



Automated MRI measures predict progression to Alzheimer's disease

Rahul S. Desikan^{a,*}, Howard J. Cabral^b, Fabio Settecase^c, Christopher P. Hess^c,
William P. Dillon^c, Christine M. Glastonbury^c, Michael W. Weiner^{c,d},
Nicholas J. Schmansky^a, David H. Salat^a, Bruce Fischl^{a,e,f}; the Alzheimer's Disease
Neuroimaging Initiative[†]

^a Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA

^b Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

^c Department of Radiology, University of California – San Francisco, San Francisco, CA, USA

^d Department of Veteran Affairs, San Francisco, CA, USA

^e Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA

^f Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA

Received 23 January, 2010; received in revised form 21 April 2010; accepted 22 April 2010

Abstract

The prediction of individuals with mild cognitive impairment (MCI) destined to develop Alzheimer's disease (AD) is of increasing clinical importance. In this study, using baseline T1-weighted MRI scans of 324 MCI individuals from two cohorts and automated software tools, we employed factor analyses and Cox proportional hazards models to identify a set of neuroanatomic measures that best predicted the time to progress from MCI to AD. For comparison, cerebrospinal fluid (CSF) assessments of cellular pathology and positron emission tomography (PET) measures of metabolic activity were additionally examined. By 3 years follow-up, 60 MCI individuals from the first cohort and 58 MCI individuals from the second cohort had progressed to a diagnosis of AD. Cox models on the first cohort demonstrated significant effects for the medial temporal factor [Hazard Ratio (HR) = 0.43 {95% confidence interval (CI), 0.32–0.55}, $p < 0.0001$], the fronto-parietoccipital factor [HR = 0.59 {95% CI, 0.48–0.80}, $p < 0.001$], and the lateral temporal factor [HR = 0.67 {95% CI, 0.52–0.87}, $p < 0.01$]. When applied to the second cohort, these Cox models showed significant effects for the medial temporal factor [HR = 0.44 {0.32–0.61}, $p < 0.001$] and lateral temporal factor [HR = 0.49 {0.38–0.62}, $p < 0.001$]. In a combined Cox model, consisting of individual CSF, PET, and MRI measures that best predicted disease progression, only the medial temporal factor [HR = 0.53 {95% CI, 0.34–0.81}, $p < 0.001$] demonstrated a significant effect. These findings illustrate that automated MRI measures of the medial temporal cortex accurately and reliably predict time to disease progression, outperform cellular and metabolic measures as predictors of clinical decline, and can potentially serve as a predictive marker for AD.

© 2010 Elsevier Inc. All rights reserved.

Keywords: Computational MRI; AD; Mild cognitive impairment; Clinical biomarker

Predicting the time to progress from mild cognitive impairment (MCI) to clinical Alzheimer's disease (AD) is of increasing importance as therapeutic interventions for the prevention or delay of dementia onset are developed. Structural MRI provides visualization of the macroscopic tissue atrophy that results from the cellular changes underlying AD and as such, offers one potential, noninvasive method for early detection and prediction of AD.

A number of prior structural MRI studies have employed either manual region of interest (ROI) (Devanand et al., 2007; Fleisher et al., 2008; Jack et al., 1999; Killiany et al., 2002) or automated whole brain approaches (Karas et al., 2008; Whit-

* Corresponding author at: Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Building, 149, 13th Street, Room, 2301 Charlestown, Ma, USA 02129. Tel: (617) 726-4897; fax: (617) 726-7422.

E-mail address: rahul@nmr.mgh.harvard.edu (R.S. Desikan).

[†] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators is available at www.loni.ucla.edu/ADNI/Collaboration/ADNI_Citation.shtml.

well et al., 2007; Whitwell et al., 2008) to identify those MCI individuals at greatest risk for AD. Though these methods offer several advantages, manual ROI methods have limited clinical use because they do not allow for the timely analysis of regions across the entire brain or in large datasets while many whole brain approaches cannot evaluate the disease state in a single individual. More recent ROI and whole brain studies have demonstrated that automated MRI-based computational measures can successfully discriminate those MCI individuals who progress to AD from those MCI individuals who do not progress (Bakkour et al., 2009; Misra et al., 2009; Querbes et al., 2009; Vemuri et al., 2009). However, only a handful of studies have attempted to examine the time to progress from MCI to AD and most of these have used manual ROI approaches (Desikan et al., 2009a; Devanand et al., 2007; Jack et al., 1999).

Advances in image analysis algorithms have led to the development of structural MRI-based software tools that can automatically parcellate the entire brain into anatomic regions and quantify the tissue properties in these regions for a single individual (Desikan et al., 2006; Desikan et al., 2009b; Fischl et al., 2002). In this study, we investigated the feasibility of using these automated MRI-based software tools as a predictive marker for AD. Using baseline MRI scans from 162 MCI individuals, we employed factor analyses and Cox proportional hazards models to identify a set of neuroanatomic regions that best predicted the time to progress from MCI to AD. We then examined the predictive consistency of these automated MRI measures on a second cohort of 162 MCI individuals. For comparison, we additionally evaluated cerebrospinal fluid (CSF) assessments of cellular pathology and positron emission tomography (PET) measures of metabolic activity.

1. Methods

1.1. Overview

All participants were selected from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). The ADNI is a large multisite 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 nonprofit 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 AD. The Principal Investigator of this initiative is Michael Weiner MD, and ADNI is the result of many coinvestigators from a broad range of academic institutions and private corporations, with subjects recruited from over 50 sites across the USA and Canada. For more information, please see www.adni-info.org.

1.2. Clinical Assessments and Group Characteristics

The institutional review boards of all participating institutions approved the procedures for this study. Written informed consent was obtained from all participants or surrogates. Each participant was selected using eligibility criteria that are described in detail elsewhere (www.adni-info.org/index.php?option=com_content&task=view&id=9&Itemid=43). The degree of clinical severity for each participant 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, which included a set of questions regarding the functional status of the participant, along with a standardized neurologic, psychiatric, and health examination. The MiniMental State Examination (MMSE) (Folstein et al., 1975) and a complete neuropsychological battery were also conducted.

Participants were selected from the ADNI database if they were clinically classified as the amnesic subtype of MCI, based on the revised MCI criteria (Petersen, 2004). These included 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. From a total of over 400 MCI individuals available from the ADNI database, 324 individuals with the amnesic subtype of MCI were selected. For the purposes of the present study, this larger sample was randomly split (using a random number generator employing a uniform distribution) into two equal samples (training and validation cohorts). No statistical differences in demographic variables were noted between the two samples.

1.3. Cohort 1 – training cohort

At baseline, 162 MCI individuals were examined. At follow-up (mean follow-up time of 1.89), 60 of these individuals met clinical criteria for probable AD (MCI-Converters) (McKhann et al., 1984) while 102 remained mildly impaired (MCI-Nonconverters). Of those mildly impaired at follow-up, 46 had CDR-SB scores that increased but their impairments had not progressed to the point where they received a diagnosis of AD, 32 had CDR-SB scores that remained stable and 24 had CDR-SB scores that declined. Approximately 54 % of these mildly impaired subjects ($n = 55$) had a CDR-SB of two or higher, and approximately 46% ($n = 47$) had a CDR-SB score of 0.5–1.5. The mean age, educational status, CDR-SB scores, MiniMental State

Table 1
Descriptive statistical information for the participants in the study

Diagnostic group	Training cohort		Validation cohort	
	MCI-converters	MCI-nonconverters	MCI-converters	MCI-nonconverters
Sample size	60	102	58	104
Age	75.73 (6.86)	74.27 (7.47)	74.95 (7.21)	75.58 (7.22)
Percent female	42%	34%	40%	32%
MMSE	26.58 (1.83)	27.49 (1.73)	26.68 (1.70)	27.04 (1.71)
CDR-SB	1.85 (0.93)	1.53 (0.87)	1.83 (1.05)	1.46 (0.79)
Percent APOE-ε4	65%	50%	67%	48%
Follow-up Time (years)	1.44 (0.60)	1.89 (0.66)	1.50 (0.68)	1.90 (0.62)

Means listed with standard deviations in parentheses.

examination (MMSE) scores, gender distribution, percent APOE-ε4, and follow-up time are shown in Table 1.

1.4. Cohort 2 – validation cohort

At baseline, 162 MCI individuals were examined. At follow-up (mean follow-up time of 1.90), 58 of these individuals met clinical criteria for probable AD (MCI-Converters) (McKhann et al., 1984) while 104 remained mildly impaired (MCI-Nonconverters). Of those mildly impaired at follow-up, 52 had CDR-SB scores that increased but their impairments had not progressed to the point where they received a diagnosis of AD, 23 had CDR-SB scores that remained stable, and 29 had CDR-SB scores that declined. Approximately 42 % of these mildly impaired subjects (n = 44) had a CDR-SB of two or higher, and approximately 58% (n = 60) had a CDR-SB score of 0.5–1.5.

1.5. MRI image acquisition

All ADNI MRI scans were acquired at multiple sites using either a GE, siemens, or Philips 1.5-T system. Multiple high-resolution T1-weighted volumetric MP-RAGE scans were collected for each subject and the raw DICOM images were downloaded from the public 1ADNI site (www.loni.ucla.edu/ADNI/Data/index.shtml). Parameter values vary depending on scanning site and can be found at www.loni.ucla.edu/ADNI/Research/Cores/.

1.6. Automated image analysis procedures

All MRI scans were processed, with little to no manual intervention, using the FreeSurfer software package, freely available at surfer.nmr.mgh.harvard.edu. A single, raw MPRAGE MRI acquisition for each participant was downloaded from the ADNI database and normalized for intensity inhomogeneities to create an image volume with high contrast-to-noise (Dale et al., 1999). This volume was used to locate the gray/white matter boundary (white matter surface) (Fischl et al., 1999a) and this in turn, was then used to locate the gray/CSF boundary (gray matter surface) (Fischl and Dale, 2000). Cortical thickness measurements were then obtained by calculating the distance between the gray and the white matter surfaces at each point (per hemisphere) across the entire cortical mantle (Fischl and Dale, 2000).

This cortical thickness measurement technique has been previously validated via histological (Rosas et al., 2002) as well as manual measurements from MRI (Salat et al., 2004). The reliability of these cortical thickness measures as well as the other image analysis procedures presented here has been demonstrated across different scanner manufacturers and upgrades, varying contrast with noise ratio, and the number of MPRAGE MRI acquisitions used (Fennema-Notestine et al., 2007; Han et al., 2006; Jovicich et al., 2009).

The neocortex of the brain on the MRI scans was then automatically subdivided into 32 gyral-based ROIs (per hemisphere, total of 64 neocortical measures). 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 gray 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. Of note, none of the herein presented 324 datasets required any manual editing. All the MRI scans were processed on a Linux cluster with 128 nodes, each with a two quad-core Xeon 5472 CPU (Intel Corporation, Santa Clara CA, USA) and 32 GB of RAM. Processing time for each MRI scan was approximately 11–14 hours. This cluster allows for the processing of 1,024 MRI scans simultaneously and all 324 MRI scans used in the present study were processed in one day.

In total, 34 neocortical and non-neocortical ROIs were used in this study. For all the analyses performed here, the mean thickness (only neocortical regions) and the volume (only non-neocortical regions) of the right and the left hemispheres, for each ROI, were added together. 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).

1.7. Cerebrospinal fluid measures

From the current ADNI sample, several individuals (total $n = 168$, training cohort $n = 85$, validation cohort $n = 83$) underwent lumbar puncture for CSF biomarker evaluation. Methods for CSF acquisition and biomarker measurement have been reported previously for this sample (Shaw et al., 2009). In brief, CSF was collected and stored at -80°C at the University of Pennsylvania ADNI Biomarker Core Laboratory. Amyloid beta fibrils from peptides 1–42 ($A\beta_{1-42}$), total tau (t-tau), and phosphorylated tau (p-tau) were measured using the multiplex xMAP Luminex platform (Luminex, Corp, Austin, TX) with Innogenetics (INNOBIA AlzBio3, Ghent, Belgium) immunoassay kit-based reagents.

1.8. Positron emission tomography measures

From the current ADNI sample, several individuals (total $n = 153$, training cohort $n = 76$, validation cohort $n = 77$) underwent PET imaging with the metabolic tracer [^{18}F] fluorodeoxyglucose (FDG-PET). Methods for FDG-PET data collection and analysis using ROIs have been reported previously for this sample (Jagust et al., 2009). In brief, FDG-PET data were first intensity normalized to a reference ROI that was comprised of the averaged pons and cerebellar vermis. For the FDG-PET analyses, ROIs were defined using coordinates from the Montreal Neurological Institute atlas. Each individual's PET scan was then spatially normalized to the SPM5 PET template, and mean FDG counts were extracted from each ROI. These ROI mean counts were then averaged to form a single "composite" FDG ROI that was the variable used in all FDG-PET analyses. For the purposes of the present study, we used ROIs from bilateral inferior temporal and angular gyri as well as posterior cingulate and precuneus cortices since prior studies have demonstrated significant differences between AD individuals and normal elderly controls in these regions (for details see Jagust et al., 2009). All FDG ROI values used in the present study were downloaded from the public ADNI site (www.loni.ucla.edu/ADNI/Data/index.shtml).

1.9. Statistical analysis

Data reduction was conducted on the training cohort ROI measures using factor analysis. First, a correlation matrix was generated using Pearson's r to determine the correlation coefficient between the individual ROI measures. Those measures with very strong relationships ($r > .80$) were removed from further analyses (Pett et al., 2003). Next, the factors along with their loading coefficients and the cumulative variance of the factors were extracted from the analysis. We applied a minimum eigenvalue criterion of 1.0 together with the examination of a scree plot to determine how many factors to retain and these were rotated (varimax) to maximize the relationship between variables. The factor

loadings derived from the training cohort ROI measures were then applied to the validation cohort ROI measures.

The automated MRI-derived factors, along with the PET and CSF measures, were further analyzed using Cox proportional hazards models. These analyses tested whether specific predictors (e.g. medial temporal factor) are associated with time to a diagnosis of AD. In these models, the hazard ratio indicates the change in risk per 1 unit change in the predictor. For instance, if the hazard ratio is 0.43 for the medial temporal factor, each 1 SD decrease in the volume and thickness of this factor increases risk by 57%.

The primary focus of these survival analyses was time from study entry to the endpoint of interest, which was the diagnosis of AD. All proportional hazards models included age, education, gender, and APOE- $\epsilon 4$ carrier status as covariates. For the training cohort, we entered the automated MRI-derived factors into a multivariate Cox model to determine which of these measures best predicted the time to diagnosis. Then, for the validation cohort, we entered only those neuroanatomic factors that were the best predictors in the training cohort data into a second proportional hazards model and compared the respective hazard ratios derived from the training and validation cohorts for each of these measures. To determine the prediction accuracy of these neuroanatomic factors, we calculated the area under the curve (AUC), sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios individually for the training and validation cohorts.

In an additional set of analyses, we entered the individual CSF biomarkers and FDG-ROIs into separate Cox models to determine if these measures predicted time to diagnosis. Given the small number of combined available measurement for the CSF and PET measures (total MCI = 78, MCI-Converters = 30, MCI-Nonconverters = 48), we pooled together data from the training and validation cohorts for these proportional hazards analyses. Prediction accuracies for the combined CSF and PET measures that best predicted time to event were calculated as described above. Individual CSF and PET measures significant at the < 0.05 alpha level were combined with those automated MRI-derived factors that best predicted time to event and entered into a final Cox model to determine which individual measures, in combination, best predicted time to AD diagnosis. For each of the multivariate Cox models, we descriptively evaluated the proportional hazards assumption by checking whether the negative log of survival probabilities associated with higher levels of each independent variable were constant multiples of those of the lower levels, across the entire range of event time. In addition we tested interaction terms that included the log of time to further verify the assumption of proportional hazards.

2. Results

The factor analysis on the training cohort ROI measures first revealed that none of the measures demonstrated a

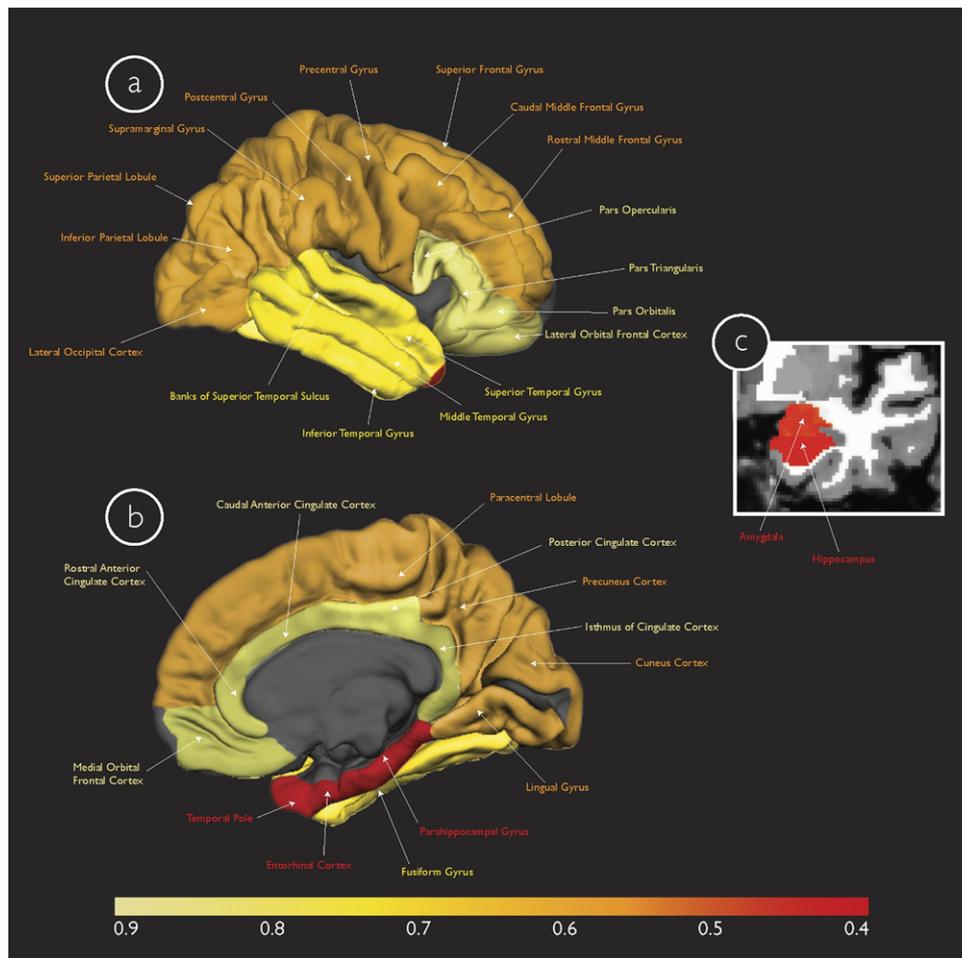


Fig. 1. Multivariate Cox model results for the four neuroanatomic factors derived from the training cohort displayed on the gray matter surface (only one hemisphere is shown) in (a) lateral and (b) medial views, and (c) in the coronal view of a T1-weighted MRI image. These factors include the 1) medial temporal factor (individual ROIs illustrated in red), 2) fronto-parietoccipital factor (individual ROIs illustrated in orange), 3) lateral temporal factor (individual ROIs illustrated in bright yellow), and 4) fronto-cingulate factor (individual ROIs illustrated in faded yellow). The color scale at the bottom illustrates the magnitude of risk (hazard ratio) associated with progressing from MCI to AD, with faded yellow indicating regions with the lowest risk and red indicating regions with the highest risk (please see text for specific hazard ratios for each of the factors).

Pearson's r of greater than 0.8 and therefore all ROI measures were included for further analyses. Four factors were extracted after rotation and confirmed using scree plots. The total explained variance for these factors was 73.11%. These four factors included: 1) the medial temporal factor, consisting primarily of the amygdala (Factor Loading (FL) = 0.84), hippocampus (FL = 0.78), entorhinal cortex (FL = 0.76), temporal pole (FL = 0.68), and parahippocampal gyrus (FL = 0.62), 2) the fronto-parietoccipital factor, consisting primarily of the superior parietal gyrus (FL = 0.87), cuneus cortex (FL = 0.77), caudal middle frontal gyrus (FL = 0.77), inferior parietal cortex (FL = 0.73) and lateral occipital cortex (FL = 0.73), 3) the fronto-cingulate factor, consisting primarily of the rostral anterior (FL = 0.78) and caudal anterior (FL = 0.74) portions of the cingulate cortex and the medial (FL = 0.73) and lateral (FL = 0.66) portions of the orbital frontal cortex, and 4) the lateral temporal factor, consisting primarily of the middle temporal gyrus (FL = 0.69), inferior temporal

gyrus (FL = 0.69), banks of the superior temporal sulcus (FL = 0.62) and superior temporal gyrus (FL = 0.54). **Figures 1 and 2** anatomically illustrate the factors used in this study.

For the training cohort, the multivariate proportional hazards model for the automated MRI-derived factors demonstrated significant effects for the medial temporal factor [Hazard Ratio (HR) = 0.43 {95% confidence interval (CI), 0.32–0.55}, $p < 0.0001$], the fronto-parietoccipital factor [HR = 0.59 {95% CI, 0.48–0.80}, $p < 0.001$], and the lateral temporal factor [HR = 0.67 {95% CI, 0.52–0.87}, $p < 0.01$]. For the validation cohort, the multivariate proportional hazards model showed significant effects for the medial temporal factor [HR = 0.44 {0.32–0.61}, $p < 0.001$] and lateral temporal factor [HR = 0.49 {0.38–0.62}, $p < 0.001$]. The prediction accuracy for the training cohort was AUC = 0.82, sensitivity = 74%, specificity = 84%, positive predictive value = 77%, negative predictive value = 82%, positive likelihood ratio = 4.65, negative likelihood

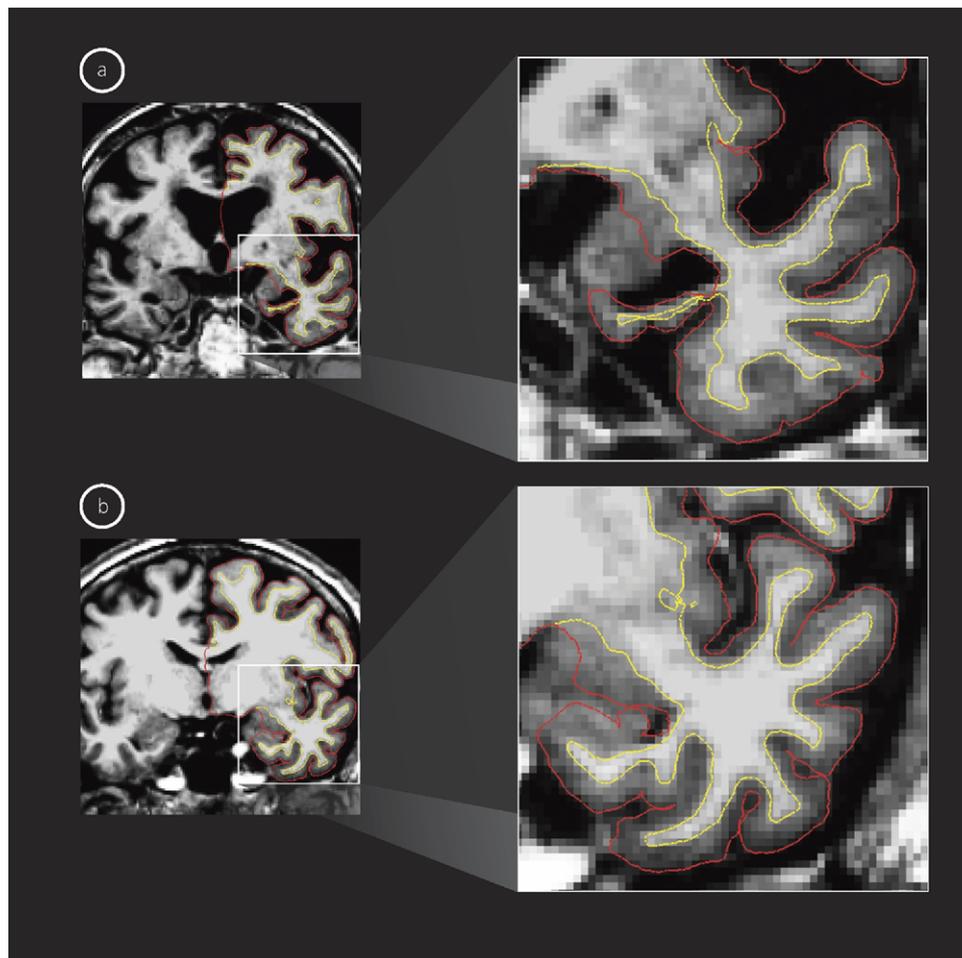


Fig. 2. T1-weighted MRI images in the coronal view showing marked differences in the medial temporal lobe, specifically hippocampal volume and entorhinal cortex thickness, for a representative (a) MCI-Converter and (b) MCI-Nonconverter. The red overlay shows the gray/CSF boundary and the white overlay shows the gray/white matter boundary and the distance between these surfaces represents the cortical thickness.

ratio = 0.30. The prediction accuracy for the validation cohort was AUC = 0.84, sensitivity = 87%, specificity = 66%, positive predictive value = 60%, negative predictive value = 82%, positive likelihood ratio = 2.53, negative likelihood ratio = 0.39.

The additional CSF and FDG-ROI multivariate proportional hazards models demonstrated significant effects for $A\beta_{1-42}$ [HR = 0.99 {95% CI, 0.98–0.99}, $p < 0.05$] and metabolic measurements from the inferior temporal cortex [HR = 0.002 {95% CI, 0.00–0.19}, $p < 0.01$]. The prediction accuracy of these two measures was AUC = 0.70, sensitivity = 93%, specificity = 48%, positive predictive value = 53%, negative predictive value = 92%, positive likelihood ratio = 1.78, negative likelihood ratio = 0.13. The final multivariate Cox model demonstrated that only the medial temporal factor [HR = 0.53 {95% CI, 0.34–0.81}, $p < 0.001$] represented the best measure to predict time to progress from MCI to AD. The lateral temporal factor [HR = 0.83 {95% CI, 0.47–1.40}, $p > 0.1$], $A\beta_{1-42}$ [HR = 0.99 {95% CI, 0.98–1.00}, $p > 0.1$] and metabolic measure-

ments from the inferior temporal cortex [HR = 0.24 {95% CI, 0.00–35.2}, $p > 0.1$] did not demonstrate significant effects. The addition of the neuroanatomic factors significantly improved ($p < 0.001$) the prediction accuracy compared with CSF and FDG-ROI measures alone (AUC = 0.83, sensitivity = 90%, specificity = 69%, positive predictive value = 64%, negative predictive value = 92%, positive likelihood ratio = 2.90, negative likelihood ratio = 0.32). To ensure that the results presented here were not subject to sample size related issues, we evaluated the four neuroanatomic factors as predictors in an independent Cox model using only those MCI individuals with CSF and PET measures (total MCI = 78, MCI-Converters = 30, MCI-Nonconverters = 48) and found that for this smaller subsample, the medial temporal factor continued to demonstrate significant effects [HR = 0.40 {95% CI, 0.19–0.58}, $p < 0.01$].

The probability of progressing from MCI to AD based on the value of the medial temporal factor (volume and thickness), derived from the training and validation cohorts, is

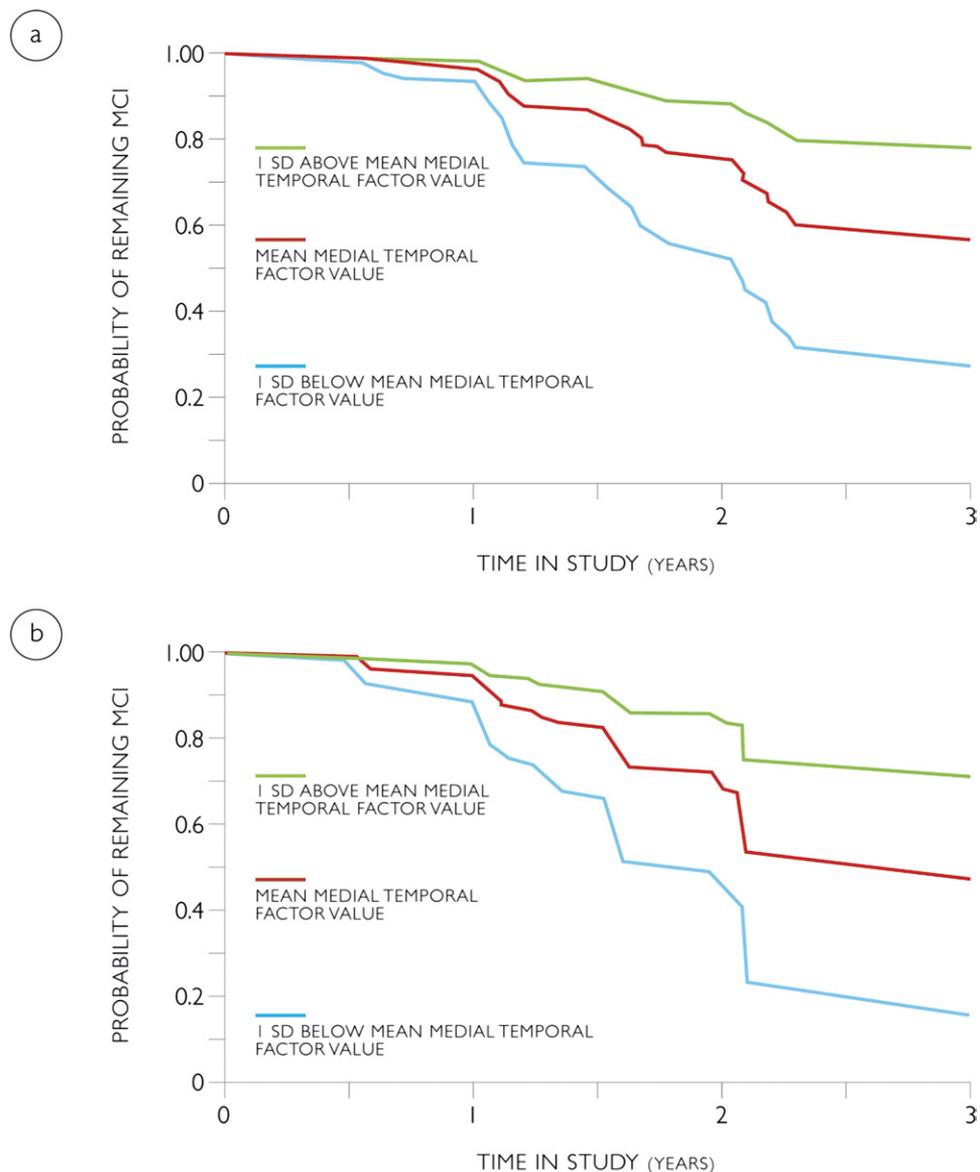


Fig. 3. Predicted survival plots estimating the probability of progressing from MCI to AD in the training (a) and validation (b) cohorts based on the value of the medial temporal factor (volume and thickness), shown at the mean (red lines) and one standard deviation above (green lines) and below (blue lines) the mean.

illustrated using predicted survival plots based on the multivariate Cox models (Figure 3). Based on these models, by 3 years from baseline, an MCI individual with a medial temporal factor value at the MCI average has a 44% to 52% probability of progressing to AD. In comparison, by 3 years from baseline, an MCI individual with a medial temporal factor value 1 SD below the MCI average has a 74% to 85% probability of progressing to AD and an MCI individual with a medial temporal factor value 1 SD above the MCI average has a 22% to 28% probability of progressing to AD.

3. Discussion

These results demonstrate that automated MRI measurements of medial temporal cortex thickness and volume

successfully predict the time to progress from MCI to AD, demonstrate robust reliability and consistency across multiple cohorts, and outperform CSF and PET measures as predictors of clinical disease progression. Taken together, these findings indicate the importance of using automated MRI-based software tools as a predictive marker for Alzheimer's disease.

Automated MRI measures can successfully quantify the risk associated with progressing to AD. Among MCI individuals, atrophy of the medial temporal factor regions (consisting primarily of entorhinal cortex thickness, hippocampal volume, amygdala volume, temporal pole thickness, and parahippocampal gyrus thickness) was associated with a greater than 50% risk increase in the training and validation

cohorts. Based on these multivariate models, by 3 years from baseline, MCI individuals with significant atrophy of the medial temporal factor regions (1 SD below the MCI average) are three times as likely to progress to AD, compared with MCI individuals with preserved medial temporal factor regions (1 SD above the MCI average). These time to progress risk assessments are consistent with prior structural MRI studies of MCI progression using the hippocampus and entorhinal cortex (Apostolova et al., 2006; Desikan et al., 2009a; Devanand et al., 2007; Jack et al., 1999; Kantarci et al., 2009) and demonstrate that limbic and pro-isocortical regions, such as the amygdala and temporal pole, are additionally important as predictors for the earliest stages of AD.

The results presented here indicate that unbiased dimension reduction approaches, such as factor analyses, can be used on a large number of automated MRI measures to derive a set of neuroanatomic factors that can successfully predict the time to progress from MCI to AD. The fact that the regions of the 1) medial temporal, 2) frontal parietal, 3) lateral temporal, and 4) frontal cingulate cortices combined together to form individual factors is consistent with the hierarchical pathology of AD (Arnold et al., 1991; Braak and Braak, 1991; Kemper, 1994) and suggests the potential of using automated MRI measures for the *in vivo* staging of Alzheimer's pathology. As disease predictors, though the value of the hazard ratios were comparable in magnitude, the narrow confidence intervals for the individual neuroanatomic factors, compared with the wide confidence intervals for the individual ROIs (data not presented), indicates that factor based automated MRI assessments are more precise than individual ROIs for quantifying the risk associated with disease progression.

The results from the multivariate proportional hazards models further illustrate that the training cohort derived medial temporal factor loadings are highly consistent and reproducible. In predicting the time to progress from MCI to AD, the hazard ratio for the medial temporal factor was similar across the training and validation cohorts indicating the reliability of this measure across multiple MCI cohorts. In contrast, the hazard ratio for the lateral temporal factor was different across the training and validation cohorts suggesting that for disease prediction this measure is not as reliable as the medial temporal factor. Furthermore, when the medial temporal factor regions were used for estimating survival, a similar number of MCI individuals from the training and validation cohorts progressed to AD at 1, 2, and 3 years from baseline. This shows that the training cohort derived medial temporal factor loading values are clinically applicable and can be used for predicting disease progression in populations other than that from which the training cohort were drawn.

As a predictor of clinical disease progression, atrophy of the medial temporal cortex outperforms cellular measures of pathology and metabolic assessments of the cerebral cortex. Though CSF-based $A\beta_{1-42}$ and PET-based metabolic mea-

surements from the inferior temporal cortex individually predicted progression to AD, when combined with the neuroanatomic factors, these measures could not predict clinical decline as successfully as the medial temporal factor. These findings are in accordance with prior MCI studies demonstrating better disease prediction with structural neuroimaging markers than either CSF (Vemuri et al., 2009) or PET measures (Walhovd et al., 2009), although a recent meta-analysis did find that PET is a slightly better disease predictor than MRI (Yuan et al., 2009). Given the high negative predictive value of 92% and the low negative likelihood ratio of 0.13, these findings suggest the potential for using CSF and PET measures as a potential screening tool. However, given the comparable AUC and positive predictive value for the combined CSF, PET and MRI measures compared with the MRI measures alone (both for the training and validation cohorts), our results suggest that further studies are needed to determine whether the diagnostic benefits of using multiple modalities outweigh the costs.

Why do morphometric measures of atrophy better predict disease progression than cellular or metabolic markers of pathology? One explanation may involve the temporal evolution of the AD disease process where $A\beta$ and metabolic changes precede gray matter disturbances (Frisoni et al., 2010; Jack et al., 2010). Recent studies have shown that amyloid deposition is present before the onset of cognitive decline (Sperling et al., 2009) and potentially contributes to metabolic and network-wide disruptions within cerebral cortical regions (Buckner et al., 2005, 2009; Hedden et al., 2009). As such, $A\beta$ and FDG-PET measures are likely more useful earlier in the disease process, possibly as a screening tool, than as a late predictive marker of MCI progression. Another explanation could be that automated MRI-based measures are a more stable indicator of long-term neuronal injury than CSF or FDG-PET markers. Structural MRI measures are not subject to diurnal variation, unlike proteins measured from the cerebrospinal fluid (De Leon et al., 2002) and likely reflect neuronal loss better than metabolic measures, which largely indicate fluctuations in synaptic activity and action potential propagation (Attwell and Laughlin, 2001).

The methods described here can be implemented in a clinical setting as a predictive marker for AD. Using these software tools a single volumetric baseline T1-weighted MRI scan can be completely processed, with minimal manual intervention. The factor loadings presented here can then be applied to the final output values of entorhinal cortex thickness, hippocampal volume, amygdala volume, temporal pole thickness, and parahippocampal gyrus thickness to quantify the 1, 2, and 3 year probability of a single individual progressing from MCI to AD.

The present study has limitations. This study was conducted on a carefully examined cohort of MCI individuals selected using criteria for inclusion in clinical trials. As

such, these results need to be validated on a larger, community-based, volunteer cohort that would be more representative of a clinical setting. A second limitation is the relatively short follow-up times, which could have resulted in including individuals with a greater degree of impairment thus increasing the likelihood of finding significant differences. Prior work from our group has shown that MRI measures can successfully predict progression to AD in milder MCI individuals with longer follow-up times (greater than 5 years) (Desikan et al., 2009a). Considered together, these results suggest that MRI measures have significant value as a predictive marker both in the early and later stages of disease progression. Another concern is that the smaller sample size of MCI individuals with PET and CSF measures may contribute to decreased power and non-significant effects for these two measures compared with the medial temporal factor. In an independent Cox model using only those MCI individuals with PET and CSF measures, the medial temporal factor continued to demonstrate a significant effect, comparable in magnitude (Hazards Ratio of 0.40) to the risk noted in the larger training (Hazards Ratio of 0.43) and validation cohorts (Hazards Ratio of 0.44) indicating that the main findings from this study are likely not subject to sample size related issues.

The early identification and prediction of those cognitively impaired individuals destined to develop Alzheimer's disease is of significant importance as therapies for altering the course of the illness or delaying dementia onset are developed. The results from this study demonstrates that automated MRI-based neuroanatomic measures can identify individuals in the earliest stages of the disease process and have promise in the clinical setting as a biomarker for tracking Alzheimer's disease progression.

Disclosure statement

The authors have no actual or potential conflicts of interest.

Acknowledgements

The authors thank Reisa Sperling for her insightful comments on this manuscript. This work was supported by grants from the National Center for Research Resources (P41-RR14075, R01 RR 16594-01A1 and the NCRB Birn Morphometric Project BIRN002, U24 RR021382), the National Institute for Biomedical Imaging and Bioengineering (R01 EB001550), the Mental Illness and Neuroscience Discovery (MIND) Institute, and the National Institute on Aging (P50 AG05681, P01 AG03991, and AG021910). Data collection and sharing for this project was funded by the Alzheimer's disease Neuroimaging Initiative (Alzheimer's disease Neuroimaging initiative (ADNI); Principal Investigator: Michael Weiner; NIH grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National

Institute of Biomedical Imaging and Bioengineering (NIBIB), and through generous contributions from the following: Pfizer, Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck & Co., Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation, plc, Forest Laboratories, and the Institute for the Study of Aging, with participation from the US Food and Drug Administration. Industry partnerships are coordinated through the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles.

References

- Apostolova, L.G., Dutton, R.A., Dinov, I.D., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M., 2006. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Arch Neurol* 63, 693–699.
- Arnold, S.E., Hyman, B.T., Flory, J., Damasio, A.R., Van Hoesen, G.W., 1991. The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cereb Cortex* 1, 103–116.
- Attwell, D., Laughlin, S.B., 2001. An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab* 21, 1133–1145.
- Bakkour, A., Morris, J.C., Dickerson, B.C., 2009. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. *Neurology* 72, 1048–1055.
- Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol (Berlin)* 82, 239–259.
- Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. 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* 23, 724–738.
- Buckner, R.L., Snyder, A.Z., Shannon, B.J., laRossa, G., Sachs, R., Fotenos, A.F., Sheline, Y.I., Klunk, W.E., Mathis, C.A., Morris, J.C., Mintun, M.A., 2005. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *J Neurosci* 25, 7709–7717.
- Buckner, R.L., Sepulcre, J., Talukdar, T., Krienen, F.M., Liu, H., Hedden, T., Andrews-Hanna, J.R., Sperling, R.A., Johnson, K.A., 2009. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J Neurosci* 29, 1860–1873.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- De Leon, M.J., Segal, S., Tarshish, C.Y., DeSanti, S., Zinkowski, R., Mehta, P.D., Convit, A., Caraos, C., Rusinek, H., Tsui, W., Saint Louis, L.A., deBernardis, J., Kerkman, D., Qadri, F., Gary, A., Lesbre, P., Wisniewski, T., Poirier, J., Davies, P., 2002. Longitudinal cerebrospinal fluid tau load increases in mild cognitive impairment. *Neurosci Lett* 333, 183–186.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for

- subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.
- Desikan, R.S., Cabral, H.J., Fischl, B., Guttman, C.R., Blacker, D., Hyman, B.T., Albert, M.S., Killiany, R.J., 2009a. Temporoparietal MR Imaging measures of atrophy in subjects with mild cognitive impairment that predict subsequent diagnosis of Alzheimer's disease. *AJNR Am J Neuroradiol* 30, 532–538.
- Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., Fischl, B., 2009b. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* 132, 2048–2057.
- Devanand, D.P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G.H., Honig, L.S., Mayeux, R., Stern, Y., Tabert, M.H., de Leon, M.J., 2007. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology* 68, 828–836.
- Fennema-Notestine, C., Gamst, A.C., Quinn, B.T., Pacheco, J., Jernigan, T.L., Thal, L., Buckner, R., Killiany, R., Blacker, D., Dale, A.M., Fischl, B., Dickerson, B., Gollub, R.L., 2007. Feasibility of multi-site clinical structural neuroimaging studies of aging using legacy data. *Neuroinformatics* 5, 235–245.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.
- Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Map* 8, 272–284.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci USA* 97, 11050–11055.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.
- Fleisher, A.S., Sun, S., Taylor, C., Ward, C.P., Gamst, A.C., Petersen, R.C., Jack, C.R., Jr, Aisen, P.S., Thal, L.J., 2008. Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment. *Neurology* 70, 191–199.
- Folstein, M., Folstein, S., McHugh, P., 1975. Mini-Mental State. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12, 189–198.
- Frisoni, G.B., Fox, N.C., Jack, C.R., Jr, Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol* 6, 67–77.
- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32, 180–194.
- Hedden, T., Van Dijk, K.R., Becker, J.A., Mehta, A., Sperling, R.A., Johnson, K.A., Buckner, R.L., 2009. Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. *J Neurosci* 29, 12686–12694.
- Jack, C.R., Jr, Petersen, R.C., Xu, Y.C., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Waring, S.C., Tangalos, E.G., Kokmen, E., 1999. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology* 52, 1397–1403.
- Jack, C.R., Jr, Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* 9, 119–128.
- Jagust, W.J., Landau, S.M., Shaw, L.M., Trojanowski, J.Q., Koeppe, R.A., Reiman, E.M., Foster, N.L., Petersen, R.C., Weiner, M.W., Price, J.C., Mathis, C.A. Alzheimer's Disease Neuroimaging Initiative, 2009. Relationships between biomarkers in aging and dementia. *Neurology* 73, 1193–1199.
- Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B., Fischl, B., 2009. 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* 46, 177–192.
- Karas, G., Sluimer, J., Goekoop, R., van der Flier, W., Rombouts, S.A., Vrenken, H., Scheltens, P., Fox, N., Barkhof, F., 2008. Amnesic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease. *AJNR Am J Neuroradiol* 29, 944–949.
- Kemper, T.L., 1994. Neuroanatomical and neuropathological changes in normal aging and in dementia In: Albert, M. Knofel, editors. *J. Clinical Neurology of Aging*. Oxford University Press, New York.
- Killiany, R.J., Hyman, B.T., Gomez-Isla, T., Moss, M.B., Kikinis, R., Jolesz, F., Tanzi, R., Jones, K., Albert, M.S., 2002. MRI measures of entorhinal cortex vs hippocampus in preclinical. *Adv Neurol* 58, 1188–1196.
- McKhann, G., Drachman, D., Folstein, M.F., Katzman, R., Price, D., Stadlan, E., 1984. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force. *Neurology* 34, 939–944.
- Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *Neuroimage* 44, 1415–1422.
- Morris, J.C., 1993. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 43, 2412–2414.
- Petersen, R., 2004. Mild cognitive impairment. *J Intern Med* 256, 183–194.
- Pett, M.A., Lackey, N.R., Sullivan, J.J., 2003. Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research. Sage, Thousand Oaks, CA.
- Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Démonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P. Alzheimer's Disease Neuroimaging Initiative, 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* 132, 2036–2047.
- Rosas, H.D., Liu, A.K., Hersch, S., Glessner, M., Ferrante, R.J., Salat, D.H., van der Kouwe, A., Jenkins, B.G., Dale, A.M., Fischl, B., 2002. Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58, 695–701.
- Salat, D.H., Buckner, R.L., Snyder, A.Z., Greve, D.N., Desikan, R.S., Busa, E., Morris, J.C., Dale, A.M., Fischl, B., 2004. Thinning of the cerebral cortex in aging. *Cereb Cortex* 14, 721–730.
- Shaw, L.M., Vanderstichele, H., Knopik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M., Trojanowski, J.Q., Alzheimer's Disease Neuroimaging Initiative, 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* 65, 403–413.
- Sperling, R.A., Laviolette, P.S., O'Keefe, K., O'Brien, J., Rentz, D.M., Pihlajamaki, M., Marshall, G., Hyman, B.T., Selkoe, D.J., Hedden, T., Buckner, R.L., Becker, J.A., Johnson, K.A., 2009. Amyloid deposition is associated with impaired default network function in older persons without dementia. *Neuron* 63, 178–188.
- Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack, C.R., Jr. Alzheimer's Disease Neuroimaging Initiative, 2009. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurology* 73, 294–301.
- Walhovd, K.B., Fjell, A.M., Amlie, I., Grambaite, R., Stenset, V., Bjørnerud, A., Reinvang, I., Gjerstad, L., Cappelen, T., Due-Tønnessen, P.,

- Fladby, T., 2009. Multimodal imaging in mild cognitive impairment: Metabolism, morphometry and diffusion of the temporal-parietal memory network. *Neuroimage* 45, 215–223.
- Wechsler, D., 1987. Wechsler Memory Scale-Revised. The Psychological Corporation, San Antonio.
- Whitwell, J.L., Przybelski, S.A., Weigand, S.D., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack, C.R., Jr, 2007. 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. *Brain* 130, 1777–1786.
- Whitwell, J.L., Shiung, M.M., Przybelski, S.A., Weigand, S.D., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack, C.R., Jr, 2008. MRI patterns of atrophy associated with progression to AD in amnesic mild cognitive impairment. *Neurology* 70, 512–520.
- Yuan, Y., Gu, Z.X., Wei, W.S., 2009. Fluorodeoxyglucose-positron-emission tomography, single-photon emission tomography, and structural MR imaging for prediction of rapid conversion to Alzheimer disease in patients with mild cognitive impairment: a meta-analysis. *AJNR Am J Neuroradiol* 30, 404–410.