Prediction of Cognitive Decline in Healthy Older Adults using fMRI

John L. Woodard\textsuperscript{a}, Michael Seidenberg\textsuperscript{b}, Kristy A. Nielsen\textsuperscript{c,d}, J. Carson Smith\textsuperscript{b}, Piero Antuono\textsuperscript{d}, Sally Durgerian\textsuperscript{d}, Leslie Guidotti\textsuperscript{b}, Qi Zhang\textsuperscript{d}, Alissa Butts\textsuperscript{c}, Nathan Hantke\textsuperscript{e}, Melissa Lancaster\textsuperscript{b} and Stephen M. Rao\textsuperscript{f,∗}

\textsuperscript{a}Department of Psychology, and Institute of Gerontology, Wayne State University, Detroit, MI, USA
\textsuperscript{b}Department of Psychology, Rosalind Franklin University of Medicine and Science, North Chicago, IL, USA
\textsuperscript{c}Department of Psychology, Marquette University, Milwaukee, WI, USA
\textsuperscript{d}Department of Neurology, Medical College of Wisconsin, Milwaukee, WI, USA
\textsuperscript{e}Department of Human Movement Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
\textsuperscript{f}Neurological Institute, Cleveland Clinic, Cleveland, OH, USA

Handling Associate Editor: Diana Woodruff-Pak

Accepted 20 April 2010

Abstract. Few studies have examined the extent to which structural and functional MRI, alone and in combination with genetic biomarkers, can predict future cognitive decline in asymptomatic elders. This prospective study evaluated individual and combined contributions of demographic information, genetic risk, hippocampal volume, and fMRI activation for predicting cognitive decline after an 18-month retest interval. Standardized neuropsychological testing, an fMRI semantic memory task (famous name discrimination), and structural MRI (sMRI) were performed on 78 healthy elders (73\% female; mean age = 73 years, range = 65 to 88 years). Positive family history of dementia and presence of one or both apolipoprotein E (APOE) \(\varepsilon4\) alleles occurred in 51.3\% and 33.3\% of the sample, respectively. Hippocampal volumes were traced from sMRI scans. At follow-up, all participants underwent a repeat neuropsychological examination. At 18 months, 27 participants (34.6\%) declined by at least 1 SD on one of three neuropsychological measures. Using logistic regression, demographic variables (age, years of education, gender) and family history of dementia did not predict future cognitive decline. Greater fMRI activity, absence of an APOE \(\varepsilon4\) status, and larger hippocampal volume were associated with reduced likelihood of cognitive decline. The most effective combination of predictors involved fMRI brain activity and APOE \(\varepsilon4\) status. Brain activity measured from task-activated fMRI, in combination with APOE \(\varepsilon4\) status, was successful in identifying cognitively intact individuals at greatest risk for developing cognitive decline over a relatively brief time period. These results have implications for enriching prevention clinical trials designed to slow AD progression.

Keywords: Aging, apolipoprotein E, cognitive decline, fMRI, hippocampal volume, neuroimaging, memory

INTRODUCTION

Alzheimer’s disease (AD) neuropathology begins decades before the onset of observable symptoms [1]. Initiating interventions after symptom onset may be too late to make a meaningful impact on disease course. Clinical trials designed to prevent or slow AD progression have dramatically intensified the search for
valid preclinical biomarkers. Extant biomarker studies have demonstrated success in predicting conversion from mild cognitive impairment (MCI) to AD using neuropsychological testing [2–5]; structural magnetic resonance imaging (sMRI) measurement of hippocampal volume [6–8] and rate of atrophy [9–11]; sMRI of entorhinal cortex volume [11–14]; cerebrospinal fluid (CSF) indices including elevated isoprostane [15–17], elevated total tau and phosphorylated tau [18–20], and low amyloid-β (Aβ)12 levels [15,21–23]; positron emission tomography (PET) involving regional glucose metabolism [24,25] and amyloid imaging using the 11C Pittsburgh Compound B [26–28]; and task-activated functional magnetic resonance imaging (fMRI) [29,30]. A more challenging task for prevention trials, however, is to identify biomarkers capable of identifying asymptomatic older persons at-risk for developing cognitive decline within the time frame required of a prevention trial (2–3 years).

The apolipoprotein E (APOE) ε4 allele is a well-known risk factor for late onset AD [31,32], and healthy APOE ε4 carriers have demonstrated faster cognitive decline than non-carriers [33–35]. However, the biomarker potential for APOE alone is limited, given that the APOE ε4 allele frequency is less than 40% among AD cases [36,37], and it has a low positive predictive value for AD diagnosis [38–40]. Using test-retest intervals of approximately three years, studies using fluorodeoxyglucose (FDG) PET [34] and CSF tau181/β42 and ptau181/β42 ratios [41] have shown promise for predicting cognitive decline in otherwise healthy older adults. The relative invasiveness of these latter two approaches may preclude their routine use in screening large numbers of cognitively intact participants for inclusion in prevention trials.

Less invasive magnetic resonance imaging (MRI) techniques provide more practical alternatives for identifying cognitively intact older adults at risk for future cognitive decline. sMRI studies have demonstrated that smaller hippocampal and entorhinal cortex volumes at baseline predict cognitive decline in healthy elders [42–47]. A task-activated fMRI study [48] has also shown that increased number and spatial extent of activated regions at baseline can predict memory decline after a two-year retest interval. Genetic risk in middle aged women (family history of AD and at least one APOE ε4 allele) has been associated with decreased fMRI activation in extrastriate and posterior inferotemporal cortex at baseline, together with further decreases after four years in these regions as well as left inferior frontal and premotor cortex [49]. However, this decreased fMRI signal was not associated with cognitive decline in this study. In contrast, using a word categorization task during fMRI with APOE ε4 carriers, nine older adults showing cognitive stability on episodic memory testing after five years demonstrated increased left inferior parietal activation at baseline relative to nine participants who demonstrated episodic memory decline; greater blood oxygen level dependent (BOLD) fMRI response in this region was associated with better memory performance after five years [50]. However, no hippocampal volume differences were observed at baseline between stable and declining participants. No study to date has directly compared the relative sensitivity of sMRI and fMRI approaches, particularly over a relatively brief interval (e.g., 1–2 years).

In this study, we compared the ability of sMRI and fMRI to predict cognitive decline over 18 months in a sample of cognitively intact older adults with varying degrees of AD risk, based on family history of dementia and APOE ε4 allele carrier status. The sMRI technique involved measurement of hippocampal volumes. The fMRI task required the discrimination of famous from unfamiliar names. Our previous studies using this task reported activation of a semantic memory system, including bilateral hippocampi, posterior cingulate, middle frontal gyrus, and lateral temporoparietal junction [51–53]. The task can be performed with a high degree of accuracy (> 90% correct) even in symptomatic amnestic MCI patients [54]. In a cross-sectional study [55], we demonstrated that the brain activation patterns of healthy elders at risk for developing AD (APOE ε4, family history) could be differentiated using this task. The current longitudinal prospective study used logistic regression to compare the relative efficacy of sMRI and fMRI, alone and in combination, for predicting cognitive decline after an 18-month retest interval. Because a greater potential exists for accelerated cognitive decline among APOE ε4 carriers [33–35], we examined APOE genotype as an additional predictor of decline.

MATERIALS AND METHODS

Participants

Participants were 78 healthy older adults (73% female; Mage = 73 years, SD = 4.9 years; Meducation = 14.9 years, SD = 2.7 years). The participants were drawn from a larger sample of 459 community-dwelling adults who were recruited
via newspaper advertisements. Following telephone screening, 92 participants met study inclusion and exclusion criteria, and 81 persons agreed to undergo APOE genotyping from blood samples, a neuropsychological evaluation, and an fMRI scanning session. MRI data were not able to be obtained for three participants. Family history was defined as a report of a clear clinical diagnosis of AD or a reported history of gradual decline in memory and other cognitive functions, confusion, or judgment problems without a formal diagnosis of AD prior to death in a first-degree relative. One participant reported a diagnosis of AD in a second degree relative, with some mild cognitive changes noted in a parent prior to the parent’s death. Because our study examined the influence of AD risk factors on prediction of cognitive decline, half of the participants were purposely selected because they had a positive family history of AD. We expected that enrichment of our sample with persons with a positive family history of AD would also increase the number of persons who were purposely selected because they had a positive family history of AD. We expected that enrichment of our sample with persons with a positive family history of AD would also increase the number of persons who were APOE ε4 positive, because APOE ε4 tends to be more common among individuals with a positive AD family history than among those with a negative AD family history [56,57].

Family history of dementia was present in 51.3% of participants, and 33.3% of the sample carried the APOE ε4 allele. All participants underwent neuropsychological evaluation (see below) and were cognitively intact when entering the study. Informed consent was obtained consistent with the Declaration of Helsinki and institutional guidelines established by the Medical College of Wisconsin Human Subjects Review Committee; all participants received financial compensation.

**Neuropsychological assessment and APOE genotyping**

All participants underwent baseline neuropsychological testing, fMRI scanning, and APOE genotyping. The neuropsychological battery included the Mini-Mental State Examination [58], Mattis Dementia Rating Scale-2 (DRS-2) [59,60], Rey Auditory Verbal Learning Test (RAVLT) [61], Geriatric Depression Scale [62], and Lawton Instrumental Activities of Daily Living Scale (ADL) [63]. Alternate forms of the DRS-2 [64,65] and RAVLT [66] were used. APOE genotype was determined using a PCR method [67]. DNA was isolated with Gentra Systems Autopure LS for Large Sample Nucleic Acid Purification. All participants underwent a follow-up neuropsychological examination after approximately 18 months.

### Definition of cognitive decline

We defined cognitive decline as a reduction from baseline performance of at least one SD on at least one of the three principal outcome indices (DRS-2, RAVLT Sum of Trials 1–5 [T1-5], RAVLT Delayed Recall [DR]). Residualized change scores were computed for each cognitive measure by predicting T2 scores using T1 scores; this procedure adjusts for baseline performance, practice effects, and regression to the mean [68–70]. Participants with standardized residuals of $−1.0$ or lower were assigned to the cognitively declining group; the remaining participants were classified as cognitively stable.

### fMRI task

For the fame discrimination task [53], stimuli consisted of 30 famous and 30 unfamiliar names randomly interspersed with 20 presentations of a centrally placed crosshair in order to introduce “jitter” into the fMRI time series (interstimulus interval = 4 sec). Participants made a right index or right middle finger key press for famous or unfamiliar names, respectively. Accuracy and reaction time were recorded, and nonparametric signal detection indices were calculated [71]. The imaging run began and ended with 12 sec of fixation and was 5 min and 44 sec in duration.

### Image acquisition

Whole-brain, event-related fMRI was conducted on a General Electric (Waukesha, WI) Signa Excite 3.0 Tesla short bore scanner equipped with a quad split quadrature transmit/receive head coil. Echoplanar images were collected using an echoplanar pulse sequence (TE = 25 ms; flip angle = 77 degrees; field of view (FOV) = 24 cm; matrix size = 64 × 64; TR = 2 s). Thirty-six contiguous axial 4-mm-thick slices provided coverage of the entire brain (voxel size = 3.75 × 3.75 × 4 mm). High-resolution, three-dimensional spoiled gradient-recalled at steady-state (SPGR) anatomic images were acquired (TE = 3.9 ms; TR = 9.5 ms; inversion recovery (IR) preparation time = 450 ms; flip angle = 12 degrees; number of excitations (NEX) = 2; slice thickness = 1.0 mm; FOV = 24 cm; resolution = 256 × 224). Foam padding was used to reduce head movement within the coil.
Functional images were generated with the Analysis of Functional NeuroImages (AFNI) software package [72]. Each image series was time shifted to the beginning of the TR and spatially registered to reduce head motion effects using a rigid body iterative linear least squares method. A deconvolution analysis was used to extract separate hemodynamic response functions (HRFs) for correctly recognized famous and unfamiliar names. HRFs were modeled for the 0–16 second period post-stimulus onset. Motion parameters and incorrect trials were incorporated into the model as nuisance regressors. Area under the curve (AUC) was calculated by summing the hemodynamic responses at time points 4, 6, and 8 seconds post trial onset, a measure of the curve peak yielding maximum signal-to-noise. Anatomical and functional scans were transformed into standard stereotaxic space [73]. To compensate for anatomical variation, functional images were blurred using a 6 mm Gaussian full-width half-maximum filter.

Spatial extent of activation for cognitively stable and declining groups

Voxelwise t-tests were used to generate separate statistical parametric maps for the stable and declining groups. These maps indicate regions where the AUCs for famous and unfamiliar names were significantly different. The statistical threshold was based on an individual voxel probability \( p = 0.005 \) coupled with a minimum cluster volume (0.73 ml). These values were derived from 3,000 Monte Carlo simulations [74] and correspond to a whole brain family-wise error threshold of \( p < 0.05 \).

Functional ROI analysis

A separate voxelwise t-test, comparing famous and unfamiliar names, was conducted on all 78 participants using the identical statistical threshold. This method identified significant cluster volumes, which we refer to as functional regions of interest (fROIs). For each participant, an “average AUC” was calculated for all voxels within each fROI. These data were then subjected to a principal components analysis (PCA) to further reduce the number of regions that would serve as predictors in the logistic regression analysis (see below).

Hippocampal volume measurement

Left and right hippocampal volumes were created using Freesurfer [75,76] and manually edited on T1-weighted SPGR images by two raters blinded to participant group membership. Using coronal views, the mask was further refined by excluding the fimbria and alveus and retaining the hippocampus (uncal apex, cornu ammonis, subiculum, gyrus of retzius, and fasciola cinerea). Hippocampal volumes were normalized by dividing by the total intracranial volume. Intraclass correlation for the two raters was 0.87. The left and right hippocampal volumes were then summed to create a single score.

Data analysis

Statistical analyses were performed using R, version 2.9.0. Group differences on demographics, total hippocampal volume, and neuropsychological and fMRI task performance were compared using t-tests and \( r^2 \) effect size measures or Fisher’s Exact tests, as appropriate. Logistic regression tested the ability of specific baseline variables to discriminate between stable and declining participants. To avoid overfitting the data and to maintain a reasonable subjects-to-variables ratio for each model, we restricted the set of predictors to no more than four variables. Our models tested the effects of age, education, and gender (Model 1); APOE \( \varepsilon4 \) status and dementia family history (Model 2); hippocampal volume (Model 3); fMRI activation (Model 4). Models 5 and 6 examined the additive effect of APOE \( \varepsilon4 \) status with either hippocampal volume or fMRI activation, respectively. Model 7 combined APOE \( \varepsilon4 \) status with both imaging predictors. The ability of these models to differentiate between stable and declining participants was assessed using the Nagelkerke \( R^2 \) and the concordance or C index (related to the area under the receiver operating characteristic curve [77]). The Nagelkerke \( R^2 \) assesses the importance of the predictors in a given model relative to a “perfectly fitting” null model [78]. The C index reflects the proportion of all possible pairs of declining and stable subjects in which the declining participant in the pair had a higher predicted probability of decline than the stable participant [77]. Values of \( R^2 \) and C for each logistic regression model were validated with a bootstrapping analysis using 5000 resamples in order to assess each model’s accuracy of prediction of decline across the entire range of probabilities [77]. This approach yielded bootstrap-corrected values for \( R^2 \) and
Results

Identification of cognitive decline

A total of 27/78 (34.6%) participants showed a one SD decline on at least one of the three neuropsychological indices (DRS-2, RAVLT Trials 1–5, and RAVLT Delayed Recall). These participants constituted the cognitively declining group and the remaining participants formed the stable group. Figure 1 illustrates performance changes on the neuropsychological outcome measures for the stable and declining groups. As expected, the stable group showed no significant neuropsychological change after 18 months, while the declining group demonstrated significant reductions on each of the three neuropsychological indices.

Subjective memory complaints were present in 33.3% of the declining group. Of the declining participants, 2 (7.6%) satisfied criteria for MCI [79]. No participant demonstrated impaired ADL skills at follow-up. Declining participants did not differ from stable participants on age, education, gender, or neuropsychological outcomes.

C. Bootstrapping is the most efficient model validation procedure, as it does not require holding out any data for cross-validation, and each phase of model development (including assessment of the degree of overfitting the data) is revalidated using repeated resampling from the entire sample [77].
Table 1
Sample characteristics, neuropsychological performance and fMRI behavioral data for stable and declining groups

<table>
<thead>
<tr>
<th>Sample Characteristics</th>
<th>Stable (n = 51)</th>
<th>Declining (n = 27)</th>
<th>p</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
<td><strong>Range</strong></td>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td><strong>Sample Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>72.7</td>
<td>5.1</td>
<td>65–88</td>
<td>73.7</td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.1</td>
<td>2.5</td>
<td>12.0–23.0</td>
<td>14.6</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (M)</td>
<td>13</td>
<td>25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (F)</td>
<td>38</td>
<td>75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family History Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (−)</td>
<td>27</td>
<td>53%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1+ (≥1)</td>
<td>24</td>
<td>47%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE ε4 Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (−)</td>
<td>39</td>
<td>76%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1+ (≥1)</td>
<td>12</td>
<td>24%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retest Interval (days)</td>
<td>55</td>
<td>43.5</td>
<td>489–763</td>
<td>560.6</td>
</tr>
<tr>
<td>Total Hippocampal Volume</td>
<td>4.5</td>
<td>0.55</td>
<td>3.2–5.9</td>
<td>4.2</td>
</tr>
<tr>
<td><strong>Neuropsychological Testing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>29.4</td>
<td>0.8</td>
<td>27–30</td>
<td>28.9</td>
</tr>
<tr>
<td>DRS-2 Total</td>
<td>140.7</td>
<td>3.2</td>
<td>127–144</td>
<td>139.7</td>
</tr>
<tr>
<td>RAVLT Trials 1–5</td>
<td>50.6</td>
<td>8.8</td>
<td>33–66</td>
<td>46.8</td>
</tr>
<tr>
<td>RAVLT DR</td>
<td>10.1</td>
<td>2.6</td>
<td>4.0–15.0</td>
<td>9</td>
</tr>
<tr>
<td>Lawton IADL</td>
<td>5</td>
<td>0</td>
<td>5–5</td>
<td>5</td>
</tr>
<tr>
<td>GDS</td>
<td>2.7</td>
<td>2.5</td>
<td>0–8</td>
<td>1.4</td>
</tr>
<tr>
<td><strong>fMRI Task Performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Correct Famous</td>
<td>93.1</td>
<td>6.8</td>
<td>73–100</td>
<td>90.4</td>
</tr>
<tr>
<td>% Correct Unfamiliar</td>
<td>96.9</td>
<td>4.6</td>
<td>77–100</td>
<td>95.2</td>
</tr>
<tr>
<td>d’</td>
<td>5.8</td>
<td>1.2</td>
<td>3.4–7.6</td>
<td>5.4</td>
</tr>
<tr>
<td>RT Famous</td>
<td>1266</td>
<td>222</td>
<td>790–1886</td>
<td>1270</td>
</tr>
<tr>
<td>RT Unfamiliar</td>
<td>1670</td>
<td>354</td>
<td>912–2602</td>
<td>1611</td>
</tr>
</tbody>
</table>

Note: MMSE=Mini-Mental State Examination; DRS-2 = Mattis Dementia Rating Scale-2; RAVLT = Rey Auditory Verbal Learning Test; DR = delayed recall; IADL=Instrumental Activities of Daily Living; GDS=Geriatric Depression Scale; d’=signal detection discrimination; RT=Reaction Time. * = data were constant.

Logical retest interval (Table 1). However, the APOE ε4 allele was over twice as prevalent (51.9% versus 23.6%) among declining (3 ε2/ε3, 10 ε3/ε3, 14 ε3/ε4) than stable (5 ε2/ε3, 34 ε3/ε3, 11 ε3/ε4, 1 ε4/ε4) participants.

Baseline neuropsychological testing and fMRI task performance

On baseline measures (Table 1), no significant differences were observed on the MMSE, DRS-2, RAVLT Trials 1–5, and RAVLT Delayed Recall between the stable and declining groups after controlling for multiple comparisons (Bonferroni adjusted alpha level = 0.0125; 0.05/4 tests). The stable group reported significantly more depressive symptoms on the GDS, but no participant in either group was in the clinically depressed range. None of the participants reported ADL impairments at baseline.

On the fMRI fame discrimination task, no differences were observed in accuracy, RT, or on a signal detection measure of discriminability (d’) between the stable and declining groups. For both groups, mean accuracy exceeded 90% for identification of famous names and rejection of unfamiliar names.

Baseline sMRI

Declining participants had a significantly smaller total hippocampal volume at baseline than cognitively stable participants (Table 1).

Baseline fMRI

Figure 2 presents significant clusters based on a voxelwise analysis performed separately for the stable and declining groups. The spatial extent of activated voxels is greater in the stable than declining group, with most of the differences reflecting more activation during recognition of famous names relative to unfamiliar names. The declining group showed a smaller amount of activated tissue, with some regions showing the opposite pattern.

Figure 3A represents the results of the voxelwise analysis performed on the entire sample. This analysis, restricted to the famous > unfamiliar name comparison, yielded eight fROIs (Table 2). A PCA was conducted on the average AUCs of these fROIs, yielding two components accounting for 73% of total variance (Table 2). Five fROIs loaded significantly on a “Cortical” component, shown in green in Fig. 3A, where-
Fig. 2. Group differences in activation derived from the comparison of the famous versus unfamiliar names condition: Famous > Unfamiliar represented in red; Unfamiliar > Famous in blue. Note the greater spatial extent of activation in the Famous > Unfamiliar names comparison in the stable group.

as two regions loaded on the “Hippocampal” component (purple regions in Fig. 3A). The right cerebellum did not demonstrate significant loadings [80] on either component and was dropped from the analysis. The unfamiliar > famous name comparison resulted in four fROIs and a single PCA component accounting for 63.3% of the total variance. This component did not predict cognitive decline and is not discussed further.

Figure 3B presents a graph of the fMRI signal response to famous and unfamiliar names compared to fixation (rather than just the comparison of these conditions) to address the question of whether the effect is driven primarily by activation to famous names or the response to novel names. For cognitively stable participants, both the Cortical and Hippocampal fMRI signal demonstrated positive changes in the AUC in response to famous names and a decreased AUC in response to unfamiliar names. In contrast, the cognitive-ly declining participants showed the opposite pattern, with greater AUC in response to unfamiliar names and reduced AUC when presented with unfamiliar names. Using a mixed-design ANOVA that tested the effects of group (stable vs. declining) and stimulus type (famous vs. unfamiliar), significant group by stimulus type interactions were observed for the Cortical (F(1,76) = 8.88, p < 0.004) and Hippocampal (F(1,76) = 8.11, p = 0.006) fMRI components.

Logistic regression analyses

Seven logistic regression models were evaluated. For each model, bootstrap-corrected R² and C values are presented in Table 3. For each predictor within a model, coefficients, standard errors, and significance levels are shown in Table 3, and odds ratios with 95% confidence intervals are presented in Fig. 4. Models 1 and 2 indicate that age, education, gender and family history of AD were not significant predictors of future cognitive decline. For Models 2–7, APOE status, cortical and hippocampal fMRI activation, and hippocampal volume each contributed significantly to the prediction of cognitive decline. Although Model 7 demonstrates the largest R² (0.293) and C index (0.789), only two of the four predictors were statistically significant (cortical fMRI activation and APOE status), whereas the remaining two predictors (hippocampal fMRI activation and hippocampal volume) were not. Model 5 (R² = 0.285; C' = 0.787) was the second best model, with APOE genotype and both cortical and hippocampal fMRI activation each contributing significantly to the prediction of future cognitive decline.

The Adequacy Index [77] is a recommended way of comparing the adequacy of a set of predictors across models. It is unitless and is represented by ratio of the −2 log likelihood statistic for testing a subset of predictors for the model of interest to the -2 log likelihood ratio statistic for testing the joint significance of the full set of predictors. It ranges between 0 (no predictive information for the subset of predictors) to 1 (complete predictive information for the subset of predictors). Using the full set of predictors in Model 7, the Adequacy Indexes for Models 2–6 are presented in Table 3. Model 1 was not included as there was no significant predictor of decline using demographic variables. The fMRI measures alone (Model 3) account for 46% of the total explanatory power for the set of variables, compared to hippocampal volume alone (Model 4), which accounts for only 27% of the total explanatory power. Perhaps more dramatically, Model 5, which uses the
Table 2
Activation foci for famous versus unfamiliar name subtraction (Famous > Unfamiliar)*

<table>
<thead>
<tr>
<th>Region #</th>
<th>Region</th>
<th>Cortical Component Loadings</th>
<th>Hippocampal Component Loadings</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Stable Component Group Loadings</th>
<th>Declining Component Group Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bilateral Posterior Cingulate Cortex, Precuneus</td>
<td>0.884</td>
<td>0.257</td>
<td>−0.9</td>
<td>−52.4</td>
<td>24.6</td>
<td>37.05</td>
<td>1.44</td>
</tr>
<tr>
<td>2</td>
<td>Left Angular Gyrus</td>
<td>0.889</td>
<td>0.266</td>
<td>−45.4</td>
<td>−55.9</td>
<td>23.7</td>
<td>26.14</td>
<td>18.68</td>
</tr>
<tr>
<td>3</td>
<td>Left Superior Frontal Gyrus</td>
<td>0.805</td>
<td>0.018</td>
<td>−17.8</td>
<td>29.5</td>
<td>41</td>
<td>30.84</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Right Angular Gyrus</td>
<td>0.815</td>
<td>0.275</td>
<td>46.2</td>
<td>−59.6</td>
<td>27.1</td>
<td>18.68</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Right Superior, Middle Frontal Gyrus</td>
<td>0.839</td>
<td>0.009</td>
<td>22.5</td>
<td>16.5</td>
<td>46.8</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Left Parahippocampal Gyrus, Hippocampus</td>
<td>0.269</td>
<td>0.898</td>
<td>−21.9</td>
<td>−21</td>
<td>−10.7</td>
<td>2.29</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Right Parahippocampal Gyrus, Hippocampus</td>
<td>0.053</td>
<td>0.946</td>
<td>23.9</td>
<td>−23.2</td>
<td>−11.6</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Right Cerebellum (Crus 1)</td>
<td>0.425</td>
<td>0.268</td>
<td>11.3</td>
<td>−74.8</td>
<td>−21.9</td>
<td>1.03</td>
<td></td>
</tr>
</tbody>
</table>

Note: Critical value ($2r_{crit,p=0.01}$) used to identify significant component loadings was 0.560 [80]. * = PCA conducted on four negative activation (Unfamiliar > Familiar) clusters (Left Precentral Gyrus; Bilateral Supplementary Motor Area; Right Insula; Left Middle Occipital Gyrus) revealed one component that did not predict decline; this component was dropped from subsequent analyses.

Fig. 3. A) Regions comprising the Cortical (green) and Hippocampal (purple) fMRI activation principal components for the Famous > Unfamiliar names comparison. B) Cortical and Hippocampal fMRI signals (areas under the curve) contrasting famous name recognition versus fixation and unfamiliar name identification versus fixation for cognitively stable and declining participants.
**DISCUSSION**

Clinical trials involving pharmacological and lifestyle (exercise, cognitive enrichment, diet) interventions are being considered to prevent or delay the onset of AD, even before symptoms emerge. For clinical trials to be maximally successful, enrichment of the sample with elders at the greatest risk for experiencing cognitive decline over the course of a typical clinical trial (2–3 years) is essential. Results of our prospective study indicate that combining genetic risk and MRI biomarkers can effectively identify such individuals, even after a relatively brief 18-month retest interval. Specifically, we were able to correctly order 78.9% of possible pairs of stable and declining participants using a combination of APOE genotype, cortical and hippocampal fMRI, and hippocampal volumes. APOE genotype and fMRI (cortical and hippocampal) predictors alone correctly ordered 78.7% of possible pairs. In contrast, hippocampal volume, alone or combined with APOE status, correctly ordered only 68.7% and 70.2% of pairs, respectively. Without the benefit of imaging data, family history of dementia and APOE status correctly ordered only 61.5% of possible pairs (chance prediction = 50%). Overall, our findings suggest that the combination of fMRI and APOE genotype

---

**Fig. 4.** Odds ratios and 95% confidence intervals for seven logistic regression models. Odds ratios whose 95% confidence intervals overlap with 1.0 (represented by vertical dashed line) are not statistically significant. Odds ratios > 1 indicate greater probability of decline with increasing value of predictor; odds ratios < 1 indicate reduced probability with increasing predictor values.

fMRI measures plus APOE genotype status accounts for 87% of the explanatory power compared to Model 6 (hippocampal volume plus APOE genotype status), which accounts for only 43% of the explanatory power.
Table 3

Results of logistic regressions

<table>
<thead>
<tr>
<th>Model</th>
<th>Likelihood Ratio</th>
<th>Adequacy Index</th>
<th>R²</th>
<th>C Index</th>
<th>Variables</th>
<th>Coeff</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>1.69</td>
<td>- - -</td>
<td>0</td>
<td>0.585</td>
<td>Age</td>
<td>0.036</td>
<td>0.051</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Education</td>
<td>-0.088</td>
<td>0.093</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
<td>-0.142</td>
<td>0.546</td>
<td>0.795</td>
</tr>
<tr>
<td>Model 2</td>
<td>6.33</td>
<td>0.252</td>
<td>0.063</td>
<td>0.615</td>
<td>Family Hx</td>
<td>0.143</td>
<td>0.524</td>
<td>0.785</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ApoE</td>
<td>1.207</td>
<td>0.533</td>
<td>0.014</td>
</tr>
<tr>
<td>Model 3</td>
<td>11.53</td>
<td>0.46</td>
<td>0.155</td>
<td>0.713</td>
<td>Cortical fMRI</td>
<td>-0.67</td>
<td>0.293</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hipp. fMRI</td>
<td>-0.642</td>
<td>0.302</td>
<td>0.034</td>
</tr>
<tr>
<td>Model 4</td>
<td>6.76</td>
<td>0.27</td>
<td>0.098</td>
<td>0.687</td>
<td>Hipp. Volume</td>
<td>-1.16</td>
<td>0.48</td>
<td>0.016</td>
</tr>
<tr>
<td>Model 5</td>
<td>21.77</td>
<td>0.869</td>
<td>0.285</td>
<td>0.787</td>
<td>Cortical fMRI</td>
<td>-0.874</td>
<td>0.309</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hipp. fMRI</td>
<td>-0.699</td>
<td>0.323</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ApoE</td>
<td>1.846</td>
<td>0.612</td>
<td>0.003</td>
</tr>
<tr>
<td>Model 6</td>
<td>10.8</td>
<td>0.431</td>
<td>0.132</td>
<td>0.702</td>
<td>Hipp. Volume</td>
<td>-0.991</td>
<td>0.49</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ApoE</td>
<td>1.056</td>
<td>0.528</td>
<td>0.045</td>
</tr>
<tr>
<td>Model 7</td>
<td>25.06</td>
<td>1</td>
<td>0.293</td>
<td>0.789</td>
<td>Cortical fMRI</td>
<td>-0.929</td>
<td>0.336</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hipp. fMRI</td>
<td>-0.644</td>
<td>0.348</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hipp. Volume</td>
<td>-0.945</td>
<td>0.545</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ApoE</td>
<td>1.558</td>
<td>0.64</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Note: Adequacy index reflects the total explanatory power of a subset of predictors relative to a model containing the total set of predictors (Model 7) using the ratio of the likelihood ratio of the model of interest to the likelihood ratio of the model containing the total set of predictors.

status holds promise for successfully screening at-risk, but asymptomatic, participants for prevention trials.

Our results would appear to be at odds with a similar prospective fMRI study [48], in which increased brain activation was associated with lower scores on episodic memory tasks after a two year retest interval. It is important to note two important methodological differences between the two studies. First, the number of participants who underwent follow-up neuropsychological testing in the earlier study (n = 14) was considerably smaller than those in the current study (n = 78). Second, the previous study used an effortful episodic learning and recall task and did not report task performance during fMRI scanning. It is conceivable that declining participants performed more poorly at baseline on the fMRI task than those who were stable over the retest interval. Such differences in task performance, if present, could have a meaningful impact on the pattern of brain activation, especially since error trials could not be eliminated from the blocked design trial format used in the previous fMRI study. In contrast, the current event-related study used a low effort, high accuracy (> 90% correct) semantic memory task in which the few error trials that did occur were excluded from the final image analyses.

In a previous study [55], we reported greater semantic memory activation in cognitively intact, APOE ε4 carriers relative to non-carriers. As in the current study, we defined semantic memory activation by a greater BOLD response to famous than unfamiliar name stimuli. Based solely on the cross-sectional results reported in our previous study, one might predict that greater semantic activation would be a predictor of future cognitive decline. However, in our prior study, we did not segregate declining from stable participants within each of the two risk groups. The current longitudinal results suggest that having increased semantic memory activation may paradoxically afford a protective effect against future cognitive decline in both high and low risk individuals. This effect is illustrated in Fig. 5. The 12 APOE ε4 carriers in the stable group demonstrated greater cortical activation in response to familiar than unfamiliar names; in contrast, the 14 declining APOE ε4 carriers exhibited greater activation to unfamiliar than to familiar names. Among the non-carriers, a similar pattern was observed, albeit with less overall semantic memory activation for the group as a whole. Among the 39 stable non-carriers, the degree of cortical activation was comparable for famous and unfamiliar names, whereas the 13 declining non-carriers demonstrated greater activation for unfamiliar than famous names.

Our finding that increased baseline fMRI activation is protective against future cognitive decline in cognitively intact elders is consistent with prior studies reporting increased task-related BOLD signal in parietal cortex in cognitively stable participants after five years [50,81]. Increased activation may reflect greater cognitive reserve in asymptomatic persons, particularly in regions subserved by the cholinergic system.
Fig. 5. Percent MR signal intensity (± SEM) for stable and declining APOE ε4 carriers (ε4+) and non-carriers (ε4-). Positive values reflect greater BOLD response aggregated across activated cortical regions in response to famous relative to unfamiliar names; negative values reflect greater BOLD response to unfamiliar relative to familiar names.

Increased brain activation in these regions has been observed following administration of cholinesterase inhibitors in MCI and AD patients [82–87]. We speculate that improved cognitive reserve, possibly manifested by increased neuronal firing rate or recruitment of additional supportive neuronal regions, permits continued functioning at a higher level in the face of early neurodegenerative changes. Persons who have lost this propensity for functional compensation are at increased risk of future cognitive deterioration.

Our famous name recognition task activates brain regions (posterior cingulate gyrus, posterior inferior parietal cortex, middle temporal gyrus, fusiform and parahippocampal gyri, hippocampus, and medial superior frontal gyrus) commonly associated with the “default mode network” (DMN) [88,89]. The DMN is frequently correlated with uncontrolled semantic processing resulting from task-unrelated thoughts that occur during resting scan conditions. Prior work by Binder and colleagues [90,91] has demonstrated considerable overlap between brain systems associated with the resting state DMN and those activated by controlled semantic memory processing tasks. Recent studies [92–94] have suggested that disruption of the DMN can occur in early AD. Not surprisingly, our results have shown that participants who have experienced cognitive decline after 18 months also demonstrate reduced baseline semantic memory activation in cortical regions that overlap with the DMN (see Fig. 5). Moreover, the AUCs corresponding to the fMRI signal in both cortical and hippocampal regions were reduced in response to famous names and increased in response to unfamiliar names for cognitively declining participants (Fig. 3B). Cognitively stable participants showed the opposite pattern. Future studies are required to determine the relative sensitivity of baseline measurements of the resting state DMN versus task-activated semantic memory processes in predicting future cognitive decline among asymptomatic persons.

Baseline hippocampal volume, corrected for intracranial volume, significantly predicted future cognitive decline, both alone and combined with APOE genotype. However, its predictive accuracy was not as strong as the combination of fMRI and APOE genotype. Considerable inter-individual variability in hippocampal volumes occurs in cross-sectional studies of cognitively intact elders, and hippocampal volume may sometimes be inversely related to cognitive abilities [95]. A recent meta-analysis concluded that the relationship between hippocampal size and episodic memory performance across the lifespan was weak [96]. While rate of hippocampal atrophy has provided more compelling evidence of a relationship with cognitive decline in healthy older adults [97–99], the requirement of two measurement periods separated by up to two or more years makes this biomarker impractical for widespread use for enriching prevention trials.

This study adds to the growing body of literature showing that combinations of biomarkers show greater predictive accuracy compared to individual biomarkers [16,100]. A stepwise combination of biomarkers might be considered for balancing invasiveness, cost-containment, and predictive accuracy when used in the context of identifying at-risk, but otherwise healthy, participants for prevention trials. For instance, a prevention trial screening process might include APOE genotyping as a first step, followed by task-related fMRI activation performed in APOE ε4 carriers. More invasive tests, such as CSF biomarkers and PET imaging (FDG or amyloid), could then be administered as further selection criteria for enriching study samples. However, because we did not perform these additional tests, we cannot state conclusively whether these procedures provide incremental predictive accuracy beyond the combination of APOE genotyping and task-activated fMRI.

It is important to acknowledge other limitations of this study. Our neuropsychological battery focused on cognitive abilities most likely to be affected in early AD and may have missed significant changes in other cognitive domains among our stable participants. Our
study also defined cognitive decline based on change in neuropsychological test scores rather than on a change in diagnostic category, i.e., conversion to MCI or early AD. In our opinion, the rate of conversion from intact cognition to MCI/AD is too low to be used as a meaningful outcome variable in prevention trials. Nevertheless, the extent to which increased baseline fMRI activation is specific to predicting early AD-related changes or more general age-related cognitive decline will await long-term follow-up studies. Finally, despite the fact that all participants performed within normal limits at baseline on all cognitive measures, baseline neuropsychological performance demonstrated non-significant trends for lower performance in the declining group on the two RAVLT measures and on the MMSE. Thus, it is conceivable that several participants in the declining group were actively undergoing cognitive decline. However, our outcome measure was based on the degree of cognitive decline from baseline performance (1 SD or more) rather than absolute levels of performance. Furthermore, our regression-based approach to defining cognitive change controlled each participant’s follow-up level of performance for his or her baseline level of performance. Small baseline differences would be unlikely to account for the dramatic cognitive change in the declining group relative to the stable group as depicted in Fig. 1.

In summary, our study provides evidence of the ability of task-related fMRI, in combination with APOE genotype, to predict future cognitive change in healthy older adults. This combination of static genetic propensity to develop AD and an fMRI approach that measures brain activity during a low-effort, high accuracy task, can be valuable for enriching a prevention trial with healthy persons at high risk of impending cognitive decline. Biomarker combinations tapping different aspects of pathological changes associated with AD that are widely available, easily implemented, minimally invasive, and relatively inexpensive will likely assume increasing importance in future clinical trials designed to prevent or slow AD progression.

ACKNOWLEDGMENTS

This study was supported by grants from the National Institutes on Aging (AG022304), Medical College of Wisconsin General Clinical Research Center (RR00058), and the W.M. Keck Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.


REFERENCES


