Reliability of fMRI Associative Encoding Memory Paradigm in Non-Demented Elderly Adults

Deepti Putcha[1,2,3], **Kelly O'Keefe**[1,2,3], **Pete LaViolette**[2,3], **Jackie O'Brien**[1,2,3], **Doug Greve**[2,3], **Dorene M. Rentz**[1,3], **Joseph Locascio**[3], **Alireza Atri**[2,3,4], and **Reisa Sperling**[1,2,3]

[1]Center for Alzheimer Research and Treatment, Department of Neurology, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, 02115

[2]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, 02114

[3]Department of Neurology, Massachusetts General Hospital, Boston, MA 02144

[4]Center for Cognitive Translational Neuroscience, Geriatric Research Education Center, Bedford Veterans Administration Medical Center, Bedford, MA 01730

Abstract

Functional MRI holds significant potential to aid in the development of early interventions to improve memory function, and to assess longitudinal change in memory systems in aging and early Alzheimer's disease. However, the test-retest reliability of hippocampal activation and of "beneficial" deactivation in the precuneus has yet to be fully established during memory encoding tasks in older subjects. Using a mixed block and event-related face-name associative encoding paradigm, the reliability of hippocampal activation and **default network** deactivation was assessed over a four-to-six week inter-scan interval in 27 older individuals who were cognitively normal (Clinical Dementia Rating Scale= 0; n=18) or very mildly impaired (CDR=0.5; n=9). Reliability was assessed in whole brain maps and regions-of-interest using both a full task paradigm of six functional runs as well as an abbreviated paradigm of the first two functional runs, which would be advantageous for use in clinical trials. We found reliable hippocampal signal response across both block and event-related designs **in the right hippocampus**. Comparable reliability in hippocampal activation was found in the full and the abbreviated paradigm. Similar reliability in hippocampal activation was observed across both CDR groups overall, but the CDR 0.5 group was more variable in left hippocampal activity. Task-related deactivation in the precuneus demonstrated much greater variability than hippocampal activation in all analyses. Overall, these results are encouraging for the utility of fMRI in "Proof of Concept" clinical trials investigating the efficacy of potentially therapeutic agents for treatment of age-related memory changes, cognitive impairment, and early Alzheimer's disease.

Keywords

test-retest; hippocampus; precuneus; reliability; mild cognitive impairment; aging

Correspondence: Reisa Sperling, M.D. Center for Alzheimer's Research and Treatment Brigham and Women's Hospital 221 Longwood Avenue, Boston, MA 02115 reisa@rics.bwh.harvard.edu.

INTRODUCTION

Functional MRI (fMRI) has become an important research tool in studying the neural correlates of cognitive processes in normal and diseased brains. Specifically, task-related blood oxygen level dependent (BOLD) fMRI has been instrumental in elucidating the neural systems involved in episodic memory, but thus far remains primarily a tool for cognitive neuroscience research. fMRI has great potential for utility in pharmacological trials, particularly for candidate cognitive-enhancing therapeutics targeted at age-related memory impairment and early Alzheimer's disease (AD). An essential step in validating the use of fMRI for "proof of concept" AD clinical trials is the quantitative assessment of test-retest reliability.

Of particular interest in the older population is the reproducibility of hippocampal activation during episodic memory tasks. A number of imaging studies **in young subjects** have demonstrated greater activity in the hippocampus and other structures of the medial temporal lobe, as well as in pre-frontal cortices, during the encoding of stimuli that are later successfully remembered (Brewer, et al. 1998; Sperling, et al. 2003b; Wagner, et al. 1998). The hippocampus is also specifically implicated in episodic encoding across a range of impairment in older subjects. Alterations in fMRI activity in the hippocampus and related structures in the medial temporal lobe have been observed in low performing "normal" older adults (Daselaar, et al. 2003; Miller, et al. 2008a), subjects at genetic risk for AD (Bondi, et al. 2005; **Bookheimer, et al. 2000**; Han, et al. 2007; Trivedi, et al. 2006), across the continuum of MCI (Celone, et al. 2006; Dickerson, et al. 2004; **Johnson, et al. 2006**; Machulda, et al. 2003; **Mandzia, et al. 2009**; Miller, et al. 2008b), and in mild AD (Dickerson, et al. 2005; Golby, et al. 2005; Machulda, et al. 2003; Remy, et al. 2005; Sperling, et al. 2003b).

There is additional evidence to suggest that successful memory formation necessitates not only hippocampal engagement, but also coordinated activation and deactivation of various structures in a distributed memory network that is altered in the process of aging. Deactivation is defined here as decreases in signal during the task as compared with a fixation baseline or another control condition. Regions that typically demonstrate deactivation during the encoding of novel information, as well as other cognitive tasks requiring attention to external stimuli, have been characterized as the "default network" (Buckner, et al. 2008; Fransson 2006; Raichle, et al. 2001). Recent fMRI studies suggest that deactivation in key nodes of the default network, in particular precuneus/posterior cingulate regions, may be beneficial to successful encoding (Daselaar, et al. 2004), and that this process may be altered in the process of aging (Duverne, et al. 2009; Miller, et al. 2008a) and early Alzheimer's disease (Celone, et al. 2006; Pihlajamaki, et al. 2009). The relationship between functional activity in the hippocampus and the precuneus/posterior cingulate continues to be studied in an attempt to elucidate the neural underpinnings of cognitive impairment in early Alzheimer's disease.

It is important to establish an accurate assessment of test-retest reliability of functional activity in these regions in order to carry out meaningful longitudinal fMRI studies, as well as to quantify change in neural activity associated with pharmacological treatments that may impact memory function. Relatively few studies have assessed test-retest reliability in functional imaging. Many of these studies examined reliability within a single scan session, or are limited to simple sensorimotor or visual tasks. The inter-scan intervals that have been studied extensively are also either on the order of a few days or weeks (Kong, et al. 2006; Rombouts, et al. 1997), or a year or more (Aron, et al. 2006). Little is known about reproducibility over a middle range inter-scan interval that might be used in early phase clinical trials. Reproducibility studies have also focused mainly on younger, healthy subjects

over a range of various sensorimotor (Kiehl and Liddle 2003; Machielsen, et al. 2000; Rombouts, et al. 1997; Specht, et al. 2003), language (Eaton, et al. 2008; Rutten, et al. 2002), or memory tasks (Sperling, et al. 2002). The literature assessing the reliability of functional activity in the hippocampus and other medial temporal lobe structures related to associative memory, as well as deactivations in the precuneus region in elderly non-demented subjects along a spectrum of cognitive impairment, is still nascent. A recent study by Clement and Belleville (2009) evaluated the test-retest reliability of fMRI activation in cognitively normal subjects and individuals with mild cognitive impairment (MCI) six weeks apart during both a phonological processing task and verbal episodic memory encoding and retrieval. They found that MCI subjects showed comparable reproducibility measures to that of normal older controls in the average of activity in multiple regions, including Broca's area, bilateral prefrontal cortex, precuneus, posterior cingulate cortex, and the hippocampus. However, they did observe some variability in the hippocampal response across conditions and clinical group. Reproducibility of fMRI activity in working memory tasks has also been explored in other patient populations, such as schizophrenia (Manoach, et al. 2001), producing variable estimates depending on brain region.

Reliability of task-related signal also depends on a number of variables, including but not limited to: transient environmental and physiological fluctuations, behavioral performance of the subjects, and the inter-scan interval (Liou, et al. 2003). Increased intra-individual variability, reflecting within-person fluctuations in behavioral performance, is commonly observed in aging and in cognitively impaired subjects (MacDonald, et al. 2009), and may contribute to lower test-retest reliability measures in older subjects than in younger subjects. It is therefore essential to quantify fMRI test-retest reliability to aid in optimal design and interpretation of fMRI experiments in older populations.

Studies assessing reliability have employed various techniques, including but not limited to the Dice spatial overlap coefficient (as adapted by Rombouts et al., 1997), and the intra-class correlation coefficient (ICC). Previous studies have suggested that spatial overlap ratios are higher in sensorimotor tasks than in higher level cognitive tasks (Clement and Belleville, 2009). Additionally, the ICC values reported by these studies appear to vary greatly depending on which region of the brain is being studied.

Our primary objective in this study was to evaluate the test-retest reliability of memory related fMRI activity, with a specific focus on the hippocampus and the posteromedial regions of the default network, as these regions have emerged as being both the most critical in memory formation and greatly affected by aging and early AD. We examined test-retest reliability of activity at the whole brain map level, as well as within these *a priori* regions of interest (ROIs). We utilized a clinically relevant, cross-modal associative memory paradigm that compares the encoding of novel face-name pairs to viewing highly familiarized repeated face-name pairs, as difficulty remembering proper names remains the most common complaint of older individuals (Zelinski and Gilewski 1988). Our previous work with this paradigm in young subjects has indicated robust hippocampal activation with good reliability over short term intervals (Sperling, et al. 2002).

One of our ultimate goals is to validate fMRI for use in short "Proof of Concept" clinical trials involving potentially therapeutic agents for treatment of age-related memory impairment and very early Alzheimer's disease. We chose an inter-scan interval of four-to-six weeks because this interval is typical of early phase clinical trials utilizing cognitive enhancing agents, and avoids the confound of potential disease progression in prodromal AD over a longer time interval. Older subjects with cognitive impairment who are recruited for such trials may have limited tolerance for lengthy scans, particularly accounting for multiple other MR sequences that are required for safety evaluations and volumetric

analyses. We therefore investigated the reproducibility of an abbreviated version of this fMRI paradigm that could be easily integrated into a standard safety and volumetric imaging protocol. Additionally, we investigate the contribution of behavioral performance to the reproducibility of fMRI activity by comparing a block design with an event-related design that assesses successful memory encoding. We evaluated reproducibility across whole brain maps and within specific anatomic ROIs defined from each individual's structural MRI data to determine the reliability of fMRI activity at both regional and whole brain network levels in these older subjects.

MATERIALS AND METHODS

Participants

Twenty-seven right-handed non-demented older adults (10 males, 17 females; mean age 72.4 years, range: 61–83) consented to participate in this study (Table I). All subjects were screened for neurological and psychiatric illness. Subjects were recruited from on-going longitudinal studies of aging and from neurology clinics and were screened for contraindications to MRI. At study entry, the Mini-Mental State Examination (MMSE) was administered and each subject scored between 27–30. Upon entering the study, subjects were classified based on the Clinical Dementia Rating (CDR) scale. CDR scores were based on an interview with the subject and the subject's healthy study partner with whom the subject had daily contact. Eighteen subjects were classified as having a CDR of 0 (cognitively normal), and nine as CDR 0.5 (mildly impaired). The assignment of CDR 0.5 in this study was based on reports of subjective memory complaints, corroborated on by a study partner. These subjects did not have significant memory impairment evidenced by neuropsychological testing, and thus would not meet Petersen criteria for amnestic MCI (Petersen, 2004). However, the CDR 0.5 subjects might be considered to be "pre-MCI" on the basis of their subjective memory complaints and mild memory impairment. The Partners Human Research Committee at Brigham and Women's Hospital and Massachusetts General Hospital approved all study procedures.

Neuroimaging Procedure

Each subject was scanned twice within an interval of 4–6 weeks (mean inter-scan interval 5.15 +/− 1.5 weeks). Each scanning session was conducted at the same time in the mid-morning using identical preparation and scanning procedures. Each session included 25 minutes of structural imaging sequences, followed by the fMRI memory paradigm which was a mixed block and event-related design adapted from a previously published fMRI block design paradigm (Celone, et al. 2006) and a subsequent memory event-related design (Miller, et al. 2008a).

Subjects were scanned during encoding runs consisting of alternating blocks of Novel and Repeated face-name pairs (40 seconds each), interspersed with blocks of visual fixation on a white crosshair (25 seconds). The stimulus duration for face-name pairs within each block were also jittered with very brief periods of visual fixation, with optimal timings derived from OptSeq (Greve, 2002). Faces were each displayed for 4.5 seconds against a black background with a fictional first name printed in white letters underneath. During the presentation of each face-name pair, subjects were asked to press a button indicating whether they thought the name was a "good name for the face" or a "bad name for the face", a purely subjective assignment designed to ensure attention to the task and enhance associative encoding (Sperling, et al. 2001). Before each run, subjects were explicitly instructed to try to remember which name was associated with which face. Face-name stimuli were randomly intermixed with trials of visual fixation (a white crosshair centered on a black background) that varied in duration from 0.3s to 2.2s. Visual fixation was

presented for 25s between each block of seven Novel and seven Repeated face-name pairs. We use two different face-name stimulus sets for each imaging session, which are taken from seven stimulus sets that have been previously validated in both young and older subjects to show equivalent post-scan memory performance (Sperling et al., 2002; Miller et al., 2008a). Stimuli were presented using MacStim 2.5 software (WhiteAnt Occasional Publishing, West Melbourne, Australia). Images were projected through a collimating lens onto a mirrored screen attached to the head coil. Responses were collected using a fiber-optic response box held in the right hand. Cushions were in place to help minimize subject movement, and headphones were used to communicate with the subjects during the scan and to dampen scanner noise.

Post-scan Memory Testing

After the scanning session, subjects underwent a forced-choice recognition memory test outside the scanner in which the 84 novel faces and 2 repeated faces seen during the scan session were presented on a computer screen outside the scanner room. Each face was paired with two names underneath in a counter-balanced design: one that was correctly paired with the face during the scan session, and one that was paired with a different face during the scan session. Subjects were asked to indicate which of the two names was correct and also to indicate if they had high or low confidence regarding their answer choice.

Image Acquisition

Subjects were scanned using a Siemens Trio 3T scanner (Siemens Medical Systems, Erlangen, Germany). T1-weighted structural images were acquired using a Magnetization Prepared- Rapid Acquisition Gradient Echo (MP-RAGE) sequence: repetition time (TR) = 2300 msec, echo time (TE)= 2.98 msec, inversion time (TI) = 900 msec, flip angle (FA) = 9 degrees, field of view (FOV) = 256mm, matrix 240 × 256, slice thickness- 1.20 mm, 160 sagittal slices (right to left). Blood oxygen level dependent (BOLD) fMRI data were acquired using a T2*-weighted gradient-echo echo-planar imaging (EPI) sequence: TR = 2000ms, TE= 30ms, FA= 90 degrees, FOV = 200mm, matrix = 64 × 64 (in-plane resolution 3.1×3.1 mm^2). Thirty oblique coronal (anterior to posterior) slices with 5.0 mm thickness and an interslice gap of 1.0 mm were acquired, oriented perpendicularly to the anterior-posterior commissural line. A total of six functional runs per scan session were acquired, each consisting of 127 whole-brain acquisitions with 5 TRs discarded for T1 stabilization. The total scanning time for all six functional runs was 25.44 minutes, with the acquisition time of each run being 4.24 minutes.

Image Processing and Analysis

MP-RAGE images were processed through the FreeSurfer pipeline (http://surfer.nmr.mgh.harvard.edu). As part of this semi-automated pipeline, preprocessing of structural volumes included an affine registration to Talairach space, bias field correction, and removal of skull and dural voxels surrounding the brain. Each volume underwent minimal manual intervention (e.g. manual skull strip, guiding the cortical segmentation procedure) to ensure that the white matter and the pial surfaces were properly reconstructed. All other processing steps were fully automated using the default parameters.

Functional MRI data were preprocessed using Statistical Parametric Mapping (SPM2; Wellcome Department of Cognitive Neurology, London, UK) for Matlab (The Mathworks, Inc, Natick, Massachusetts, USA). Functional data were realigned using INRIAlign, a motion correction algorithm unbiased by local signal changes, normalized to the standard SPM2 EPI template, re-sampled into 3 mm isotropic resolution in MNI305 space, and then smoothed using an 8 mm Gaussian kernel. No scaling was implemented for global effects. A high pass filter of 260s was used to remove low frequency signal (e.g. drifts across entire

fMRI run). The data were then modeled by convolving a canonical haemodynamic response function with the onsets from encoded face-name pairs.

We measured the technical quality of each scan by calculating the signal-to-noise ratio (SNR) of each fMRI run, and then calculated an average SNR over all six functional runs for each subject at each scan session. We conducted a paired-t test to determine if the SNR over the whole group was comparable between the scan sessions. We also examined head motion parameters in six directions, to ensure that no scans exceeded three standard deviations in any direction. We utilized this liberal threshold for movement parameters to emulate a clinical trial situation in older and cognitively impaired subjects.

For the block design analyses, trials were categorized as either novel (N) or repeated (R) face-name pairs, and compared to Fixation (F). Activation contrasts of interest defined from a block design were novel versus repeated face-name pairs (NvR) and novel face-name pairs versus a fixation cross (NvF). These contrasts were selected on the basis of previous work, by our group and others, which has elicited robust hippocampal activation when comparing the encoding of novel stimuli compared to the viewing of repeated stimuli in memory encoding paradigms (Sperling, et al. 2003a; Stern, et al. 1996). Using an event-related analysis, encoding trials could also be classified on the basis of whether or not they were subsequently correctly identified (hit vs. miss), and with high (HC) or low confidence (LC) on the post-scan recognition memory test (Miller, et al. 2008a). Contrasts of interest defined from the event-related analyses were high confidence hits versus repeated face-name pairs (HCHvR) and high confidence hits versus a fixation cross (HCHvF). We did not use a "Hits vs. Misses" contrast for this study because this mixed block and event-related design did not yield robust hippocampal activation at either scanning session, even with young subjects. This is likely due to the limited jitter possible within the block timing constraints, as well as the relatively small number of stimuli in the miss trials in normal subjects.

Both block and event-related analyses were based on SPM2 mixed-effects linear models. To assess memory task-related activation, NvR, NvF, HCHvR, and HCHvF contrasts were created for each subject. To quantify task-related deactivation, fixation versus all novel and repeated faces (FvALL) and fixation versus high-confidence hits (FvHCH) contrasts were created on the basis of previous work showing that passive fixation contrasted to active task involving cognitively engaging external stimuli highlights beneficial deactivation in the midline parietal region (Daselaar, et al. 2004; Lustig, et al. 2003; McKiernan, et al. 2003; Pihlajamaki, et al. 2008). We first examined whole-brain voxel-wise activation and deactivation at the group level. One-sample t-tests were run separately on all of these contrasts. At the next level, mixed-effects paired t-tests on the whole group were utilized to determine voxel-by-voxel differences between scanning sessions. Results were considered to be statistically significant at p<0.001 (uncorrected) with a minimum extent threshold of 5 voxels.

Test-Retest Reliability Analysis Within ROIs

We utilized *a priori* functionally defined ROIs in the hippocampus and precuneus, constrained by individually defined anatomic regions. We sought to test whether the signal measured from these areas during the encoding of novel face-name pairs was reproducible over a time interval typical of inter-assessment intervals in early phase AD clinical trials. We examined reliability for both the novel face-name pairs compared to repeated pairs in a block design, and also specifically for face-name pairs that were subsequently "successfully" encoded in an event-related design. The functional volumes from each scanning session were aligned to the structural volume to determine which functional voxels lay within the FreeSurfer-derived, individual, anatomically defined hippocampus and precuneus ROIs (see Fig. 1). The ROIs were further constrained with a union mask of

hippocampal activation and precuneus deactivation present at either scanning session created from cross-sectional group maps of each specific contrast, thresholded at p<0.001 with an extent of 5 voxels. This process created ROIs that were specific to individual subject anatomy that were also focused on subregions that were engaged during the task in at least one of the two sessions. We also performed analyses on lateral parietal deactivation using a post-hoc ROI to determine if this has any better reliability than observed in the precuneus.

Magnitude of fMRI activation, defined as percent signal change (PSC) in the BOLD signal, was extracted from each ROI at the individual level to determine the reproducibility of fMRI activity between scanning sessions. Paired *t*-tests were again used to analyze significant differences in activation between Scan Session 1 and Scan Session 2. Intra-class Correlation Coefficients (ICCs), a statistic that quantifies the consistency or reproducibility of measurements or raters, were calculated to quantify the test-retest reliability of behavioral performance as well as fMRI task-related activations and deactivations. The ICC has numerous variations. We employed the one termed "ICC (2,1)" by Shrout and Fleiss (1979) computed as:

$$(BMS-EMS) / (BMS + (k-1) EMS + k (TMS-EMS) /n)$$

where

BMS = the mean square (sum of squares/degrees of freedom) for the subjects,

TMS = the mean square for time, or scan session factor here,

EMS = the mean square for error,

k= number of scanning sessions of assessment, and

n= number of subjects.

We chose this definition of ICC because it assumes the scanning sessions of assessment are randomly selected from a large population of potential assessments, although the same points of time are used for all subjects. This version is also sensitive to (attenuated by) undesirable mean change across assessments as well as interaction of subjects with assessment (essentially lack of correlation between assessments).

Two types of ICCs are reported: single measure (relevant to cross-sectional studies) as well as average measure (relevant to longitudinal studies). Here, single measure reliability implies that individual ratings constitute the unit of analysis, whereas average measure reliability implies that the mean of all ratings is the unit of analysis. That is, average measure reliability gives the reliability of the average rating.

Another form of assessing reliability is to solely examine the within-subject variability. We defined within-subject variability as the standard deviation of BOLD signal changes across scanning sessions (Zandbelt, et al. 2008), which is essentially an index of subject x assessment interaction effect. The reliability metric reported is the standard deviation of the change scores (σ_{change}) **of all the individuals in the group.**

In order to examine main effects and possible interactions between run type (6 runs vs. 2 runs), scanning session (Scan Session 1 vs. Scan Session 2), hemisphere (left vs. right), and CDR subgroup (CDR 0 vs. CDR 0.5), we calculated two types of three-way mixed between-within subjects repeated measures ANOVAs on the PSC data: "run x scanning session x CDR group" as well as "hemisphere x scanning session x CDR group." All repeated measures ANOVAs were run separately for each contrast of interest.

In addition to these reliability measures of magnitude of fMRI activity, we investigated the reliability of extent of fMRI activity within the ROIs using Dice spatial overlap coefficients. The Dice coefficient is defined as twice the intersection of the area of Scan Session 1 activation and Scan Session 2 activation all divided by the union of the two areas. This measure was calculated for each subject individually, from his or her whole anatomic hippocampus and precuneus, and then averaged for each contrast at a group level. All analyses were run with the full paradigm (6 functional runs) as well as with an abbreviated paradigm (2 functional runs).

RESULTS

Participants

Independent samples *t*-tests were conducted to compare demographic data between the CDR subgroups, including age, gender, MMSE score, and years of education. None of these variables differed significantly between groups except for gender (Fischer's exact test, p = 0.039), reflecting a disproportionately high number of females in the CDR 0 group than in the CDR 0.5 group.

Memory Test Performance

A summary of performance on the post-scan face-name recognition test is presented in Table II. The CDR 0.5 subjects correctly identified 58% of trials, and correctly recognized 26% of the face-name pairs with high confidence, compared to the CDR 0 group, which correctly identified 68% of the face-name pairs, and correctly recognized 45% of the stimuli with high confidence. Notably, **independent samples *t*-tests demonstrated that** subjects classified as CDR 0.5 performed **significantly** worse than those classified as CDR 0 at both scanning sessions (Scan Session 1: t= 2.52, p=0.019; Scan Session 2: t= 3.083, p=0.005). However, within the CDR subgroups, subjects performed comparably between scanning sessions. ICCs for the percentage of answers reported correctly with high confidence demonstrate high reproducibility of the scores across groups: individual ICC (mean ICC): all subjects: 0.76 (0.86); CDR 0: 0.61 (0.76); CDR 0.5: 0.87 (0.93).

We did not find any statistically significant relationships between change in memory test performance and change in signal in either activation or deactivation across scan sessions. However, change in signal across sessions was related to baseline behavioral performance on the task during the HCHvF contrast, such that greater observed change in the BOLD signal in the hippocampus between scanning sessions was correlated with worse memory performance (lower percentage of successful recognition with high confidence) during the first scan session. This was true for the full paradigm (r= −0.427, p=0.026) as well as the abbreviated paradigm (r=−0.496, p=0.008). There was no significant relationship between change in signal during deactivation and behavioral performance at either scan session.

Signal-to-Noise Ratio Assessments of fMRI Data

Technical quality in the form of the signal-to-noise ratio (SNR) of each functional run of each subject's scanning session was calculated. All runs excepting two runs from one subject at the second scanning session had a SNR above the accepted value (>100). The overall SNR (averaged over the six functional runs) for each subject at each scanning session was above threshold. Lastly, paired *t*-test analyses suggest that the group had comparable SNR across scanning sessions (t=1.009, p=0.322, N.S).

Whole-Brain Activation

We first examined whole-brain voxel-by-voxel activation at Scan Session 1 and Scan Session 2 separately, analyzing group-level activity with both block and event-related analyses using one-sample *t*-tests thresholded at p<0.001, uncorrected. In the block design random effect group analysis one-sample *t*-test for the NvR contrast, subjects demonstrated comparable patterns of whole-brain activation at both scanning sessions. Additionally, we specifically observed activation in the hippocampus bilaterally, as well as in fusiform gyrus and prefrontal cortices, similar to results in previous experiments with this paradigm (see Fig. 2). Group-level activation maps for NvF also demonstrated a similar pattern at both scanning sessions. Interestingly, these results were also observed for NvR and NvF in the abbreviated cognitive paradigm analyzing just the first two functional runs. Likewise, group activation contrasting the encoding of HCHvR showed comparable whole brain activation in both the full paradigm and the abbreviated paradigm between the two scanning sessions (see Fig. 3). Group maps of the CDR subgroups separately demonstrated very similar results for each scanning session. In order to capture all task-related activity within the hippocampus, particularly for the ROI analyses of this small region, we utilized a relatively liberal threshold of p<0.001 (uncorrected) as our primary analysis. Whole brain activation, and in particular, hippocampal activation, was still observed with an FDR correction for multiple comparisons, p<0.05, in all contrasts and run types with the exception of the CDR 0.5 subgroup in the abbreviated paradigm.

We then examined the differences in whole brain activation between scanning sessions with paired *t*-tests in SPM2. Thresholded at an uncorrected p-value of 0.001 and extent of 5 voxels, there were no significantly different clusters identified for bidirectional tests of either Scan Session 1 > Scan Session 2 or Scan Session 2 > Scan Session 1 in any of the activation contrasts for the full paradigm. In analyses of the abbreviated paradigm, a small number of voxels demonstrated differential activity at Scan Session 2 greater than Scan Session 1, but these were located outside of the hippocampal ROI, primarily in regions around the edge of the brain or the ventricles, which may represent motion artifact due to head motion or ventricular pulsation. Similar results were observed when analyzing paired *t*-tests of the CDR subgroups separately. Overall, in the full paradigm there were no significant differences for bidirectional tests of Scan Session 1 > Scan Session 2 or Scan Session 2 > Scan Session 1 in any activation contrast. However, in the abbreviated paradigm examining NvR, the CDR 0 subgroup showed a few voxels activated at Scan Session 2 that were not activated at Scan Session 1 in the visual cortex.

In particular, we observed that the hippocampus was activated bilaterally in a similar pattern at both scanning sessions for both block and event-related analyses, and for the full and abbreviated paradigm. For NvR, peak MNI coordinates were: Scan Session 1 Left [−24 −18 −18], Right [27 −18 −15] and Scan Session 2 Left [−27 −6 −21], Right [24 −18 −12]. In event-related analyses, peak MNI coordinates for HCHvR were: Scan Session 1 Left [−30 −27 −9], Right [27 −18 −15] and Scan Session 2 Left [−27 −9 −18], Right [27 −18 −15].

ROI Test-Retest Analysis of Activation in the Hippocampus

In addition to whole-brain analyses, we investigated *a priori* region specific reproducibility using functionally defined ROIs constrained by each individual's neuroanatomy. Percent signal change during the various block- and event-specific contrasts was extracted from each subject's hippocampus ROI at both scanning sessions. The results from the NvR analysis from the full paradigm are shown in Fig. 4.

A three-way mixed between-within subjects repeated analysis of variance was conducted to assess the impact of number of runs (6 vs. 2), scanning sessions (Scan Session 1 vs. Scan

Session 2), and CDR group status (0 vs. 0.5) on subject's activation measures. There was no significant interaction between run number and scanning session, Wilks Lambda=0.963, F = 0.956, p= 0.338, partial eta squared = 0.037. Likewise, there were no significant interactions between run number and CDR subgroup, or scanning session and CDR subgroup, and no significant three-way interactions between run number, CDR subgroup, and scanning session. Additionally, there were no significant main effects of scanning session, CDR subgroup, or run number.

Paired t-tests were then used to more specifically investigate any evidence of difference in the magnitude of activation for the whole group, and for each CDR subgroup, across time. None of the paired t-tests indicated a significant difference in magnitude of activation within the hippocampal ROIs for NvR, NvF, HCHvR, or HCHvF. Intra-class correlation coefficients (ICCs) were calculated from the values of Scan Session 1 and Scan Session 2 percent signal change as a measure of signal reliability (See Table III). Across the various contrasts in both block and event-related designs, ICCs for the whole sample were moderate for the left hippocampus (**0.35** - 0.6) and moderate-high (0.6– 0.8) in the right hippocampus. Importantly, ICCs were comparable for both the full and abbreviated cognitive paradigm.

Dice spatial overlap coefficients were on average 0.37 in left hippocampus and 0.36 in right hippocampus for NvR. The within-subject variability (σ_{change}) was also calculated. Again, it was observed that there was less variability in the right hippocampus than the left, with lower σ_{change} in the right hippocampus observed than in the left across contrasts in the full paradigm (σ_{change} in NvR : left= 0.25, right= 0.20; in HCHvR: left= 0.31 right= 0.20) as well as in the abbreviated paradigm (σ_{change} in NvR : left= 0.29, right= 0.22; in HCHvR: left= 0.28, right= 0.20).

Whole-Brain Deactivation

Although the primary focus of this study was on hippocampal activity, we also investigated deactivation at Scan Session 1 and Scan Session 2 separately by contrasting fMRI activity that was greater during the presentation of a fixation cross than during the encoding of all novel and repeated face-name pairs (FvALL), as well as during the encoding of subsequently correctly recognized face-name pairs with high confidence (FvHCH).

The FvALL contrast was preferred over an FvN contrast because we saw more evidence of consistent deactivation in the precuneus at baseline across all older subjects when examining fixation compared to all novel and repeated face-name pairs as opposed to fixation compared to just novel face-name pairs. In examining the FvALL contrast, the group demonstrated deactivation in the precuneus region at Scan Session 1 and Scan Session 2 in the full paradigm (peak MNI coordinates were: Scan Session 1 Left [−3 −45 45], Right [3 −42 51]; Scan Session 2 Left [−24 −51 9], Right [24 −54 21]), but only in the left precuneus and only at Scan Session 2 in the abbreviated paradigm (see Figure 5; MNI coordinates x,y,z:[−9 −81 36]). There were more voxels deactivating during FvALL at Scan Session 2 as compared to Scan Session 1. To see comparable patterns of deactivation between scanning sessions, and to be able to compare the full to the abbreviated paradigm at both scanning sessions, it was necessary to drop the threshold to the level of p<0.01. Areas of the lateral parietal region were also significantly deactivated within this contrast at the group level, thus we also extracted signal estimates from a lateral parietal ROI (see Figure 6). Similar results were observed in group-level activation maps of the contrast FvHCH. Again, more voxels deactivated at Scan Session 2 than at Scan Session 1. With the abbreviated paradigm, few voxels appeared to be deactivating at Scan Session 1, and at Scan Session 2, only the left precuneus showed deactivation.

Next, we examined whole brain deactivation differences between scanning sessions with paired *t*-tests in SPM2. Thresholded at an uncorrected p-value of 0.001 with an extent of 5 voxels, group maps of deactivation at Scan Session 1 versus Scan Session 2, bidirectionally, did not show significant differences in the precuneus/posterior cingulate area when contrasting FvALL. However, when analyzed within CDR subgroups, the CDR 0 group showed evidence of small clusters of greater deactivation in the Scan Session 2 > Scan Session 1 paired *t*-tests for both FvALL and FvHCH contrasts.

ROI Test-retest Analysis of Deactivation in the Precuneus

Similar to the analyses of hippocampal activation, percent signal change in the FvALL and FvHCH contrasts were extracted from each subject's precuneus ROI, and compared using paired *t*-tests. Overall, the results demonstrated that the magnitude and location of deactivation within the precuneus ROI was highly variable. For the FvALL contrast, deactivation was measurable for most subjects in the full paradigm, but not in the abbreviated paradigm examining just the first two runs. Likewise, no significant deactivation was observed for FvHCH in the abbreviated paradigm.

Again, a three-way mixed between-within subjects repeated analysis of variance was conducted to assess the impact of number of runs (6 vs. 2), scanning sessions (Scan Session 1 vs Scan Session 2), and CDR group status (0 vs. 0.5) on subject's activation measures. Because of the lack of measurable data for all contrasts and runs, we could only perform three ANOVAs using deactivation contrasts. In examining a "run x CDR subgroup x scanning session" interaction for the contrast of FvHCH in the right precuneus, there were no significant interactions between run number and scanning session. However, both run number and scanning session had significant main effects, run number: Wilks' Lambda = 0.546, F=21.6, p=0.000; scanning session: Wilks' Lambda = 0.851, F= 4.55, p= 0.043. There were no significant main effects or interactions in examining a "hemisphere x CDR subgroup x scanning session" ANOVA for the full paradigm FvALL contrast or for the abbreviated paradigm FvHCH contrast.

To explicitly investigate any evidence of difference in deactivation, we again conducted paired *t*-tests for FvALL and FvHCH contrasts. The full paradigm demonstrated that there were no significant differences overall between Scan Session 1 and Scan Session 2 deactivation for the whole group. However, deactivation in the right precuneus for just the CDR 0.5 group in the full paradigm showed a trend for significance (p=0.097). Because both the intra-subject as well as inter-subject magnitude of deactivation was so variable, ICC values for precuneus deactivations were low (0.2 – 0.4), signifying poor test-retest reliability.

Dice spatial overlap coefficients in FvALL also showed less overlap in deactivation compared to the results for activation in the hippocampus, averaging 0.28 in the left precuneus and 0.31 in the right precuneus (ranging from 0 to 0.55). Overall within-subject variability, σ_{change}, was higher for precuneus deactivation than for hippocampal activation (σ_{change} in FvALL : left = 0.45, right= 0.34; in FvHCH: left= 0.33).

Power Analyses

We conducted a series of power calculations based on hippocampal activation data. Our goal is to model acute drug effects, as opposed to modeling longitudinal changes in AD related decline. For proof-of-concept studies of acute pharmacologic effects, we would hope to be able to detect a moderate effect size of 50% change in fMRI signal. Using block design data from the right hippocampus, power analyses indicate that in order to detect a 50% difference in hippocampal activity with a power of 0.8 and two-sided alpha of 0.05, approximately 25

subjects would be required using the full 6 run paradigm and 35 subjects for the abbreviated paradigm of just the first two functional runs. Event-related analyses yielded similar power estimates, with 26 subjects needed for the full paradigm and 29 subjects for the abbreviated paradigm.

DISCUSSION

This study investigated the test-retest reliability of an fMRI face-name associative encoding paradigm in cognitively intact and mildly memory-impaired older individuals over a four to six week inter-scan interval. Importantly, good reliability of activation patterns in whole brain maps and specifically in the hippocampus was observed over this intermediate inter-scan interval in elderly non-demented older adults. Furthermore, we observed good reproducibility of hippocampal activation by using an abbreviated fMRI paradigm that would be suitable to add on to standard safety and volumetric protocols in a clinical trial. Substantially more variability was observed in the pattern and magnitude of deactivation within the medial parietal (precuneus) regions of the default network using the same analytic methods, suggesting that task-induced deactivation may be less reliable overall than hippocampal activation, in older and cognitively impaired subjects.

We investigated percent signal change in regions of interest using a method that would allow us to sample from regions engaged in the task within each individual's anatomy. We restricted our analyses to voxels that were actually activating or deactivating at either scanning session. We further defined the ROIs by the regions where these activated areas lay within each individual's anatomically defined hippocampus and precuneus regions. This was done to better account for differences in individual anatomy, as older subjects might have early regional atrophy in the medial temporal lobe and the medial parietal cortices. Notably, when analyses were repeated using only MNI-based, structurally defined ROIs, we observed very similar results to those reported here using individual anatomically defined ROIs. Thus it may not be critical to include individual volumetric information in fMRI analyses when only normal or mildly impaired subjects are included. However, the need for volumetrically individualized fMRI analysis may be required when studying more impaired MCI or AD patients, who likely have greater regional atrophy.

Behavioral performance on the post-scan memory test did not demonstrate any statistical differences between scanning sessions, and may reflect the fact that most of our subjects performed fairly well at both scanning sessions. Nevertheless, it is important to note that event-related contrasts demonstrated similar signal reliability as block design contrasts since event-related designs may prove more useful in the evaluation of individuals who demonstrate significant cognitive change over time. However, a disadvantage to event-related designs is the inherent loss of statistical power as well as insufficient bin sizes for some contrasts based on inter-subject variability, particularly when utilizing an abbreviated version of this paradigm.

One important objective of this study was to establish the reliability of hippocampal signal during an associative memory task, given this region's importance in memory function and previous concerns about the inter-subject variability of hippocampal activity (Daselaar, et al. 2003; Rombouts, et al. 1997). Similar to results of reliability studies of fMRI paradigms investigating auditory working memory (Wei, et al. 2004) and fear responses in the amygdala (Johnstone, et al. 2005), we found that paired-associate encoding-related hippocampal activation has good reproducibility, based both on ICC values and within-subject variance measures. ICCs for hippocampal ROIs mostly ranged between 0.4–0.8, showing moderate to high signal reliability for this associative encoding task. Individual ICCs from the left hippocampus (~0.5), was generally lower than the ICCs from right

hippocampus (~0.7). These ICCs are in the range of those reported in other imaging reliability studies (Eaton, et al. 2008; Johnstone, et al. 2005; Manoach, et al. 2001; Wei, et al. 2004), and are also consistent with results from a recent fMRI study reporting on a memory paradigm in older subjects (Clement and Belleville 2009). It is perhaps not surprising that we observed somewhat better signal reliability from the right hippocampus, as the literature implicates the right hippocampus in playing a crucial role in spatial encoding (Schacter, et al. 1996; Sperling, et al. 2001; Squire, et al. 1992). Furthermore, our paradigm may specifically probe the role of the right hippocampus in novel face encoding (Chua, et al. 2007; Sperling, et al. 2003b). Our recent longitudinal studies in aging and MCI may implicate the right hippocampus as being predictive of clinical decline (O'Brien, et al. 2010), suggesting that the right (non-dominant) hippocampus may be more vulnerable to the process of neurodegeneration.

Dice spatial overlap coefficients calculated from each individual's anatomically defined hippocampus had low to moderate values of 0.3 – 0.5, and were similar to those reported in other studies (Clement and Belleville 2009; Machielsen, et al. 2000; Rombouts, et al. 1997). These results are consistent with suggestions in the literature that statistical comparisons of magnitude of signal in activation contrasts may be more reliable than spatial voxel comparisons (Clement and Belleville 2009). As spatial overlap is partially determined by proportion of the region activated, this may have particular implications for the hippocampus, and more specifically, for our task, which primarily activates the anterior hippocampal formation. Given the marked inter-subject variability in the morphology of the hippocampus and the extent of activation, the reproducibility of fMRI signal in this region may not be best evaluated using spatial overlap methods (Rombouts, et al. 1997).

Establishing hippocampal reliability with this associative memory paradigm in older cognitively normal and mildly impaired subjects should be informative for future work investigating and quantifying cognitive decline of older subjects at risk for Alzheimer's disease. The comparison of reliability between a full associative memory paradigm and an abbreviated version of the same paradigm suggests that the short version may have sufficient reliability to be usefully incorporated into clinical trial MRI protocols. Interestingly, after separating the subjects into two groups based on their Clinical Dementia Rating score, it was observed that although the groups both showed overall moderately high signal reliability, the CDR0 group showed higher reliability than the CDR 0.5 group in the left hippocampus. However, reliability was comparable in the right hippocampus across both block and event-related designs. Our findings are generally consistent with that of Clement and Belleville (2009), who also observed overall a similar degree of reproducibility in older controls and MCI subjects in single-measure ICC across an average of multiple regions, although they did observe variability in hippocampal activation across conditions and clinical group.

Interestingly, we observed a dissociation between the reliability of behavioral memory measures and fMRI activity in comparing CDR 0 and CDR 0.5 groups. We observed that although CDR 0.5 subjects show a higher ICC for behavioral memory performance across scan sessions than CDR 0 subjects, the CDR 0.5 subjects also show decreased ICC with respect to right hippocampal activation. We speculate that these results could reflect a high between-subject variability among CDR 0.5 subjects in objective memory performance, as these individuals are known to be clinical heterogeneous. However, it is also possible that CDR 0.5 subjects may demonstrate slightly greater within-subject variability in neural activity over short time frames, reflecting evidence of early vulnerability of the neural systems engaged in memory encoding.

A secondary objective of this study was to examine the reproducibility of task-related deactivation, as the default network has received increasing interest in the aging and

neurodegenerative disease literature. The precuneus/posterior cingulate area is documented as exhibiting "beneficial" task-related deactivation, or increase of negative signal (Miller, et al. 2008a). Some studies further report a left lateralization in the areas deactivating in response to task (Binder, et al. 1999; Mazoyer, et al. 2001), a finding that was also demonstrated in the current study. While deactivation was present at both scanning sessions in the full paradigm, the locations and magnitude of deactivations varied greatly between subjects, as well as within subjects across time. One potential explanation for this is that areas of the precuneus involved in task-related deactivations may inherently demonstrate more individual regional variability in activity than the anatomically smaller and more reliably activated hippocampal formation.

It is also thought that the normal aging process disrupts coordinated intrinsic activity between different components of the default network, most prominently seen in the precuneus/posterior cingulate region of the brain (Andrews-Hanna, et al. 2007; Damoiseaux, et al. 2007; Grady, et al. 2006; Persson, et al. 2007) as well as task-induced deactivations (Sambataro, et al. 2010; Lustig, 2003). In particular, recent data suggest that the presence of amyloid pathology, even in cognitively intact older subjects, may disrupt normal default network activity during the resting state (Hedden, et al. 2009; Sheline, et al. 2010). Interestingly, older subjects with high amyloid burden, as well as cognitively normal older subjects with genetic risk factors for AD demonstrate significant reductions in deactivation during cognitive tasks, with some subjects even demonstrating paradoxical activation in this area, similar to reports in MCI and AD (Fleisher, et al. 2009; Lustig, et al. 2003; Petrella, et al. 2007; Pihlajamaki, et al. 2009; Sperling, et al. 2009). For these reasons, it may not be surprising that we observed much more variability in patterns of deactivation at each scanning session than previously seen in younger subjects (Gusnard and Raichle, 2001; Lustig et al., 2003). The results from the present study suggest that the intra- and inter-subject variability of default network activity during memory encoding, both in terms of magnitude as well as spatial extent, is higher than the variability observed in activation of the "task-positive" network in older subjects. Thus, although the default network may be a very sensitive indicator of early neural alterations seen in aging and prodromal AD, these regions may provide less reliable metrics for demonstrating pharmacologic effects.

As a large number of potential treatments for memory impairment are entering clinical trials, it is important to develop measures which can detect a "signal of efficacy" in a short time period that may predict subsequent effects over a longer trial. Our power calculations based on these results indicate that a relatively small number of subjects would be needed to detect significant pharmacological effects on hippocampal activity **within a six-week span**. These results are particularly encouraging for the abbreviated version of the paradigm, which is more feasible than the full paradigm to add on to early "proof of concept" trials or scans acquired for safety monitoring. Although a few more subjects would be needed to detect these effects in the abbreviated version of the paradigm, the relative benefits of adding such a short cognitive paradigm may outweigh the relative cost. The results of the power analyses lend further support to the promise of fMRI as a potential biomarker to detect acute efficacy in proof-of-concept clinical trials. We are currently including an abbreviated version of this paradigm in ongoing fMRI studies in the context of a placebo-controlled clinical trial in mild AD patients.

It is important to note the limitations and challenges of this study. One limitation inherent to fMRI studies is inconsistency in signal intensity due to extensive variability in individual subject's hemodynamic response and neurovascular coupling, particularly in older subjects (D'Esposito, et al. 2003; Miller, et al. 2002). We decided to include every subject from the initial dataset in our study, and did not exclude any subjects based on criteria of motion in either their functional scans or their structural images. The reason for this was to make it as

applicable to a clinical trial setting as possible, where it may not be feasible to choose only the scans that are optimal for analysis. However, because some older subjects had more head motion than typically accepted in studies of young subjects, our reproducibility analyses likely suffered by including those data points. Lastly, unlike our previous reliability study in young subjects (Sperling, et al. 2002), we did not control for over the counter medication (e.g. antihistamines), alcohol, or caffeine use, as this may not be feasible in a clinical trial. These factors may have also affected fMRI activity or cognitive performance during scan sessions. It is also important to note that we only assessed short-term test-retest reliability, which does not provide information about the utility of fMRI in assessing long-term change in the context of a potential disease-modifying medication in AD.

Strengths of this study include its focus on examining reliability in an aging and cognitively declining population, a demographic that is often overlooked in reliability studies. Additionally, all 27 of these subjects were scanned at the same place and at the same time of day to reduce potential secondary biases. We also used freely available analysis platforms that have been shown to be reliable and generalizeable for various types of data. Beyond demonstrating similar reliability between the hippocampal activations of cognitively normal and mildly impaired individuals at an inter-scan interval of 4–6 weeks, we also found that a more practical, abbreviated version of our memory paradigm demonstrated similar reliability to the full version.

CONCLUSION

In the setting of an aging population at risk for prodromal Alzheimer's disease, we demonstrated good reliability of hippocampal activity using a clinically relevant associative memory paradigm, over a time frame typically employed in early stage clinical trials. Furthermore, we demonstrated adequate reproducibility of hippocampal signal using an abbreviated form of the paradigm, which could be easily added to a standard clinical trial imaging session. The same level of reliability was not observed in default network regions, which typically demonstrate deactivation during encoding using this paradigm, which may reflect age-related variability in default network activity. These results suggest that fMRI may prove useful in evaluating the effects of interventions affecting cognitive performance in aging and early cognitive impairment over short time intervals in "Proof of Concept" clinical trials.

Acknowledgments

REFERENCES

Andrews-Hanna JR, Snyder AZ, Vincent JL, Lustig C, Head D, Raichle ME, Buckner RL. Disruption of large-scale brain systems in advanced aging. Neuron. 2007; 56(5):924–35. [PubMed: 18054866]

Aron AR, Gluck MA, Poldrack RA. Long-term test-retest reliability of functional MRI in a classification learning task. Neuroimage. 2006; 29(3):1000–6. [PubMed: 16139527]

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Rao SM, Cox RW. Conceptual processing during the conscious resting state. A functional MRI study. J Cogn Neurosci. 1999; 11(1):80–95. [PubMed: 9950716]

Bondi MW, Houston WS, Eyler LT, Brown GG. fMRI evidence of compensatory mechanisms in older adults at genetic risk for Alzheimer disease. Neurology. 2005; 64(3):501–8. [PubMed: 15699382]

Bookheimer SY, Strojwas MH, Cohen MS, Saunders AM, Pericak-Vance MA, Mazziotta JC, Small GW. Patterns of brain activation in people at risk for Alzheimer's disease. N Engl J Med. 2000; 343(7):450–6. [PubMed: 10944562]

Brewer JB, Zhao Z, Desmond JE, Glover GH, Gabrieli JD. Making memories: brain activity that predicts how well visual experience will be remembered. Science. 1998; 281(5380):1185–7. [PubMed: 9712581]

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci. 2008; 1124:1–38. [PubMed: 18400922]

Celone KA, Calhoun VD, Dickerson BC, Atri A, Chua EF, Miller SL, DePeau K, Rentz DM, Selkoe DJ, Blacker D, et al. Alterations in memory networks in mild cognitive impairment and Alzheimer's disease: an independent component analysis. J Neurosci. 2006; 26(40):10222–31. [PubMed: 17021177]

Chua EF, Schacter DL, Rand-Giovannetti E, Sperling RA. Evidence for a specific role of the anterior hippocampal region in successful associative encoding. Hippocampus. 2007

Clement F, Belleville S. Test-retest reliability of fMRI verbal episodic memory paradigms in healthy older adults and in persons with mild cognitive impairment. Hum Brain Mapp. 2009

D'Esposito M, Deouell LY, Gazzaley A. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. Nat Rev Neurosci. 2003; 4(11):863–72. [PubMed: 14595398]

Damoiseaux JS, Beckmann CF, Arigita EJ, Barkhof F, Scheltens P, Stam CJ, Smith SM, Rombouts SA. Reduced resting-state brain activity in the "default network" in normal aging. Cereb Cortex. 2007; 18(8):1856–64. [PubMed: 18063564]

Daselaar SM, Prince SE, Cabeza R. When less means more: deactivations during encoding that predict subsequent memory. Neuroimage. 2004; 23(3):921–7. [PubMed: 15528092]

Daselaar SM, Veltman DJ, Rombouts SA, Raaijmakers JG, Jonker C. Neuroanatomical correlates of episodic encoding and retrieval in young and elderly subjects. Brain. 2003; 126(Pt 1):43–56. [PubMed: 12477696]

Dickerson BC, Salat D, Greve D, Chua E, Rand-Giovannetti E, Rentz D, Bertram L, Mullin K, Tanzi R, Blacker D, et al. Increased hippocampal activation in mild cognitive impairment compared to normal aging and AD. Neurology. 2005; 65:404–411. [PubMed: 16087905]

Dickerson BC, Salat DH, Bates JF, Atiya M, Killiany RJ, Greve DN, Dale AM, Stern CE, Blacker D, Albert MS, et al. Medial temporal lobe function and structure in mild cognitive impairment. Ann Neurol. 2004; 56(1):27–35. [PubMed: 15236399]

Duverne S, Motamedinia S, Rugg MD. The relationship between aging, performance, and the neural correlates of successful memory encoding. Cereb Cortex. 2009; 19(3):733–44. [PubMed: 18653664]

Eaton KP, Szaflarski JP, Altaye M, Ball AL, Kissela BM, Banks C, Holland SK. Reliability of fMRI for studies of language in post-stroke aphasia subjects. Neuroimage. 2008; 41(2):311–22. [PubMed: 18411061]

Fleisher AS, Sherzai A, Taylor C, Langbaum JB, Chen K, Buxton RB. Resting-state BOLD networks versus task-associated functional MRI for distinguishing Alzheimer's disease risk groups. Neuroimage. 2009; 47(4):1678–90. [PubMed: 19539034]

Fransson P. How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations. Neuropsychologia. 2006; 44(14):2836–45. [PubMed: 16879844]

Golby A, Silverberg G, Race E, Gabrieli S, O'Shea J, Knierim K, Stebbins G, Gabrieli J. Memory encoding in Alzheimer's disease: an fMRI study of explicit and implicit memory. Brain. 2005; 128(Pt 4):773–87. [PubMed: 15705615]

Grady CL, Springer MV, Hongwanishkul D, McIntosh AR, Winocur G. Age-related changes in brain activity across the adult lifespan. J Cogn Neurosci. 2006; 18(2):227–41. [PubMed: 16494683]

Han SD, Houston WS, Jak AJ, Eyler LT, Nagel BJ, Fleisher AS, Brown GG, Corey-Bloom J, Salmon DP, Thal LJ, et al. Verbal paired-associate learning by APOE genotype in non-demented older adults: fMRI evidence of a right hemispheric compensatory response. Neurobiol Aging. 2007; 28(2):238–47. [PubMed: 16434125]

Hedden T, Van Dijk KR, Becker JA, Mehta A, Sperling RA, Johnson KA, Buckner RL. Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. J Neurosci. 2009; 29(40):12686–94. [PubMed: 19812343]

Johnson SC, Schmitz TW, Moritz CH, Meyerand ME, Rowley HA, Alexander AL, Hansen KW, Gleason CE, Carlsson CM, Ries ML, et al. Activation of brain regions vulnerable to Alzheimer's disease: the effect of mild cognitive impairment. Neurobiol Aging. 2006; 27(11):1604–12. [PubMed: 16226349]

Johnstone T, Somerville LH, Alexander AL, Oakes TR, Davidson RJ, Kalin NH, Whalen PJ. Stability of amygdala BOLD response to fearful faces over multiple scan sessions. Neuroimage. 2005; 25(4):1112–23. [PubMed: 15850729]

Kiehl KA, Liddle PF. Reproducibility of the hemodynamic response to auditory oddball stimuli: a six-week test-retest study. Hum Brain Mapp. 2003; 18(1):42–52. [PubMed: 12454911]

Kong J, Gollub RL, Rosman IS, Webb JM, Vangel MG, Kirsch I, Kaptchuk TJ. Brain activity associated with expectancy-enhanced placebo analgesia as measured by functional magnetic resonance imaging. J Neurosci. 2006; 26(2):381–8. [PubMed: 16407533]

Liou M, Su HR, Lee JD, Cheng PE, Huang CC, Tsai CH. Functional MR images and scientific inference: reproducibility maps. J Cogn Neurosci. 2003; 15(7):935–45. [PubMed: 14628755]

Lustig C, Snyder AZ, Bhakta M, O'Brien KC, McAvoy M, Raichle ME, Morris JC, Buckner RL. Functional deactivations: change with age and dementia of the Alzheimer type. Proc Natl Acad Sci U S A. 2003; 100(24):14504–9. [PubMed: 14608034]

MacDonald SW, Li SC, Backman L. Neural underpinnings of within-person variability in cognitive functioning. Psychol Aging. 2009; 24(4):792–808. [PubMed: 20025396]

Machielsen WC, Rombouts SA, Barkhof F, Scheltens P, Witter MP. FMRI of visual encoding: reproducibility of activation. Hum Brain Mapp. 2000; 9(3):156–64. [PubMed: 10739366]

Machulda MM, Ward HA, Borowski B, Gunter JL, Cha RH, O'Brien PC, Petersen RC, Boeve BF, Knopman D, Tang-Wai DF, et al. Comparison of memory fMRI response among normal, MCI, and Alzheimer's patients. Neurology. 2003; 61(4):500–6. [PubMed: 12939424]

Mandzia JL, McAndrews MP, Grady CL, Graham SJ, Black SE. Neural correlates of incidental memory in mild cognitive impairment: an fMRI study. Neurobiol Aging. 2009; 30(5):717–30. [PubMed: 17963998]

Manoach DS, Halpern EF, Kramer TS, Chang Y, Goff DC, Rauch SL, Kennedy DN, Gollub RL. Test-retest reliability of a functional MRI working memory paradigm in normal and schizophrenic subjects. Am J Psychiatry. 2001; 158(6):955–8. [PubMed: 11384907]

Mazoyer B, Zago L, Mellet E, Bricogne S, Etard O, Houde O, Crivello F, Joliot M, Petit L, Tzourio-Mazoyer N. Cortical networks for working memory and executive functions sustain the conscious resting state in man. Brain Res Bull. 2001; 54(3):287–98. [PubMed: 11287133]

McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. J Cogn Neurosci. 2003; 15(3):394–408. [PubMed: 12729491]

Miller MB, Van Horn JD, Wolford GL, Handy TC, Valsangkar-Smyth M, Inati S, Grafton S, Gazzaniga MS. Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. J Cogn Neurosci. 2002; 14(8):1200–14. [PubMed: 12495526]

Miller SL, Celone K, DePeau K, Diamond E, Dickerson BC, Rentz D, Pihlajamaki M, Sperling RA. Age-related memory impairment associated with loss of parietal deactivation but preserved hippocampal activation. Proc Natl Acad Sci U S A. 2008a; 105(6):2181–6. [PubMed: 18238903]

Miller SL, Fenstermacher E, Bates J, Blacker D, Sperling RA, Dickerson BC. Hippocampal activation in adults with mild cognitive impairment predicts subsequent cognitive decline. J Neurol Neurosurg Psychiatry. 2008b; 79(6):630–5. [PubMed: 17846109]

O'Brien JL, O'Keefe KM, Laviolette PS, Deluca AN, Blacker D, Dickerson BC, Sperling RA. Longitudinal fMRI in elderly reveals loss of hippocampal activation with clinical decline. Neurology. 2010

Persson J, Lustig C, Nelson JK, Reuter-Lorenz PA. Age differences in deactivation: a link to cognitive control? J Cogn Neurosci. 2007; 19(6):1021–32. [PubMed: 17536972]

Petrella JR, Wang L, Krishnan S, Slavin MJ, Prince SE, Tran TT, Doraiswamy PM. Cortical deactivation in mild cognitive impairment: high-field-strength functional MR imaging. Radiology. 2007; 245(1):224–35. [PubMed: 17885190]

Pihlajamaki M, DePeau KM, Blacker D, Sperling RA. Impaired medial temporal repetition suppression is related to failure of parietal deactivation in Alzheimer disease. Am J Geriatr Psychiatry. 2008; 16(4):283–92. [PubMed: 18378553]

Pihlajamaki M, O'Keefe K, Bertram L, Tanzi RE, Dickerson BC, Blacker D, Albert MS, Sperling RA. Evidence of Altered Posteromedial Cortical fMRI Activity in Subjects at Risk for Alzheimer Disease. Alzheimer Dis Assoc Disord. 2009

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. Proc Natl Acad Sci U S A. 2001; 98(2):676–82. [PubMed: 11209064]

Remy F, Mirrashed F, Campbell B, Richter W. Verbal episodic memory impairment in Alzheimer's disease: a combined structural and functional MRI study. Neuroimage. 2005; 25(1):253–66. [PubMed: 15734360]

Rombouts SA, Barkhof F, Hoogenraad FG, Sprenger M, Valk J, Scheltens P. Test-retest analysis with functional MR of the activated area in the human visual cortex. AJNR Am J Neuroradiol. 1997; 18(7):1317–22. [PubMed: 9282862]

Rutten GJ, Ramsey NF, van Rijen PC, van Veelen CW. Reproducibility of fMRI-determined language lateralization in individual subjects. Brain Lang. 2002; 80(3):421–37. [PubMed: 11896650]

Sambataro F, Murty VP, Callicott JH, Tan HY, Das S, Weinberger DR, Mattay VS. Age-related alterations in default mode network: impact on working memory performance. Neurobiol Aging. 2010; 31(5):839–52. [PubMed: 18674847]

Schacter DL, Alpert NM, Savage CR, Rauch SL, Albert MS. Conscious recollection and the human hippocampal formation: evidence from positron emission tomography. Proc Natl Acad Sci U S A. 1996; 93(1):321–5. [PubMed: 8552630]

Sheline YI, Raichle ME, Snyder AZ, Morris JC, Head D, Wang S, Mintun MA. Amyloid plaques disrupt resting state default mode network connectivity in cognitively normal elderly. Biol Psychiatry. 2010; 67(6):584–7. [PubMed: 19833321]

Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing inter-rater reliability. Psychological Bulletin. 1979; 86(2):420–428. [PubMed: 18839484]

Specht K, Willmes K, Shah NJ, Jancke L. Assessment of reliability in functional imaging studies. J Magn Reson Imaging. 2003; 17(4):463–71. [PubMed: 12655586]

Sperling R, Chua E, Cocchiarella A, Rand-Giovannetti E, Poldrack R, Schacter DL, Albert M. Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. Neuroimage. 2003a; 20(2):1400–10. [PubMed: 14568509]

Sperling R, Greve D, Dale A, Killiany R, Holmes J, Rosas HD, Cocchiarella A, Firth P, Rosen B, Lake S, et al. Functional MRI detection of pharmacologically induced memory impairment. Proc Natl Acad Sci U S A. 2002; 99(1):455–60. [PubMed: 11756667]

Sperling RA, Bates J, Cocchiarella A, Schacter D, Rosen B, Albert M. Encoding Novel Face-Name Associations: A Functional MRI Study. Human Brain Mapping. 2001; 14:129–139. [PubMed: 11559958]

Sperling RA, Bates JF, Chua EF, Cocchiarella AJ, Rentz DM, Rosen BR, Schacter DL, Albert MS. fMRI studies of associative encoding in young and elderly controls and mild Alzheimer's disease. J Neurol Neurosurg Psychiatry. 2003b; 74(1):44–50. [PubMed: 12486265]

Sperling RA, Laviolette PS, O'Keefe K, O'Brien J, Rentz DM, Pihlajamaki M, Marshall G, Hyman BT, Selkoe DJ, Hedden T, et al. Amyloid deposition is associated with impaired default network function in older persons without dementia. Neuron. 2009; 63(2):178–88. [PubMed: 19640477]

Squire LR, Ojemann JG, Miezin FM, Petersen SE, Videen TO, Raichle ME. Activation of the hippocampus in normal humans: a functional anatomical study of memory. Proc Natl Acad Sci U S A. 1992; 89(5):1837–41. [PubMed: 1542680]

Stern CE, Corkin S, Gonzalez RG, Guimaraes AR, Baker JR, Jennings PJ, Carr CA, Sugiura RM, Vedantham V, Rosen BR. The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. Proc Natl Acad Sci U S A. 1996; 93(16): 8660–5. [PubMed: 8710927]

Trivedi MA, Schmitz TW, Ries ML, Torgerson BM, Sager MA, Hermann BP, Asthana S, Johnson SC. Reduced hippocampal activation during episodic encoding in middle-aged individuals at genetic risk of Alzheimer's disease: a cross-sectional study. BMC Med. 2006; 4:1. [PubMed: 16412236]

Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science. 1998; 281(5380):1188–91. [PubMed: 9712582]

Wei X, Yoo SS, Dickey CC, Zou KH, Guttmann CR, Panych LP. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. Neuroimage. 2004; 21(3):1000–8. [PubMed: 15006667]

Zandbelt BB, Gladwin TE, Raemaekers M, van Buuren M, Neggers SF, Kahn RS, Ramsey NF, Vink M. Within-subject variation in BOLD-fMRI signal changes across repeated measurements: quantification and implications for sample size. Neuroimage. 2008; 42(1):196–206. [PubMed: 18538585]

Zelinski EM, Gilewski MJ. Assessment of memory complaints by rating scales and questionnaires. Psychopharmacol Bull. 1988; 24(4):523–9. [PubMed: 3074313]
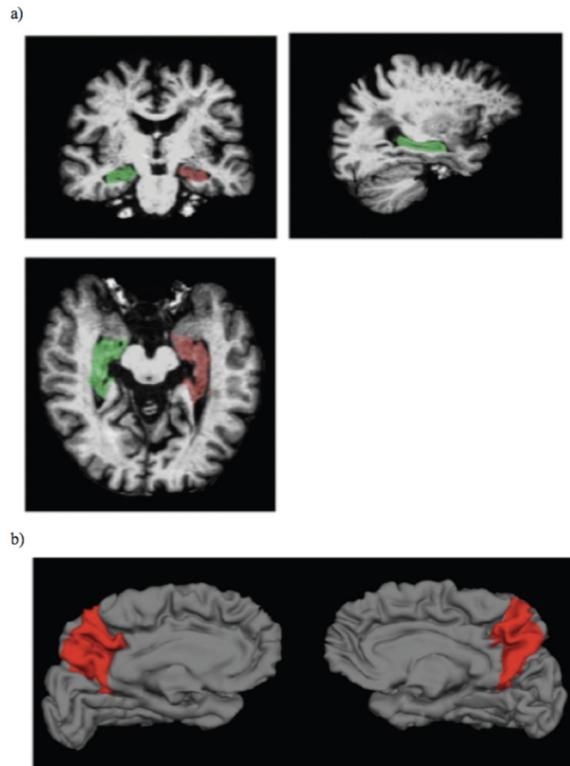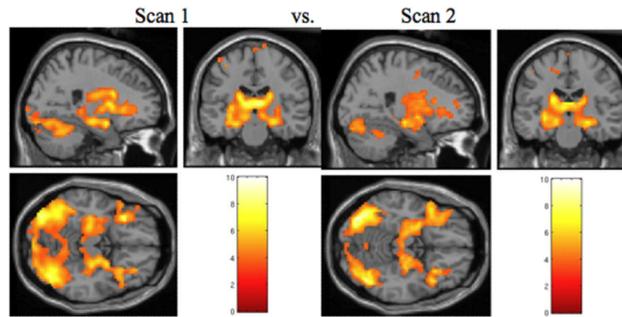
Figure 1.
An individual subject's left and right a) hippocampus and b) precuneus, as defined by
FreeSurfer.

Full Paradigm

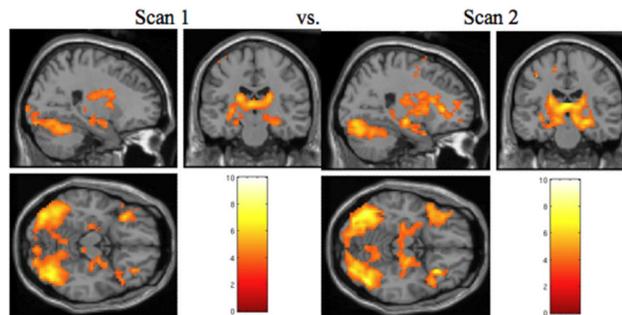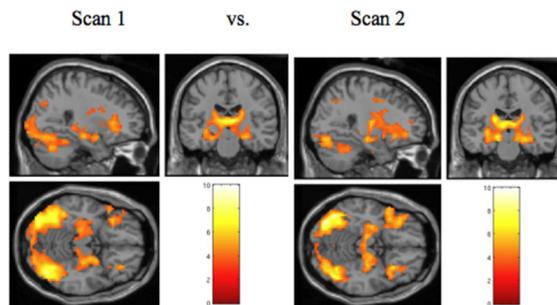
Abbreviated Paradigm

Figure 2. Block analysis- Novel versus Repeated face-name pairs
Whole Brain SPM2 maps of 27 subjects with foci at [24 −18 −15], thresholded at p< 0.001, extent = 5 contiguous voxels. Similar patterns of whole brain and hippocampal activation are observed between Scan Session 1 and Scan Session 2, and also between the full and abbreviated paradigms.

Full Paradigm

Scan 1 vs. Scan 2

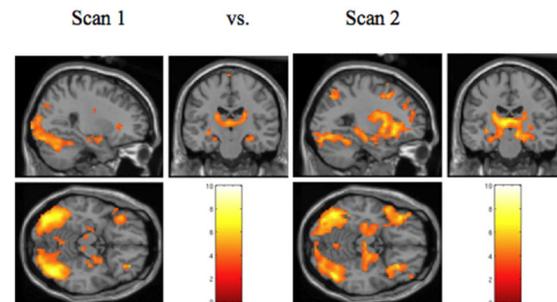**Abbreviated Paradigm**

Scan 1 vs. Scan 2

Figure 3. Event-related analysis- High-confidence Hits versus Repeated face-name pairs
Whole Brain SPM2 maps of 27 subjects with foci at [27 −18 −15], thresholded at p< 0.001, extent = 5 contiguous voxels. Again, similar patterns of whole brain, and specifically hippocampal, activation are observed between Scan Session 1 and Scan Session 2. These patterns also appear consistent between the full and abbreviated paradigms.
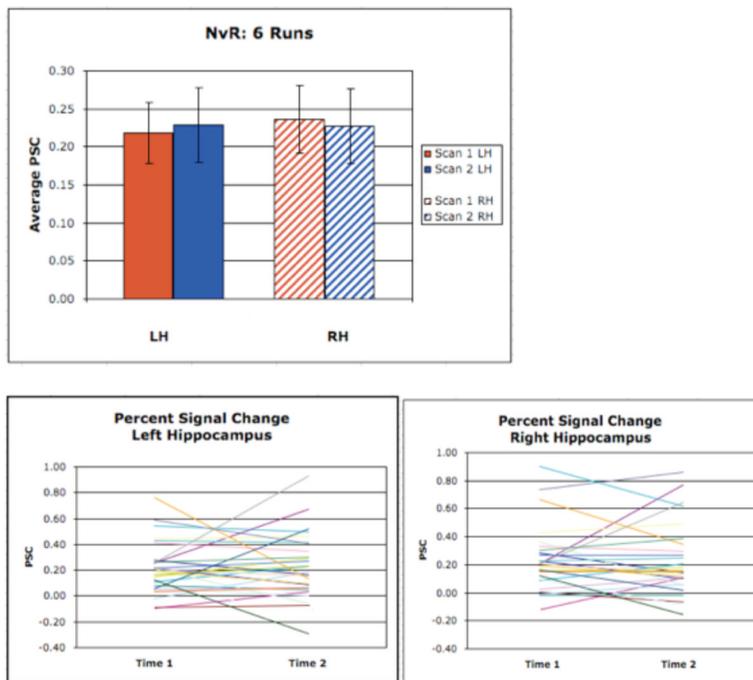
Figure 4. Hippocampal ROI results

Contrast of novel versus repeated face-name pairs extracted from the full memory paradigm. Percent signal change (PSC) was extracted from individual hippocampi and compared across scan session. PSC is very similar between scan sessions for both the left and the right hippocampi at the group level. "Spaghetti" plots (one line for each subject) show individual magnitudes of activation from Scan Session 1 to Scan Session 2.
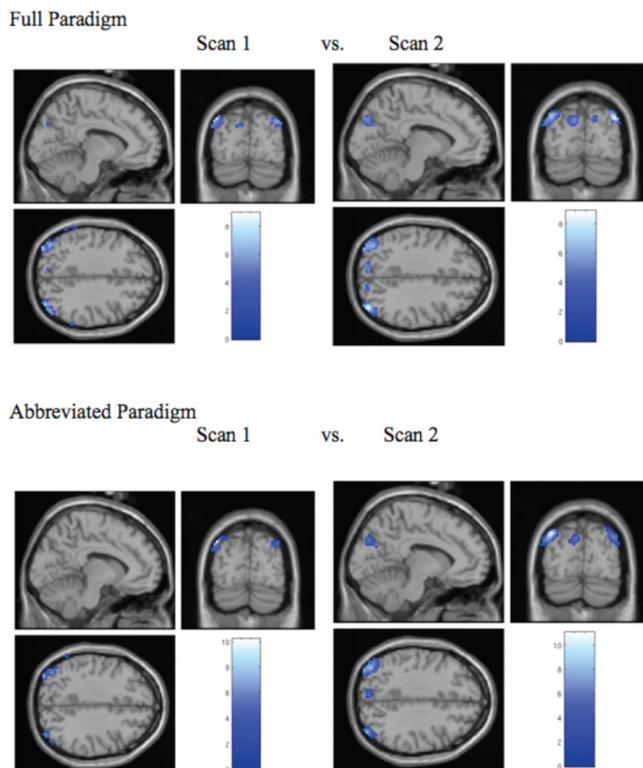
Full Paradigm

Abbreviated Paradigm

Figure 5. Event-related analysis- Fixation versus All Novel and Repeated face-name pairs
Whole Brain SPM2 maps of 27 subjects with foci at [−9 −81 36], thresholded at p< 0.001, extent = 5 contiguous voxels. Similar patterns of whole brain deactivation are observed between Scan Session 1 and Scan Session 2, with subjects showing more deactivation at Scan Session 2 than at Scan Session 1 in the precuneus. These patterns also appear comparable between the full and abbreviated paradigm at Scan Session 2, but not so at Scan Session 1.
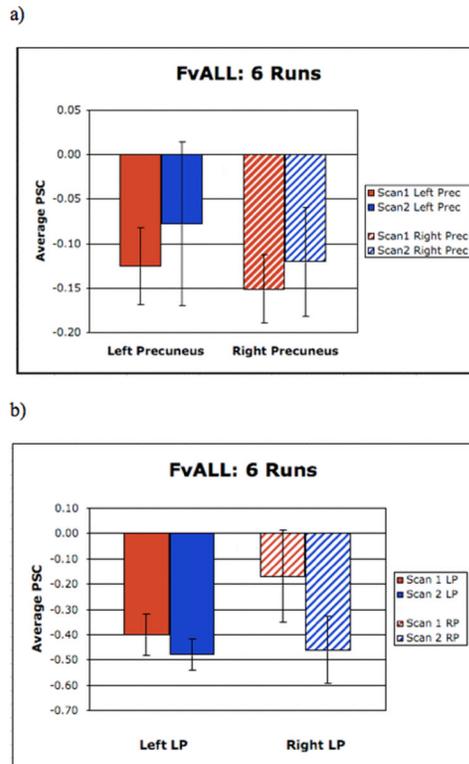
a)

b)

Figure 6. Precuneus and Lateral Parietal ROI results: Event-related analysis-Fixation versus All Novel and Repeated face-name pairs
Percent signal change was extracted from the full memory paradigm from individual a) precuneus ROIs and b) lateral parietal ROIs, and compared across scan session. Note that there is considerable variability in deactivation signal.

Table I

Subject Demographics Statistics for numeric variables are mean and standard deviation.

	ALL	Groups	
	ALL	CDR 0	CDR 0.5
N	27	18	9
M/F	10/17	4/14[*]	6/3
Age(yrs)	72.4 ± 10.1	73.8 ± 6.2	69.4 ± 15.4
CDR SOB	--	0.0 ± 0.0	1.40 ± 1.42
Education (yrs)	17.6 ± 2.8	16.9 ± 3.17	18.7 ± 1.6
MMSE	29.1 ± 1.0	29.2 ± 1.0	28.8 ± 1.1

[*] Fisher's exact test, p<0.05

Table II

Behavioral Performance

	Groups		
	ALL	**CDR 0**	**CDR 0.5**
	Average Percentage of Trial (SE)		
HC- Hits			
Scan 1	38.45 (3.33)	44.71 (3.55) **	25.93 (5.06)
Scan 2	37.08 (3.35)	41.79 (3.59) *	27.64 (6.18)
LC - Hits			
Scan 1	26.06 (2.30)	23.02 (2.57)	32.14 (4.09)
Scan 2	28.44 (2.87)	27.25 (3.36)	30.82 (5.61)
HC- Miss			
Scan 1	13.14 (1.65)	13.94 (1.86)	12.83 (3.46)
Scan 2	13.18 (1.66)	14.15 (2.18)	12.69 (2.57)
LC- Miss			
Scan 1	20.63 (2.11)	17.19 (1.97) *	27.51 (4.26)
Scan 2	19.97 (1.74)	16.21 (1.30) **	27.51 (3.41)

Independent samples *t*-test between CDR groups.

*
p<0.05,

**
p<0.005.

watermark-text $watermark-text $watermark-text

Table III

Intraclass Correlation Coefficients from the Hippocampus Reliability of percent signal change across two time points assessed four-six weeks apart.

	ICC: individual (average)					
	ALL		CDR 0		CDR 0.5	
	6 Runs	**2 Runs**	**6 Runs**	**2 Runs**	**6 Runs**	**2 Runs**
NvR						
LH	0.40 (0.57)	0.34 (0.50)	0.72 (0.83)	0.51 (0.61)	0.10 (0.18)	0.13 (0.23)
RH	0.66 (0.80)	0.64 (0.78)	0.64 (0.80)	0.60 (0.75)	0.69 (0.82)	0.70 (0.82)
NvF						
LH	0.38 (0.55)	0.40 (0.57)	0.54 (0.70)	0.39 (0.56)	0.26 (0.41)	0.39 (0.57)
RH	0.56 (0.72)	0.56 (0.73)	0.48 (0.65)	0.51 (0.73)	0.63 (0.77)	0.64 (0.78)
HCHvR						
LH	0.39 (0.56)	0.46 (0.63)	0.82 (0.94)	0.57 (0.72)	0.15 (0.26)	0.37 (0.54)
RH	0.71 (0.83)	0.70 (0.82)	0.81 (0.89)	0.70 (0.83)	0.63 (0.77)	0.70 (0.82)
HCHvF						
LH	0.38 (0.56)	0.43 (0.58)	0.72 (0.83)	0.46 (0.63)	0.26 (0.41)	0.41 (0.58)
RH	0.65 (0.78)	0.45 (0.62)	0.71 (0.83)	0.51 (0.67)	0.63 (0.77)	0.39 (0.56)