BRAIN
A JOURNAL OF NEUROLOGY

Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease*

Rahul S. Desikan,[1,2] Howard J. Cabral,[3] Christopher P. Hess,[4] William P. Dillon,[4] Christine M. Glastonbury,[4] Michael W. Weiner,[4,5] Nicholas J. Schmansky,[1] Douglas N. Greve,[1] David H. Salat,[1] Randy L. Buckner[1,6,7,10] and Bruce Fischl[1,8,9]; Alzheimer's Disease Neuroimaging Initiative

1 Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA
2 Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA
3 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
4 Department of Radiology, University of California, San Francisco, CA, USA
5 Veteran Affairs Medical Center, San Francisco, CA, USA
6 Department of Psychology, Harvard University, Cambridge, MA, USA
7 Howard Hughes Medical Institute, Chevy Chase, MD, USA
8 Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA
9 Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA
10 Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

Correspondence to: Bruce Fischl, PhD,
Massachusetts General Hospital,
NMR Center Rm. 2301,
Building 149, 13th Street,
Charlestown, MA 02129,
USA
Email: fischl@nmr.mgh.harvard.edu

Mild cognitive impairment can represent a transitional state between normal ageing and Alzheimer's disease. Non-invasive diagnostic methods are needed to identify mild cognitive impairment individuals for early therapeutic interventions. Our objective was to determine whether automated magnetic resonance imaging-based measures could identify mild cognitive impairment individuals with a high degree of accuracy. Baseline volumetric T_1-weighted magnetic resonance imaging scans of 313 individuals from two independent cohorts were examined using automated software tools to identify the volume and mean thickness of 34 neuroanatomic regions. The first cohort included 49 older controls and 48 individuals with mild cognitive impairment, while the second cohort included 94 older controls and 57 mild cognitive impairment individuals. Sixty-five patients with probable Alzheimer's disease were also included for comparison. For the discrimination of mild cognitive impairment, entorhinal cortex thickness, hippocampal volume and supramarginal gyrus thickness demonstrated an area under the curve of 0.91 (specificity 94%, sensitivity 74%, positive likelihood ratio 12.12, negative likelihood ratio 0.29) for the first cohort and an

area under the curve of 0.95 (specificity 91%, sensitivity 90%, positive likelihood ratio 10.0, negative likelihood ratio 0.11) for the second cohort. For the discrimination of Alzheimer's disease, these three measures demonstrated an area under the curve of 1.0. The three magnetic resonance imaging measures demonstrated significant correlations with clinical and neuropsychological assessments as well as with cerebrospinal fluid levels of tau, hyperphosphorylated tau and abeta 42 proteins. These results demonstrate that automated magnetic resonance imaging measures can serve as an *in vivo* surrogate for disease severity, underlying neuropathology and as a non-invasive diagnostic method for mild cognitive impairment and Alzheimer's disease.

Introduction

Mild cognitive impairment (MCI) represents a transitional period between normal ageing and clinically probable Alzheimer's disease (Petersen *et al.*, 2001). Individuals classified with the amnestic subtype of MCI are likely in the prodromal stage of Alzheimer's disease (Morris *et al.*, 2001) and up to 80% of these individuals progress to dementia after 6 years (Petersen *et al.*, 1999). Overall, the prevalence of MCI in the elderly is 19%, with estimates as high as 29% amongst individuals greater than 85 years of age (Lopez *et al.*, 2003). The neuropathologic changes of this transitional state are consistent with the density and distribution of the tau-associated neurofibrillary and abeta-associated amyloid features of very early Alzheimer's disease (Bennett *et al.*, 2005) and present long before the clinical onset of probable Alzheimer's disease (Gomez-Isla *et al.*, 1996). As therapeutic interventions become available, there is a need for developing methodologies that will serve as an *in vivo* surrogate for these pathologic changes, and thus, accurately identify those cognitively impaired individuals who are in the earliest stages of Alzheimer's disease.

Structural MRI provides visualization of the macroscopic tissue atrophy that results from the cellular changes underlying Alzheimer's disease. In order to be used as a diagnostic marker, structural MRI measures should: (i) specifically detect and quantify fundamental features of Alzheimer's pathology in individuals at an elevated risk for Alzheimer's disease (i.e. individuals with amnestic MCI) and in patients with a clinical diagnosis of Alzheimer's disease; (ii) demonstrate excellent discrimination accuracy between normal elderly controls and individuals with MCI and Alzheimer's disease; (iii) exhibit a high degree of consistency and test–retest reproducibility across multiple, independent cohorts; and (iv) correlate strongly with clinical measures of decline as well as invasive measures of cellular pathology.

Prior structural MRI studies have employed either manual region of interest (ROI) (Killiany *et al.*, 2000; Xu *et al.*, 2000; Devanand *et al.*, 2007) or automated whole-brain approaches (Scahill *et al.*, 2002; Buckner *et al.*, 2005; Dickerson *et al.*, 2009) to identify MCI and Alzheimer's disease individuals. Though these methodologies offer several strengths, they are limited in their use as a diagnostic marker due to variable discrimination accuracy and decreased test–retest reliability with the manual ROI methods, and an inability to evaluate the disease state in a single individual with the whole-brain approaches.

Recent advances in image analysis algorithms have led to the development of structural MRI-based software tools that can automatically parcellate the brain into anatomic regions and quantify the tissue atrophy in these regions for a single individual (Fischl *et al.*, 2002; Desikan *et al.*, 2006). In this study, we investigated the feasibility of utilizing these automated software tools as a diagnostic marker for Alzheimer's disease. Using structural MRI scans from a cohort of 97 participants, we first identified a set of anatomic regions that best differentiated MCI individuals from elderly controls and examined the discrimination accuracy of these regions. We then validated the accuracy and consistency of these measures on a second, independent cohort of 216 participants. Finally, we examined the relationship between these MRI-based anatomic measures and clinical measures of decline and cerebrospinal fluid (CSF) markers of cellular pathology.

Methods

Overview

A total of 313 individuals were examined in this study. The first cohort of 97 participants ('training cohort') was selected from the Open Access Series of Imaging Studies (OASIS) database (Marcus *et al.*, 2007). Informed consent for all participants was obtained in accordance with guidelines of the Washington University Human Studies Committee (St Louis, MO). Data from subsets of these participants have been published in previous studies (Salat *et al.*, 2004; Buckner *et al.*, 2004, 2005; Head *et al.*, 2005; Dickerson *et al.*, 2009). The OASIS dataset reflects a collaborative effort of investigators from a single acquisition site supported by the National Institute on Aging (NIA), the Howard Hughes Medical Institute, the Biomedical Informatics Research Network (BIRN) and the Washington University Alzheimer's Disease Research Center [Alzheimer's Disease Research Center (ADRC)]. The dataset includes multiple (2–4) structural MRI acquisitions from 416 adults, ages 18–96. For more information, please see http://www.oasis-brains.org.

The second cohort of 216 participants ('validation cohort') was selected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). The ADNI is a large multi-site collaborative effort launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organizations as a public–private partnership aimed at testing whether serial MRI, PET, other biological markers and clinical and neuropsychological assessment can be

combined to measure the progression of MCI and early Alzheimer's disease. The Principal Investigator of this initiative is Michael Weiner, MD, and ADNI is the result of many co-investigators from a broad range of academic institutions and private corporations, with subjects recruited from over 50 sites across the United States and Canada. For more information, please see http://www.adni-info.org.

Cohort 1—training cohort

Potential participants underwent a multi-stage screening procedure that has been described in detail elsewhere (Rubin et al., 1998). Briefly, eligibility requirements included living independently in the community, having an informant who could provide information about the participants' daily function, and absence of significant underlying medical, neurologic or psychiatric illness. The degree of clinical severity was evaluated by an annual semi-structured interview. This interview generates both an overall Clinical Dementia Rating (CDR) score and a measure known as the CDR Sum of Boxes (CDR-SB) (Morris, 1993). Experienced clinicians conducted independent semi-structured interviews with the participant and a knowledgeable collateral source that included a health history and neurological examination. The Mini-Mental State Examination (MMSE) (Folstein et al., 1975) and a complete neuropsychological battery (Rubin et al., 1998) were also conducted.

Participants were selected from the OASIS database if they were clinically classified as: (i) older controls (OCs) ($n = 49$)—individuals who were cognitively normal (CDR 0) or (ii) estimated to be MCI ($n = 48$)—individuals with a memory complaint who experienced very mild cognitive decline and a CDR of 0.5. Note that a CDR score of 0.5 is not synonymous with a diagnosis of amnestic-type MCI and slight differences in the sample may exist, although the two groups are expected to largely be similar and capture individuals at the transitional state between normal cognition and dementia. The mean age, gender and mean MMSE, and mean CDR-SB scores are shown in Table 1. As expected, the CDR-SB and MMSE showed a difference between the groups ($P < 0.05$). No other demographic variables differed between the groups.

Confirming our expectation that the samples are similar, based on their CDR-SB scores, the individuals in this estimated MCI group are comparable to amnestic MCI subjects used in epidemiological studies and clinical trials (Davis and Rockwood, 2004; Petersen et al., 2005). Prior work using a larger cohort of the above-described mildly impaired individuals has demonstrated that amnestic MCI represents the prodromal stage of Alzheimer's disease (Morris et al., 2001). For simplicity, we use MCI in this article to refer to the entire group of cognitively impaired subjects who did not meet criteria for probable Alzheimer's disease (CDR = 1.0).

Cohort 2—validation cohort

Each participant was selected using eligibility criteria that are described in detail elsewhere (http://www.adni-info.org/index.php?option= com_content&task=view&id=9&Itemid=43). Briefly, experienced clinicians conducted independent semi-structured interviews with the participant and a knowledgeable collateral source that included a health history, neurological examination, the MMSE (Folstein et al., 1975), the CDR-Sum of Boxes (Morris, 1993) and a comprehensive neuropsychological battery.

Participants were selected from the ADNI database if they were clinically classified as: (i) OCs ($n = 94$)—individuals who were cognitively normal (CDR 0) or (ii) MCI ($n = 57$)—individuals with MMSE scores between 24 and 30, a subjective memory complaint verified by an informant, objective memory loss as measured by education-adjusted performance on the Logical Memory II subscale (delayed paragraph recall) of the Wechsler Memory Scale-Revised (Wechsler, 1987), a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living and an absence of dementia at the time of the baseline MRI scan who within 2 years progressed to a diagnosis of probable Alzheimer's disease (CDR = 1.0). Only individuals classified as the amnestic subtype of MCI, based on the revised MCI criteria (Petersen, 2004), were selected. In addition, 65 individuals who met criteria for probable Alzheimer's disease (McKhann et al., 1984) (all were CDR 1) were also included for comparison. As expected, the CDR-SB and MMSE showed a difference between the groups ($P < 0.05$). No other demographic variables differed between the groups.

MRI image acquisition

For the training cohort (OASIS subjects), the MRI scans were acquired on a 1.5T Vision system (Siemens, Erlangen, Germany). T_1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) scans were obtained according to the following protocol: two sagittal acquisitions, FOV = 224, matrix = 256×256, resolution = $1 \times 1 \times 1.25$ mm^3, TR = 9.7 ms, TE = 4 ms, flip angle = $10°$, TI = 20 ms, TD = 200 ms (Marcus et al., 2007). Two acquisitions were averaged together to increase the contrast-to-noise ratio.

For the validation cohort (ADNI subjects), the MRI scans were acquired at multiple sites using either a GE, Siemens or Philips 1.5T system. Two high-resolution T_1- weighted volumetric MP-RAGE scans were collected for each subject and the raw DICOM images were downloaded from the public ADNI site (http://www.loni.ucla.edu/ADNI/Data/index.shtml). Parameter values vary depending on scanning site and can be found at http://www.loni.ucla.edu/ADNI/Research/Cores/.

Table 1 Descriptive statistical information for the subjects in the study

Diagnostic group	Training cohort (OASIS subjects)		Validation cohort (ADNI subjects)		
	OC	MCI	OC	MCI	Alzheimer's disease
Sample size	49	48	94	57	65
Age	76.6 (4.9)	78.0 (5.6)	76.0 (5.0)	76.4 (6.1)	76.6 (7.7)
Percent female	65%	60%	52%	40%	56%
MMSE	29.4 (0.8)	25.9 (2.9)	29.2 (1.0)	26.7 (1.7)	22.5 (2.0)
CDR-SB	0.0 (0.1)	2.8 (1.0)	0.0 (0.0)	1.5 (0.8)	5.7 (1.2)

Means are listed with standard deviations in parentheses. OC = Older controls; MCI = individuals with mild cognitive impairment; Alzheimer's disease = individuals with clinically diagnosed Alzheimer's disease; OASIS = Open Access Series of Imaging Studies; ADNI = Alzheimer's Disease Neuroimaging Initiative.

Automated image analysis procedures

All MRI scans were processed using the FreeSurfer software package, freely available at http://surfer.nmr.mgh.harvard.edu. Multiple MPRAGE MRI acquisitions for each participant were motion corrected, averaged and normalized for intensity inhomogeneities to create a single image volume with relatively high contrast to noise (Dale *et al*., 1999). This averaged volume was used to locate the grey/white matter boundary (white matter surface) and this, in turn, was then used to locate the grey/CSF boundary (grey matter surface) (Fischl *et al*., 1999a; 2000). Cortical thickness measurements were then obtained by calculating the distance between the grey and the white matter surfaces at each point (per hemisphere) across the entire cortical mantle (Fischl *et al*., 2000). This cortical thickness measurement technique has been validated via histological (Rosas *et al*., 2002) as well as manual measurements (Salat *et al*., 2004; Dickerson *et al*., 2009). The reliability of the cortical thickness measures as well as the other image analysis procedures presented here has been demonstrated across different manufacturer types, scanner upgrades, varying contrast-to-noise ratio, and the number of MPRAGE MRI acquisitions used (Han *et al*., 2006; Fennema-Notestine *et al*., 2007; Jovicich *et al*., 2009).

The neocortex of the brain on the MRI scans was then automatically subdivided into 32 gyral-based ROIs (in each hemisphere). To accomplish this, a registration procedure was used that aligns the cortical folding patterns (Fischl *et al*., 1999b) and probabilistically assigns every point on the cortical surface to one of the 32 ROIs (Desikan *et al*., 2006). In addition, two non-neocortical regions of the brain, namely the amygdala and the hippocampus, were automatically delineated using an algorithm that examines variations in voxel intensities and spatial relationships to classify non-neocortical regions on MRI scans (Fischl *et al*., 2002).

The anatomic accuracy of the grey and white matter surfaces as well as each of the individual ROIs was carefully reviewed by a trained neuroanatomist (RSD), with particular attention to the medial temporal lobe where non-brain tissue, such as dura mater and temporal bone, often needs to be excluded. All of the MRI scans were processed on a Linux cluster machine with 230 nodes, each with a 2 GHz AMD Opteron CPU (Advanced Micro Devices, Sunnyvale, CA, USA) and 4 GB RAM. Processing time for each MRI scan was ~25–40 h. The cluster machine allows for the processing of 230 MRI scans simultaneously.

In total, 34 neocortical and non-necortical ROIs were used in this study. For all of the analyses performed here, the mean thickness (only neocortical regions) and the volume (both neocortical and non-neocortical regions) of the right and the left hemispheres, for each ROI, were added together. In order to account for differences in head size, the total volume for each ROI was corrected using a previously validated estimate of the total intracranial volume (eTIV) (Buckner *et al*., 2004). Figure 1 shows all of the ROIs used in this study.

Neuropsychological measures

All ADNI subjects were administered a neuropsychological battery. Two test scores were selected for analysis in the present study because they had previously been shown to be sensitive predictors of progression from MCI to Alzheimer's disease (Estevez-Gonzalez *et al*., 2003; Blacker *et al*., 2007). These two test scores were 5 min and 30 min recall from the Rey Auditory Verbal Learning Test (AVLT) (Lezak, 1995), and time to complete part B of the Trail Making Test (Trails B) (Reitan, 1958).

Biomarker samples

From the current sample, a number of ADNI subjects ($n = 115$) underwent lumbar puncture for CSF biomarker evaluation. Three protein samples were selected for analysis in the present study because they had previously been shown to be sensitive predictors of progression from MCI to Alzheimer's disease (Hampel *et al*., 2008). These included plasma samples of tau, abeta 42 and hyperphosphorylated tau (p-tau).

Statistical analysis

A series of logistic regression models were applied on the training cohort ROI data to identify those automated measures that best discriminated the MCI individuals from the OCs. Age was included as a covariate in each of the regression models. For each measure, the *P*-value, odds ratio (OR) computed for a 1 SD difference in the predictor and area under the curve (AUC) was computed. In these analyses, the AUC functionally combines the sensitivity and specificity of the regression analyses in classifying subjects as either MCI or OCs. The value for the AUC varies from 0.5 representing no discrimination to 1.0 representing perfect discrimination (Nam *et al*., 2002).

For the training cohort, the thickness and estimate of the total intracranial volume-corrected volumes for each of the ROIs (total of 66 measures) were first entered into simple logistic regression models and only those measures that demonstrated an AUC of 0.70 or better were retained for further analysis. These retained ROI measures were entered into a second logistic regression model but only those that demonstrated an independent effect equal to an OR of 0.57 or lower (equivalent to a 75% increase in risk with smaller volume or decreased thickness) were selected. The resulting ROI measures were retained for inclusion into a final multiple regression model and the discrimination accuracy for the training cohort was derived. For the validation cohort, only those ROI measures that best discriminated the MCI group in the training cohort were entered into a multiple regression model and the discrimination accuracy was derived. In order to assess consistency and reproducibility, the logistic regression coefficients from the final model (developed from the training cohort) were applied to the validation cohort and the AUC was calculated.

Correlation coefficients were used to examine the relationship between the automated MRI measures that best discriminated the MCI group and clinical, neuropsychological and CSF biomarker evaluations in the validation cohort (ADNI subjects). Spearman's rank correlation coefficients were utilized in order to avoid making assumptions about the statistical distributions of the variables.

Results

The simple logistic regressions on the training cohort (OASIS subjects) revealed significant effects for entorhinal cortex thickness (AUC = 0.86, $P < 0.00001$) and volume (AUC = 0.80, $P < 0.0001$), inferior parietal lobule thickness (AUC = 0.71, $P < 0.0001$) and volume (AUC = 0.70, $P < 0.0001$), inferior temporal gyrus thickness (AUC = 0.72, $P < 0.0001$) and volume (AUC = 0.72, $P < 0.0001$), isthmus of cingulate cortex thickness (AUC = 0.71, $P < 0.001$), lateral occipital cortex thickness (AUC = 0.72, $P < 0.0001$), lingual cortex thickness (AUC = 0.71, $P < 0.0001$) and volume (AUC = 0.71, $P < 0.0001$), middle temporal gyrus thickness (AUC = 0.72, $P < 0.0001$) and volume (AUC = 0.72, $P < 0.0001$), parahippocampal gyrus (AUC = 0.74, $P < 0.0001$) and volume (AUC = 0.75, $P < 0.0001$), precuneus cortex (AUC = 0.70, $P < 0.0001$) and volume (AUC = 0.74, $P < 0.0001$), superior
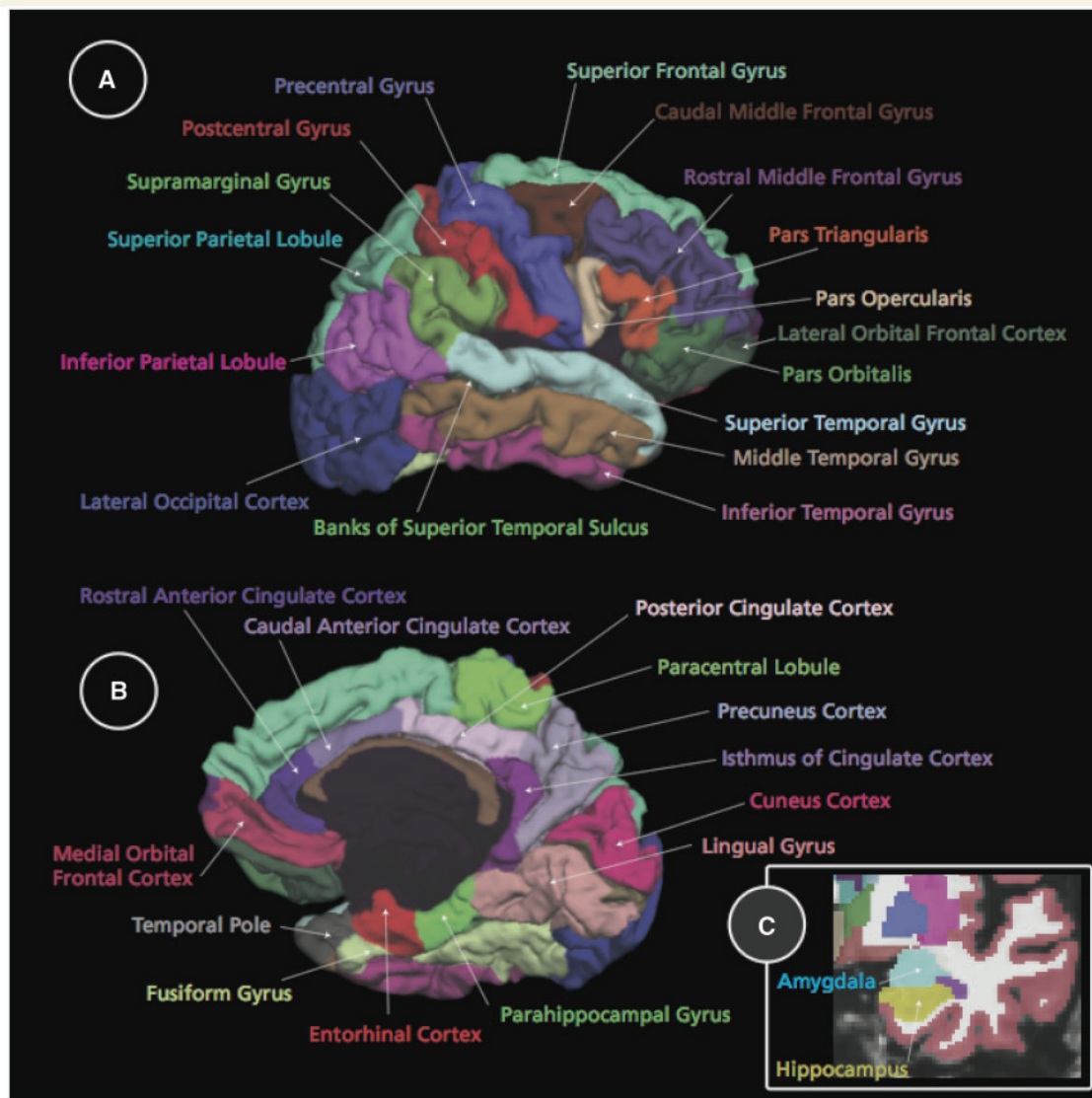
Figure 1 Three-dimensional representations of all 34 ROIs examined in the current study (only one hemisphere is shown). All of the neocortical ROIs visible in (**A**) lateral and (**B**) medial views of the grey matter surface and (**C**) the two non-neocortical regions (i.e. the hippocampus and amygdala) visible in the coronal view of a T_1-weighted MRI image.

temporal gyrus thickness (AUC = 0.78, $P < 0.0001$), supramarginal gyrus thickness (AUC = 0.75, $P < 0.0001$), temporal pole thickness (AUC = 0.75, $P < 0.0001$), amygdala volume (AUC = 0.78, $P < 0.0001$) and hippocampal volume (AUC = 0.82, $P < 0.00001$) (Fig. 2).

The final regression model, estimated from the training cohort, demonstrated that entorhinal cortex thickness (OR = 0.18, $P < 0.0001$), hippocampal volume (OR = 0.20, $P < 0.0001$) and supramarginal gyrus thickness (OR = 0.02, $P < 0.0001$) represented the best set of discriminators for MCI. In the comparison between the OCs and the MCI individuals, these three measures demonstrated an AUC = 0.91.

For the validation cohort (ADNI subjects), in the comparison between the OCs and the MCI individuals, entorhinal cortex thickness, hippocampal volume and supramarginal thickness demonstrated an AUC = 0.95. In the comparison between the OCs and patients with Alzheimer's disease, these three measures

demonstrated an AUC = 1.00. The application of the logistic regression coefficients from the model based on the training cohort to the validation cohort resulted in an AUC of 0.95. The AUC, sensitivity and specificity, negative and positive predictive values and negative and positive likelihood ratios for both the training and validation cohorts are presented in Table 2.

The correlations between the automated MRI measures and the clinical, neuropsychological and CSF measures are presented in Table 3. Figure 3 illustrates in pictorial format the nature of the relationship between one of the automated MRI measures, entorhinal cortex thickness and the three CSF biomarkers.

Discussion

The results demonstrate that automated MRI measures of entorhinal cortex thickness, hippocampal volume and supramarginal

Table 2 Discrimination results for automated MRI measures from final stepwise regression model

	Training cohort (OASIS subjects) (n = 98; MCI 48, OC 49)			Validation cohort (ADNI subjects) (n = 151; MCI 57, OC 94)		
	Regression coefficient (SE)	Odds ratio (95% CI)	P-value	Regression coefficient (SE)	Odds ratio (95% CI)	P-value
Intercept	34.9 (7.91)			31.3 (5.68)		
Entorhinal thickness	−1.58 (0.61)	0.26 (0.09−0.73)	0.0097	−2.48 (0.63)	0.13 (0.05−0.71)	<0.0001
Hippocampal volume	−1.62 (0.63)	0.28 (0.11−0.73)	0.0105	−1.29 (0.45)	0.34 (0.16−0.71)	0.0041
Supramarginal thickness	−3.90 (1.42)	0.19 (0.18−0.75)	0.0062	−2.25 (0.86)	0.42 (0.22−0.81)	0.0096
Area Under Curve (AUC)[a] (95% CI)		0.91 (0.83−0.95)			0.95 (0.90−0.97)	
Sensitivity[a] (95% CI)		73% (58−85%)			90% (79−96%)	
Specificity[a] (95% CI)		94% (83−99%)			91% (84−96%)	
Negative predictive value[a]		78%			94%	
Positive predictive value[a]		92%			85%	
Negative likelihood ratio[a]		0.29			0.11	
Positive likelihood ratio[a]		12.12			10.00	

SE = standard error; CI = confidence interval; Odds ratio is for a 1 SD difference in the independent variable.
a Derived from entorhinal thickness, hippocampal volume and supramarginal thickness.

Table 3 Correlation results from the validation cohort (ADNI subjects) between the automated MRI measures that best discriminated the MCI group and clinical, neuropsychological and CSF biomarker evaluations

Region of interest	CDR-SB	MMSE	AVLT 5 min recall	AVLT 30 min recall	Trails B	Tau	P-Tau	Abeta 42
Entorhinal cortex thickness	−0.81 (0.0001)	0.72 (0.0001)	0.66 (0.0001)	0.70 (0.0001)	−0.51 (0.0001)	−0.41 (0.0001)	−0.47 (0.0001)	0.38 (0.0001)
Hippocampal volume	−0.71 (0.0001)	0.60 (0.0001)	0.62 (0.0001)	0.62 (0.0001)	−0.51 (0.0001)	−0.37 (0.0003)	−0.44 (0.0001)	0.43 (0.0001)
Supramarginal gyrus thickness	−0.50 (0.0001)	0.43 (0.0001)	0.39 (0.0003)	0.42 (0.0001)	−0.39 (0.0001)	−0.33 (0.0003)	−0.38 (0.0003)	0.26 (0.006)

Spearman's rank correlation coefficients listed with P-values in parenthesis.

gyrus thickness identify MCI and Alzheimer's disease individuals with excellent discrimination accuracy and specificity, exhibit a high degree of consistency and reproducibility across multiple independent cohorts, and correlate strongly with clinical measures of decline as well as cellular biomarkers. Taken together, these findings suggest the feasibility of using automated, MRI-based software tools as a diagnostic marker for Alzheimer's disease.

The regression analyses presented here indicate that automated MRI measures can differentiate MCI and Alzheimer's disease from normal ageing with excellent discrimination accuracy. In the comparisons between the MCI individuals and OCs, entorhinal cortex thickness, hippocampal volume and supramarginal gyrus thickness demonstrated an average AUC of 0.91 in the training cohort and an AUC of 0.95 in the validation cohort. Using these same MRI measures, patients with mild Alzheimer's disease could be differentiated from OCs with perfect discrimination (AUC = 1.0). These AUC values are more accurate than prior MRI (Xu et al., 2000; Devanand et al., 2007; Colliot et al., 2008; Kloppel et al., 2008), FDG–PET (Mosconi et al., 2008; Jack et al., 2008) or amyloid-binding PET studies (Jack et al., 2008; Li et al., 2008). The MCI discrimination accuracies presented here are comparable to one prior PET study utilizing a radioactive amyloid and tau protein tracer (Small et al., 2006)

and two prior MRI studies where a smaller number of subjects were examined from a single cohort (Killiany et al., 2000; Davatzikos et al., 2008). Further studies are needed to determine whether combining structural MRI measures with other imaging modalities will improve diagnostic and prediction accuracy and whether the benefits of using multiple methods outweigh the costs.

To the best of our knowledge, this is the first study to demonstrate that automated software tools can be utilized to quantify the atrophy of individual anatomic regions in a highly specific and precise fashion. The fact that entorhinal cortex thickness and hippocampal volume were two of the best discriminators of MCI indicates the specificity of these automated MRI methods for identifying the two regions implicated in the earliest stages of Alzheimer's pathology (Braak and Braak, 1991; Kemper, 1994). Consistent with prior MRI studies (Scahill et al., 2002; Buckner et al., 2005; Dickerson et al., 2009), these results also highlight the relative importance of examining lateral parietal regions, such as the supramarginal gyrus, as important discriminators for the earliest stages of Alzheimer's disease.

The regression results further illustrate that these automated MRI measures are highly consistent and reproducible. In the comparison between the MCI individuals and OCs, both the training
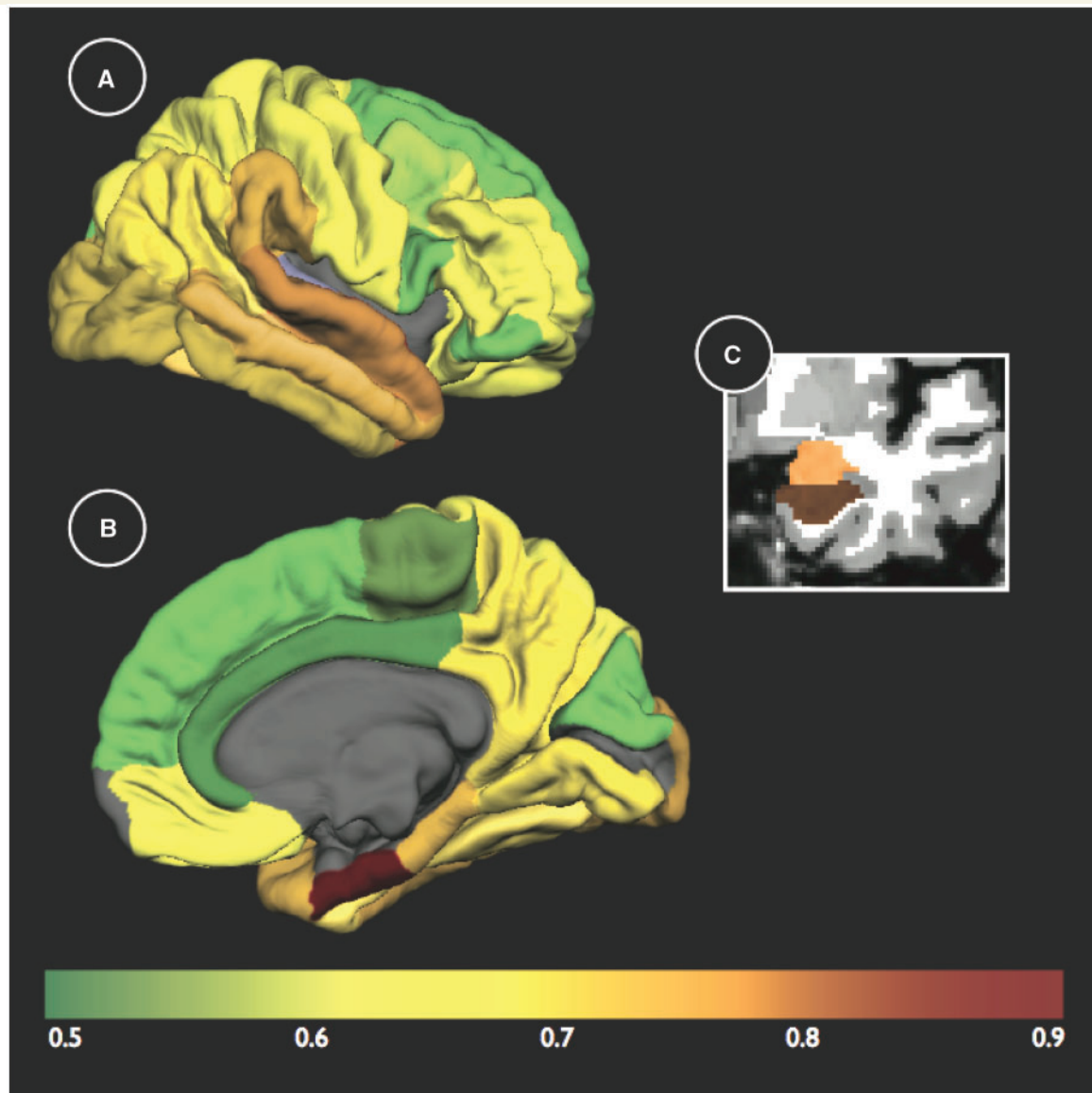
Figure 2 AUC results (neocortical thickness and non-neocortical volumes) from the first regression model (MCI versus older controls) for all of the automated ROIs from the training cohort (OASIS subjects) displayed on the grey matter surface (only one hemisphere is shown) in (**A**) lateral, (**B**) medial views and (**C**) the two non-neocortical regions (i.e. the hippocampus and amygdala) in the coronal view of a T_1-weighted MRI image. The colour scale at the bottom represents the discrimination accuracy (AUC value), with green indicating regions of lowest discrimination and brown/red indicating regions of highest discrimination (please see text for specific AUC values for each ROI).

and the validation cohorts demonstrated similar AUC values indicating the reliability of these measures across multiple independent cohorts. Furthermore, for the validation cohort, the application of the training cohort logistic regression coefficients resulted in an AUC value of 0.95, the same as the value derived without the application of these coefficients. This shows that the model based on the training cohort using the three temporoparietal measures is clinically applicable and can be reproduced in populations other than that from which the training cohort were drawn.

The correlations between the three MRI measures and measures of clinical severity (i.e. CDR-SB and MMSE) suggest the potential for using these measures as surrogate markers of underlying disease. Correlations between tests of episodic memory function (AVLT 5 and 30 min recall) and measures of entorhinal cortex thickness and hippocampal volume are consistent with the fact that declines in episodic memory function are reported as predictors of disease progression. Future studies will examine whether combining these automated MRI measures with neuropsychological assessments will better predict which MCI individuals eventually progress to Alzheimer's disease.

Correlations between the three temporoparietal measures (i.e. entorhinal cortex thickness, hippocampal volume and
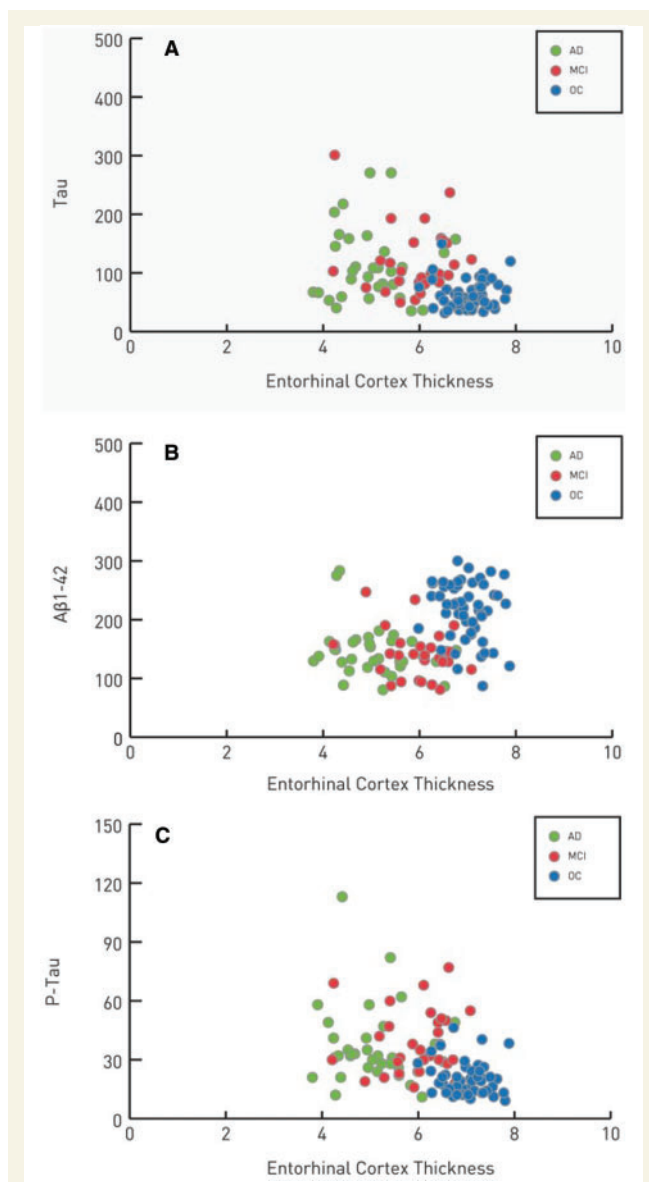
Figure 3 Scatter plots illustrating the relationship between total entorhinal cortex thickness and CSF measures of (**A**) tau protein, (**B**) abeta 42 protein and (**C**) p-tau protein for 33 Alzheimer's disease (green circles), 30 MCI (red circles) and 52 older controls (blue circles) individuals. Cortical thickness values are expressed in mm and CSF measures are expressed in picograms per mm.

supramarginal gyrus thickness) and CSF measures of tau, p-tau and abeta 42 suggest that these MRI measures are likely to be a reflection of known underlying Alzheimer's disease pathology. When considered together with the regression results, these data suggest the hierarchical fashion in which pathology affects the earliest stages of Alzheimer's disease, with tau-associated neurofibrillary changes in medial temporal regions and abeta-associated amyloid changes in the entorhinal cortex and neocortical regions (Arnold *et al*., 1991; Braak and Braak, 1991; Kemper, 1994).

The methods we have described here can be implemented in clinical practice for the diagnosis of MCI and Alzheimer's disease.

Using these software tools a single volumetric T_1-weighted MRI scan can be completely processed, with little to no manual intervention, in a relatively short amount of time. The training cohort regression coefficients presented here can then be applied to the final output values of entorhinal cortex thickness, hippocampal volume and supramarginal gyrus thickness to calculate the predictive probability of a single individual being diagnosed as either MCI or Alzheimer's disease.

The present study has limitations. Since the MCI individuals in the two cohorts were diagnosed using slightly different criteria, differences between the two MCI groups could have affected the ability to independently assess the discrimination accuracy of the automated MRI measures. Another limitation is that the two MCI cohorts had differing percentages of males and females, with the training cohort comprised of a larger number of females and the validation cohort comprised of a larger number of males. The fact that the AUC values were comparable between the two cohorts and that the application of the training cohort logistic regression coefficients resulted in the same AUC value as without the application of these coefficients, suggests that the differences observed between the two cohorts did not play a major role in affecting the main findings of this study.

One concern is that although the procedures demonstrated here generalized across clinically diagnosed Alzheimer's disease and MCI populations, these procedures may be less accurate in the clinical setting where a range of cognitive disorders and dementia subtypes are present. The fact that the current results show complete discrimination suggests that these tools would be additionally powerful in the clinical setting. Future work will examine the application of these automated MRI measures to a larger, community-based, volunteer cohort that would be more representative of a clinical setting. Another concern is regarding clinical utility and whether these automated MRI measures can predict progression from MCI to Alzheimer's disease. Recently, we have completed a study examining the feasibility of using these automated MRI measures to identify those MCI individuals, within a larger MCI cohort, at greatest risk for Alzheimer's disease. Preliminary evidence from this study indicates that these automated MRI measures can identify MCI converters from MCI non-converters with a high degree of accuracy and have significant benefit when compared to clinical and neuropsychological assessments alone for predicting progression from MCI to Alzheimer's disease (Desikan *et al*., 2009).

The identification of individuals in a transitional phase is critical for testing disease-modifying therapies and for the development of novel medications to prevent or delay Alzheimer's disease. The results from this study demonstrate that automated MRI-based neuroanatomic measures provide one cost-effective and efficient method to identify individuals in the earliest stages of Alzheimer's disease and may further serve as a quantitative and biologically meaningful endpoint in therapeutic trials.

Acknowledgements

References

Arnold SE, Hyman BT, Flory J, Damasio AR, Van Hoesen GW. The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. Cereb Cortex 1991; 1: 103–16.

Bennett DA, Schneider JA, Bienias JL, Evans DA, Wilson RS. Mild cognitive impairment is related to Alzheimer disease pathology and cerebral infarctions. Neurology 2005; 64: 834–41.

Blacker D, Lee H, Muzikansky A, Martin EC, Tanzi R, McArdle JJ, et al. Neuropsychological measures in normal individuals that predict subsequent cognitive decline. Arch Neurol 2007; 64: 862–71.

Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. Acta Neuropathology 1991; 82: 239–59.

Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 2004; 23: 724–38.

Buckner RL, Snyder AZ, Shannon BJ, LaRossa G, Sachs R, Fotenos AF, et al. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. J Neurosci 2005; 25: 7709–17.

Colliot O, Chételat G, Chupin M, Desgranges B, Magnin B, Benali H, et al. Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. Radiology 2008; 248: 194–201.

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 1999; 9: 179–94.

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging 2008; 29: 514–23.

Davis H, Rockwood K. Conceptualization of mild cognitive impairment: a review. Internatl J Geriatr Psychiatr 2004; 19: 313–19.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006; 31: 968–80.

Desikan RS, Cabral HJ, Settecase F, Hess CP, Dillon WP, Glastonbury CM, et al. Use of automated MR imaging cortical thickness and volume measurements with neuropsychological testing for predicting progression from mild cognitive impairment to Alzheimer disease. American Society for Neuroradiology 47th Annual Meeting 2009; (poster presentation).

Devanand DP, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, et al. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. Neurology 2007; 68: 828–36.

Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN, et al. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild Alzheimer's disease dementia and is detectable in asymptomatic amyloid-positive individuals. Cereb Cortex 2009; 19: 497–510.

Estévez-González A, Kulisevsky J, Boltes A, Otermín P, García-Sánchez C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: comparison with mild cognitive impairment and normal aging. Int J Geriatr Psychiatry 2003; 18: 1021–28.

Fennema-Notestine C, Gamst AC, Quinn BT, Pacheco J, Jernigan TL, Thal L, et al. Feasibility of multi-site clinical structural neuro-imaging studies of aging using legacy data. Neuroinformatics 2007; 5: 235–45.

Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. Neuroimage 1999a; 9: 195–207.

Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Map 1999b; 8: 272–84.

Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA 2000; 97: 11050–55.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002; 33: 341–55.

Folstein M, Folstein S, McHugh P. ''Mini-Mental State''. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12: 189–98.

Gomez-Isla T, Price JL, McKeel DW Jr, Morris JC, Growdon JH, Hyman BT. Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. J Neurosci 1996; 16: 4491–4500.

Hampel H, Bürger K, Teipel SJ, Bokde AL, Zetterberg H, Blennow K. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. Alzheimers Dement 2008; 4: 38–48.

Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, et al. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 2006; 32: 180–94.

Head D, Snyder AZ, Girton LE, Morris JC, Buckner RL. Frontal-hippocampal double dissociation between normal aging and Alzheimer's disease. Cereb Cortex 2005; 15: 732–9.

Jack CR Jr, Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, et al. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain 2008; 131: 665–80.

Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, et al. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 2009; 46: 177–92.

Kemper TL. Neuroanatomical and neuropathological changes in normal aging and in dementia. In: Albert M, Knoefel J, editors. Clinical neurology of aging. New York: Oxford University Press; 1994. p. 3–78.

Killiany R, Gomez-Isla T, Moss M, Kikinis R, Sandor T, Jolesz F, et al. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. Ann Neurol 2000; 47: 430–39.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008; 131: 681–9.

Lezak MD. Neuropsychological Assessment. Oxford: Oxford University Press; 1995.

Li Y, Rinne JO, Mosconi L, Pirraglia E, Rusinek H, Desanti S, et al. Regional analysis of FDG and PIB-PET images in normal aging, mild cognitive impairment, and Alzheimer's disease. Eur J Nucl Med Mol Imaging 2008; 35: 2169–81.

Lopez OL, Jagust WJ, DeKosky ST, Becker JT, Fitzpatrick A, Dulberg C, et al. Prevalence and classification of mild cognitive impairment in the cardiovascular health study cognition study: part 1. Arch Neurol 2003; 60: 1385–89.

Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 2007; 19: 1498–1507.

McKhann G, Drachman D, Folstein MF, Katzman R, Price D, Stadlan E. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-Alzheimer's disease RDA Work group under the auspices of Department of Health and Human Services Task Force. Neurology 1984; 34: 939–44.

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993; 43: 2412–14.

Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild cognitive impairment represents early-stage Alzheimer disease. Arch Neurol 2001; 58: 397–405.

Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, et al. Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. J Nucl Med 2008; 49: 390–98.

Nam B-H, D'Agostino R. Discrimination index, the area under the ROC curve. In: Huber-Carol C, Balakrishnan N, Nikulin MS, Mesbah M, editors. Goodness-of-fit tests and model validity. Boston: Birkhäuser; 2002. p. 273–77.

Petersen R, Smith G, Waring S, Ivnik R, Tangalos E, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999; 56: 303–8.

Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, et al. Arch Neurol 2001; 58: 1985–92.

Petersen R. Mild cognitive impairment. J Intern Med 2004; 256: 183–194.

Petersen R, Thomas R, Grundman M, Bennett D, Doody R, Ferrris S, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 2005; 352: 2379–88.

Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958; 8: 271–76.

Rosas HD, Liu AK, Hersch S, Glessner M, Ferrante RJ, Salat DH, et al. Regional and progressive thinning of the cortical ribbon in Huntington's disease. Neurology 2002; 58: 695–701.

Rubin EH, Storandt M, Miller JP, Kinscherf DA, Grant EA, Morris JC, et al. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. Arch Neurol 1998; 55: 395–401.

Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RS, Busa E, et al. Thinning of the cerebral cortex in aging. Cereb Cortex 2004; 14: 721–30.

Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. Proc Natl Acad Sci USA 2002; 99: 4703–07.

Small GW, Kepe V, Ercoli LM, Siddarth P, Bookheimer SY, Miller KJ, et al. PET of brain amyloid and tau in mild cognitive impairment. N Engl J Med 2006; 355: 2652–63.

Wechsler D. Wechsler memory scale-revised. San Antonio: The Psychological Corporation; 1987.

Xu Y, Jack CR Jr, O'Brien PC, Kokmen E, Smith GE, Ivnik RJ, et al. Usefulness of MRI measures of entorhinal cortex versus hippocampus in Alzheimer's disease. Neurology 2000; 54: 1760–7.