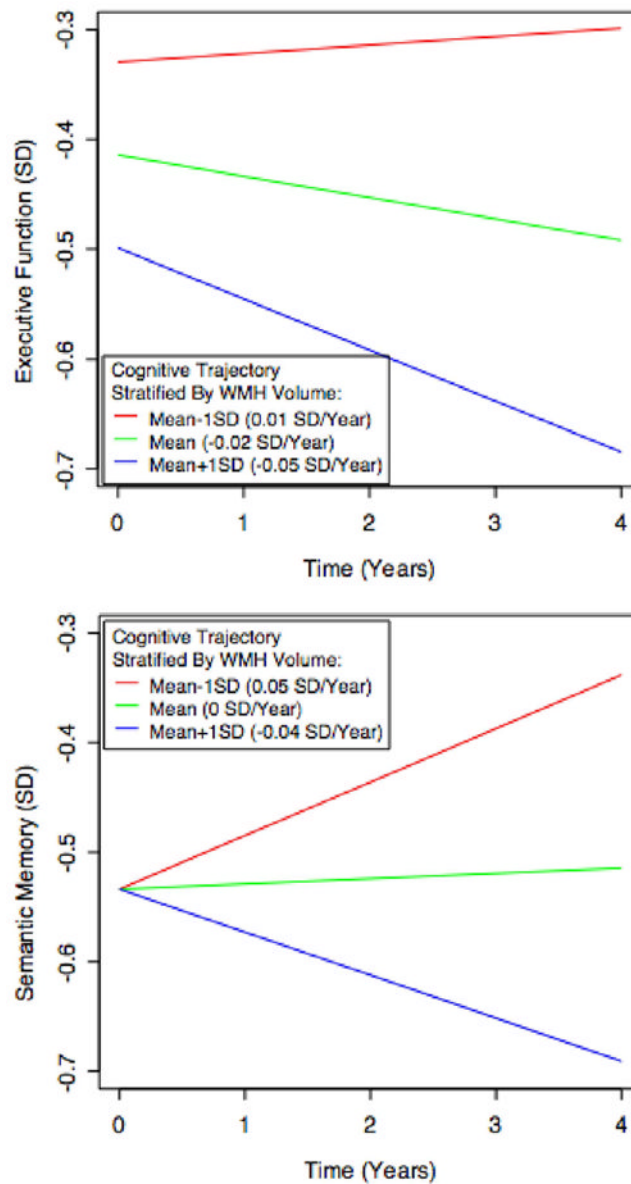


Figure 2.

Example longitudinal trajectories of executive function and semantic memory, and the effect of WMH volume in modifying those trajectories. Each line represents the estimated cognitive trajectory for a hispanic female with mean educational attainment (12.6 years), mean BV (919.1 cm³), and mean HC (3.62 cm³). The three lines per plot represent the estimated trajectories of an individual that exhibits these characteristics along with differing levels of BV: the population mean, the mean plus one standard deviation, and the mean minus one standard deviation. The executive function and semantic memory scores are scaled as z-scores; the mean of these measures over the entire population is 0 and a one-unit difference represents a difference of one standard deviation.

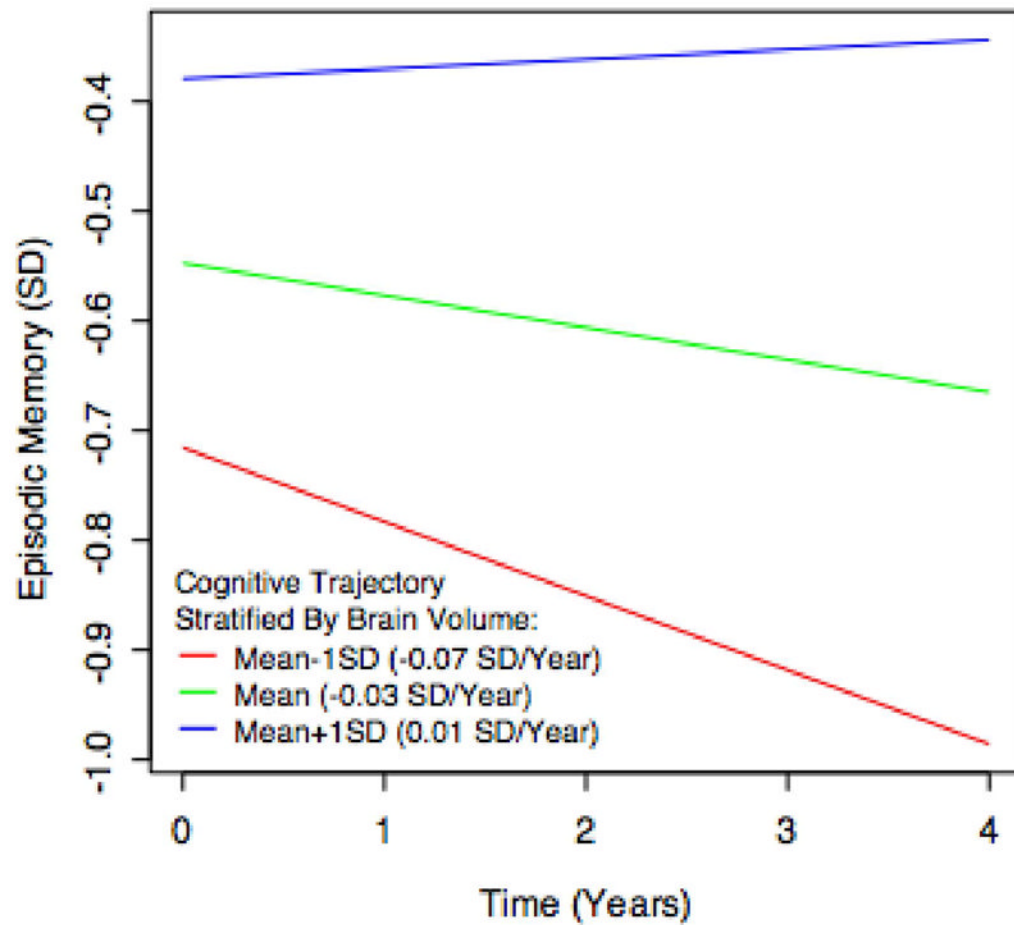


Figure 3.

Characteristic slices from MRI scans of individuals who had the minimum, median, and maximum values of left hippocampal volume, WMH volume, and brain volume. The left hippocampus is annotated by a red-to-white overlay on sagittal slices of T1-weighted images, and WMHs are annotated by a white overlay on axial slices of FLAIR images. Note that in the statistical analysis, brain and hippocampal volume were adjusted to account for inter-individual differences in intracranial volume.

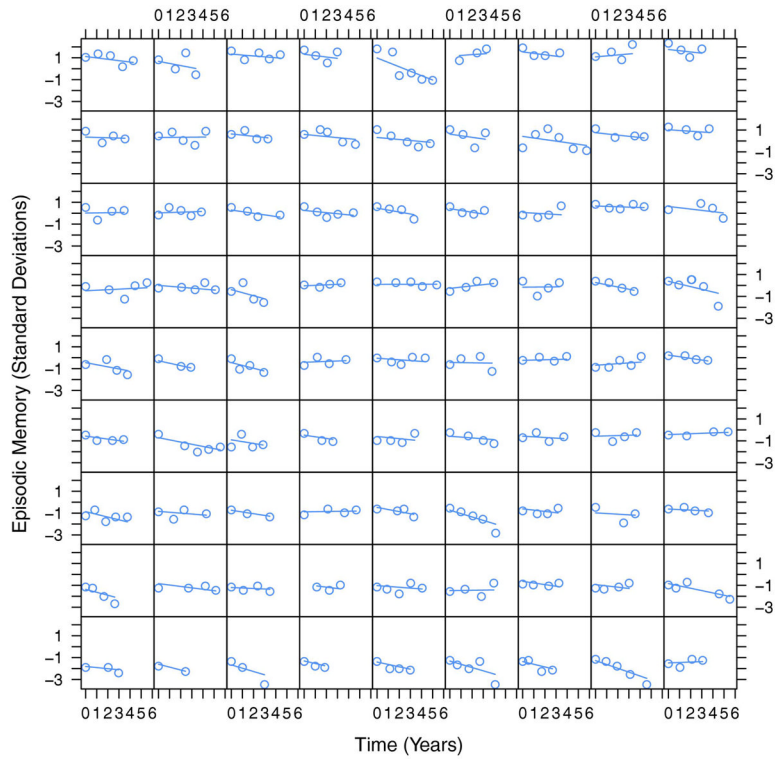


Figure 4. Episodic memory measurements and model-predicted linear episodic memory trajectories for 81 randomly-selected subjects. Each cell plots the real episodic memory measurements for a single subject as a function of time using circles; each line represents the predicted trajectory of episodic memory for the individual based on the mixed effects statistical model with demographic and MRI predictors.

Table 1
Cohort characteristics for 307 participants

Descriptive summary of the study population. Z scores for cognitive and MRI measures are broken down by baseline diagnostic category.

Variable	Summary
Demographics	
Gender (Number male, percent male)	117, 38%
Age (mean +/- stdev, range)	75.9 +/- 6.9 years, 60.0–4.5
Ethnicity (African American, Caucasian, Hispanic)	99, 129, 79
Education (mean +/- stdev)	12.6 +/- 4.5 years
Followup	
Number of cognitive evaluations (2, 3, 4, 5, >5)	92, 88, 77, 38, 12
Total duration of cognitive followup (mean +/- stdev)	2.9 +/- 1.36 years
Interval between cognitive assessments (mean +/- stdev)	1.31 +/- .40 years
Interval between baseline cognitive assessment and baseline MRI (mean +/- stdev)	.27 +/- .48 years
Baseline cognition and function	
Baseline clinical diagnosis (Normal, MCI, Demented)	163, 104, 40
Baseline MMSE (mean +/- stdev)	27.4 +/- 8.89
Baseline CDR sum of boxes (mean +/- stdev)	1.39 +/- 2.08

Domain-specific cognitive measures	Normal Z Score	MCI Z Score	Demented Z Score
Baseline episodic memory (mean z-score +/- stdev in z-score)	0.25 +/- 0.81	-0.72 +/- 0.66	-1.5 +/- 0.5
Yearly change in episodic memory (mean change in z-score per year +/- stdev)	-0.12 +/- 0.27	-0.13 +/- 0.44	-0.25 +/- 0.32
Baseline semantic memory (mean z-score +/- stdev in z-score)	0.28 +/- 0.77	-0.032 +/- 0.83	-0.8 +/- 1
Yearly change in semantic memory (mean change in z-score per year +/- stdev)	0.042 +/- 0.23	-0.086 +/- 0.25	-0.14 +/- 0.36
Baseline executive function (mean z-score +/- stdev in z-score)	0.056 +/- 0.61	-0.28 +/- 0.57	-0.95 +/- 0.69
Yearly change in executive function (mean change in z-score per year +/- stdev)	-0.0091 +/- 0.17	-0.14 +/- 0.3	-0.18 +/- 0.28

Baseline MRI	Overall	Normal Z Score	MCI Z Score	Demented Z Score
Brain volume (mean +/- stdev)	919.1 +/- 106.4 cm ³	.192 +/- .939	-.234 +/- 1.02	-.738 +/- .942
White matter hyperintensity volume (mean +/- stdev)	9.10 +/- 12.05 cm ³	-.125 +/- .961	.086 +/- 1.05	.399 +/- .944
Total hippocampal volume (left + right, mean +/- stdev)	3.62 +/- .70 cm ³	.248 +/- .901	-.184 +/- .935	-.70 +/- .924

Table 2

Mixed effects modeling results

Summary of mixed effects models of cognitive trajectories using multiple MRI predictors as fixed effects. Entries show the regression coefficient (β) for the listed fixed effect, followed by the associated p value for an F test on the marginal sum of squares for that fixed effect. Fixed effects with $p < 0.05$ are shown in bold. For baseline fixed effects, regression coefficients represent the number of standard deviations increase in the cognitive measure attributed to a 1 cc increase in the MRI variable. For fixed effects on change rate, regression coefficients represent the number of standard deviations increase in cognitive change per year attributed to a 1 cc increase in the MRI variable. Entries with a “**” represent fixed effects that were not entered into the final multivariate model because they failed significance tests earlier in the model-building process.

Baseline:		Effect of brain volume on baseline		Effect of hippocampal volume on baseline		Effect of WMH volume on baseline	
Cognitive Variable	β	p	β	p	β	p	
<i>Episodic memory</i>	0.167	0.001	0.219	<0.001	-1.135	0.007	
<i>Executive function</i>	0.19	<0.001	0.051	0.144	-1.089	0.012	
<i>Semantic memory</i>	0.17	<0.001	0.134	0.002	*	*	
Change:		Effect of brain volume on change		Effect of hippocampal volume on change		Effect of WMH volume on change	
Cognitive Variable	β	p	β	p	β	p	
<i>Episodic memory</i>	0.038	0.012	0.013	0.341	-1.016	0.253	
<i>Executive function</i>	0.011	0.349	0.016	0.136	-1.027	0.011	
<i>Semantic memory</i>	*	*	0.008	0.464	-1.045	0.001	