RESEARCH ARTICLE

International Journal of Geriatric Psychiatry

Visual rating and volumetric measurement of medial temporal atrophy in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort: baseline diagnosis and the prediction of MCI outcome

Daniel Varon[1,2], Warren Barker[1], David Loewenstein[3], Maria Greig[1], Adriana Bohorquez[1], Isael Santos[1], Qian Shen[1], Molly Harper[1], Tatiana Vallejo-Luces[1], Ranjan Duara[1,2,3,4,5] and For the Alzheimer's Disease Neuroimaging Initiative †

[1]Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL, USA
[2]Department of Neurology, Herbert Wertheim College of Medicine, Florida International University, Miami, FL, USA
[3]Department of Psychiatry, University of Miami Miller School of Medicine, Miami, FL, USA
[4]Department of Medicine, University of Miami Miller School of Medicine, Miami, FL, USA
[5]Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA
Correspondence to: W. Barker, E-mail: billy@msmc.com

†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Objective: This study aims to determine the clinical utility of visual ratings and volumetric measurements of medial temporal atrophy among subjects from the Alzheimer's Disease Neurorimaging Initiative (ADNI) cohort.

Methods: A sample of 189 subjects from the ADNI, Phase 1 (ADNI-1), was chosen as follows: 49 cognitively normal (CN), 89 with mild cognitive impairment (MCI), and 50 with Alzheimer's disease (AD). Structural MRI images were downloaded from the ADNI website, and a visual rating system (VRS) was used to obtain semi-quantitative ratings of the hippocampus (HPC) and entorhinal cortex (ERC). VRS ratings and FreeSurfer measures of the HPC and ERC were used to predict (i) baseline diagnosis and (ii) progression to AD among subjects with MCI at baseline.

Results: VRS and FreeSurfer measures of ERC were equivalent in classifying subjects at baseline, but FreeSurfer measures of HPC were superior to VRS measures for classifying CN versus MCI subjects. VRS and FreeSurfer measures of both HPC and ERC were significant predictors of progression from MCI to AD. However, VRS ratings of ERC were superior to other MRI measures. MCI subjects with minimal ERC atrophy by VRS had a threefold lower progression rate to AD at 3.2 years compared with those with mild, moderate, or severe atrophy (23% vs 63%, 69%, and 87%, respectively).

Conclusions: Visual ratings of HPC and ERC provide useful information to a physician in a clinical setting. Visual ratings of ERC may be especially useful in following patients with MCI. Copyright © 2014 John Wiley & Sons, Ltd.

Key words: Alzheimer's; MCI; MRI; atrophy; hippocampus; entorhinal cortex
History: Received 1 February 2014; Accepted 25 March 2014; Published online in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/gps.4126

Introduction

In the long pre-clinical phase of Alzheimer's disease (AD), subtle cognitive deficits signal the beginning of a prodromal phase, associated with synaptic dysfunction, cell death, and regional brain atrophy most notably in the medial temporal lobe structures such as the entorhinal cortex (ERC) and hippocampus (HPC)

(Devanand *et al.*, 2007; Jauhiainen *et al.*, 2009). Structural MRI measurements of both HPC and ERC are highly correlated with observed clinical alterations in both the mild cognitive impairment (MCI) and dementia phases of AD (Desikan *et al.*, 2010; Jack *et al.*, 2013). In a hypothetical model of Jack *et al.* (2013) of AD biomarkers, measures of atrophy from structural MRI and hypometabolism from fluorodeoxyglucose positron emission tomography (FDG PET) are the last to become abnormal but the ones that most closely track with cognitive decline.

Volumes of the HPC and ERC have been quantified on 3-D MRI scans using computerized segmentation methods, such as Individual Brain Atlases using Statistical Parametric Mapping (IBASPM) (Shen *et al.*, 2011) and FreeSurfer (http://surfer.nmr.mgh.harvard.edu/; Eggert *et al.*, 2012). These volumetric methods are suitable for research studies but are not easily adaptable for routine clinical use. Scheltens developed a user-friendly visual rating scale to assess medial temporal atrophy (MTA) (Scheltens *et al.*, 1992; Scheltens *et al.*, 1995), using a 0–4 scale based on a single coronal slice posterior to the amygdala, at the level of the mamillary bodies. The visual assessment includes the "hippocampus proper, dentate gyrus, subiculum, parahippocampal gyrus, entorhinal cortex and surrounding cerebral spinal fluid (CSF) spaces such as temporal horn and choroid fissure" (Westman *et al.*, 2011). Scores on this scale correlated with volumetric measurements of HPC, discriminated normal controls from AD as well as volumetric methods (Wahlund *et al.*, 2000; Bresciani *et al.*, 2005; Westman *et al.*, 2011; Boutet *et al.*, 2012), and predicted conversion from MCI to AD (Lehmann *et al.*, 2013). However, it is uncertain whether visual assessment of HPC is equivalent (Visser *et al.*, 2002; Westman *et al.*, 2011) or inferior (Boutet *et al.*, 2012) to volumetric assessment in predicting conversion from MCI to AD.

Scheltens' method was expanded to include independent assessments of atrophy of the HPC, ERC, and perirhinal cortex (PRC) (Duara *et al.*, 2008) as part of a visual rating system (VRS). VRS measures of both HPC (VRS-HPC) and ERC (VRS-ERC) discriminated subjects with no cognitive impairment, amnestic MCI, and AD, and correlated with deficits in episodic memory among mildly impaired subjects (Loewenstein *et al.*, 2009). All VRS measures predicted the 1-year conversion from MCI to AD. A VRS measure of ERC was the best discriminator between MCI and normal control subjects (Urs *et al.*, 2009), but it is not known if VRS-ERC is better than VRS-HPC at predicting long-term progression to AD in subjects with MCI.

Current diagnostic clinical and research guidelines for "MCI due to AD" (Albert *et al.*, 2011) and "preclincial AD" (Sperling *et al.*, 2011) include the use of biomarkers as proxies for underlying pathology. These biomarkers include measures from CSF (tau and $A\beta_{1-42}$), FDG PET scans (glucose metabolism), amyloid PET scans (amyloid deposition), and structural MRI scans (regional atrophy). MRI scans are readily available but are used most commonly for ruling out potential causes of dementia other than AD and cognitive impairment, such as strokes, normal pressure hydrocephalus, or brain tumors. The use of VRS by clinicians could enhance the accuracy of a dementia diagnosis and provide useful information about the expected rate of progression to patients and their families.

In the current study, assessments of HPC and ERC using both volumetric (FreeSurfer) and VRS methods were obtained on subjects with normal cognition, MCI, and AD in the Alzheimer's Disease Neuroimaging Initiative, Phase 1 (ADNI-1), cohort. The main goal of this study was to contrast the performance of visual and volumetric ratings of HPC and ERC in the prediction of conversion from MCI to AD. The secondary goal was to provide information for clinicians to counsel patients and families.

Methods

Subjects

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and non-profit organizations as a $60m, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The principal investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California at San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the USA and Canada. The initial goal of ADNI was to recruit 800 subjects, but ADNI

has been followed by ADNI Grand Opportunity (ADNI-GO) and ADNI, Phase 2 (ADNI-2). To date, these three protocols have recruited over 1500 adults, ages 55 to 90 years, to participate in the research, consisting of cognitively normal (CN) older individuals, people with early or late MCI, and people with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org

At the baseline visit, ADNI participants were classified as CN or having MCI, based primarily on the Mini mental state examination (MMSE) (Folstein *et al.*, 1975), Clinical Dementia Rating (CDR) and CDR sum of boxes (CDR-SB) (Berg, 1988), and Delayed Paragraph Recall of Logical Memory II (LM-Del; Wechsler, 1987). CN and MCI patients had an MMSE score between 24 and 30. A CDR global score of 0 was required in CN participants, whereas a score of 0.5 was required of MCI subjects, and 1 or greater in AD patients. The thresholds of LM-Del for classifying CN and MCI subjects were based on age and education. An additional memory measure, the Rey Auditory Verbal Learning Test (RAVLT) immediate recall score (Rey, 1964), was administered at the baseline and follow-up visits.

A stratified convenience sample of 191 subjects was chosen from the ADNI-1 cohort by using the participants' baseline data. The following three participant characteristics were stratified: cognitive diagnosis (normal, MCI, and mild AD), age (≤75 years and >75 years), and sex (male and female). The relative sizes of each of the 12 strata (age by sex by cognitive diagnoses) were chosen to match the ratios from ADNI. Table 1 shows the demographic and clinical characteristics of the sample. The images used in the study were downloaded from the Laboratory of Neuro Imaging (LONI) on 24 October 2011, whereas the clinical and imaging results were downloaded on 27 June 2013.

MRI image and volume acquisition

At the baseline visit, structural MRI scans were acquired from 1.5T scanners at multiple sites across the USA and Canada. MRI protocols were performed across a variety of scanners such as GE, Siemens, or Philips to ensure comparability. A 3-D sagittal magnetization prepared rapid gradient-echo (MPRAGE) imaging sequence was used. MRI volumes were computed using FreeSurfer version 4.3.0 by the University of California at San Francisco and downloaded in an Excel file from LONI, as described earlier. The regions used in this study were the left (ST29SV) and right HPC (ST88SV) and the left (ST83CV) and right ERC (ST24CV). The left and right ERC measures were summed, and the left and right hippocampal volumes were summed, to create bilateral volumes for HPC and ERC. All regional volumes were normalized to intracranial volume (ICV).

Visual rating system

VRS ratings were performed blind to the subjects' diagnosis and demographic information, on a standardized

Table 1 Demographic and baseline clinical data

Diagnosis	Cognitively normal ($n=50$)	MCI ($n=89$)	Alzheimer ($n=52$)	p-value
Age (years)	73.0 ± 4.1	70.5 ± 7.3	71.8 ± 6.7	ns
Gender (% female)	42	51	50	ns
Education	15.9 ± 4.1	15.7 ± 3.1	14.6.9 ± 2.1	ns
Race (% white)	88	93	90	ns
MMSE	29.1^{a} ± 1.2	27.1^{b} ± 1.8	23.6^{c} ± 1.9	<0.0001
CDR sum of boxes	0.0^{a} ± 0.1	1.7^{b} ± 0.9	4.3^{c} ± 1.5	<0.0001
Logical Memory Delayed Recall	13.7^{a} ± 3.7	3.1^{b} ± 3.4	1.5^{c} ± 2.3	<0.0001
APOE εPOE 1l Memory Delayed (%)	19	34	50	0.001
APOE-ε4+ (% with an e4 allele)				
Hippocampus				
VRS*	2.3^{a} ± 1.6	3.6^{b} ± 2.3	5.2^{c} ± 2.1	<0.0001
Volume**	0.49^{a} ± 0.06	0.41^{b} ± 0.07	0.38^{c} ± 0.06	<0.0001
Entorhinal cortex				
VRS*	1.9^{a} ± 1.9	3.9^{b} ± 2.4	5.3^{c} ± 2.1	<0.0001
Volume**	0.25^{a} ± 0.04	0.21^{b} ± 0.05	0.18^{c} ± 0.05	<0.0001

MCI, mild cognitive impairment; MMSE, Mini mental state examination; CDR, Clinical Dementia Rating; *APOE* = Apolipoprotein E; VRS, visual rating system; ns, not significant.

[a–c]Means with different letters are significantly different by post hoc Scheffe test.

*Average of right plus left visual rating of hippocampus or entorhinal cortex.

**Hippocampus or entorhinal volume (right plus left), as percent of intracranial volume.

coronal slice, perpendicular to the line joining the anterior and posterior commissures, intersecting the mammillary bodies and on adjacent slices. In VRS, ratings for HPC, ERC, and PRC are obtained in each hemisphere. However, PRC ratings were not included in this paper because it was not possible to rate several of the images for PRC. In prior studies using VRS, the average time to choose the proper images and do ratings for HPC, ERC, and PRC was 5 to 6 min per subject (Urs *et al.*, 2009).

VRS ratings are based on a 5-point scale: 0 = no atrophy, 1 = minimal atrophy, 2 = mild atrophy, 3 = moderate atrophy, and 4 = severe atrophy (Figure 1). A library of reference images that defines the anatomical boundaries of each brain structure and depicts different levels of atrophy is provided from a drop-down menu in the system interface; this facilitates a direct comparison of each structure on the subject's MRI. Excellent inter-rater (kappa, 0.75 to 0.94) and intra-rater (kappa, 0.84 to 0.94) agreement for VRS ratings of atrophy in HPC and ERC has been reported (Duara *et al.*, 2008; Urs *et al.*, 2009). For this study, a neurologist/psychiatrist (Daniel Varon) who had 4 years of experience with VRS and who was blind to the diagnosis rated all of the images downloaded from LONI. A training presentation of VRS is available online (http://www.mcisymposium.org/vrs.pps). VRS ratings for right and left were added to create an *average* VRS score for HPC and for ERC. Average atrophy for each structure was classified as minimal (less or equal to 0.5), mild (more than 0.5 and less than or equal to 1), moderate (more than 1 and less than or equal to 2), or severe (more than 2).

Statistical analysis. All analyses were performed using SAS Version 9.1 (SAS Institute, Cary, NC, USA). Group comparisons of means were analyzed using a general linear model. The Scheffe *post hoc* procedure was used to examine differences between groups. Proportions were examined using chi-square analyses. Logistic regression was used to assess the accuracy of VRS and volumetric measures of HPC and ERC for the classification of (i) CN versus MCI subjects and (ii) CN versus AD subjects. The area under the receiver operating characteristic (AuROC) curve was used to compare VRS and volumetric assessments for the aforementioned classifications. Specificity of the VRS and volumetric measures for classifying the baseline diagnosis (MCI and AD versus CN) were computed for the cut point corresponding to a minimum sensitivity of 80%, to provide a more practical understanding of the ROC curves.

Conversion rates and progression curves from MCI to AD were estimated using the Kaplan–Meier product limit from the procedure LIFETEST in SAS. Predictors of conversion from MCI to AD were evaluated using univariate and multivariate Cox regression models implemented with the procedure PHREG in SAS. Time intervals of 1.2, 2.2, 3.2, 4.2, and 5.2 years were used to accommodate the timing of the annual ADNI assessments, which were usually completed within 1 to 2 months of the anniversary of the initial visit. The *p*-value for statistical significance was set at 0.05.

Results

Baseline characteristics. The mean age of the 191 subjects was 71.5 ± 6.5 years, with a range of 55 to 81 years. There were no significant differences between the three diagnostic groups, with respect to age, gender, education, or race. MCI subjects scored

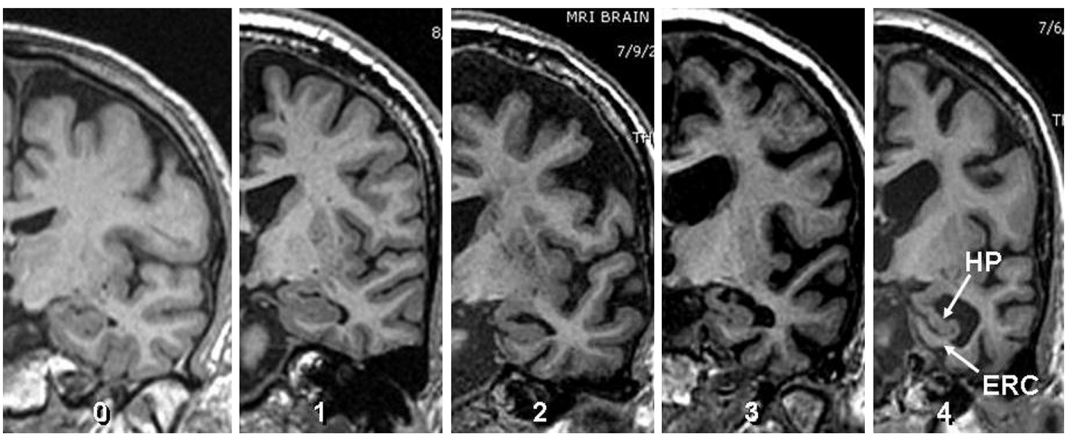

Figure 1 Images depicting four degrees of atrophy in hippocampus and entorhinal cortex according to Visual Rating Scale where 0 = no atrophy, 1 = minimal atrophy, 2 = mild atrophy, 3 = moderate atrophy, and 4 = severe atrophy (score shown corresponds to both structures). Duara *et al. Neurology* (2008) 71:1986–92.

significantly worse on the MMSE, CDR-SB, and LM-Del compared with CN subjects, but significantly better than AD subjects (Table 1). Volumetric and VRS ratings of HPC and ERC revealed more atrophy among AD compared with MCI subjects, and among MCI subjects compared with CN subjects.

Cross-sectional classification of CN versus MCI using HPC and ERC. ***HPC.*** The AuROC was greater for volumetric than VRS assessment of HPC (Table 2) (0.78 vs. 0.66; $\chi^2 = 11.6$, $p = 0.0006$). A VRS threshold score of 2 or greater (right plus left) corresponded to the highest specificity at a minimum sensitivity of 80%. This threshold resulted in a sensitivity/specificity of 83%/26% for the detection of MCI using VRS-HPC. For the volumetric measure of HPC, a sensitivity of 83% corresponded to a specificity of 46% at a cutoff of 0.00476 (volume normalized to ICV).

ERC. There were no significant differences between VRS and volumetric measures in the discrimination of normal from MCI subjects, as reflected by the AuROCs. A VRS score of 2 or greater (right plus left) corresponded to a sensitivity/specificity of 84%/40%, whereas an FS-ERC volume of 0.00256 (normalized to ICV) resulted in a sensitivity/specificity of 84%/50%. VRS-ERC was not different from VRS-HPC in the discrimination of CN from MCI.

Cross-sectional classification of CN versus AD using HPC and ERC. ***HPC.*** The AuROCs for VRS and volumetric measures were not significantly different. A VRS-HPC threshold score of 3 or greater (right plus left) corresponded to a sensitivity/specificity of 83%/56%. For volumetric measures of HPC (right plus left), the sensitivity/specificity was 83%/78% at a threshold of 0.00423 (normalized to ICV).

ERC. There was no difference in AuROCs for VRS and volumetric measures. A VRS-ERC threshold score of 3 or greater (right plus left) corresponded to a sensitivity/specificity of 83%/72%. Volumetric measures of ERC, at a threshold of 0.00223 (normalized to ICV), had a sensitivity/specificity of 83%/74%. VRS measures of ERC and HPC were not different in their ability to discriminate CN from AD.

Predictors of conversion from MCI to AD. Predictors of conversion to AD were evaluated using univariate and multivariate proportional Cox regression models, which included these independent predictor variables: CDR-SB, RAVLT, and both volumetric and VRS measures of atrophy of HPC and ERC (Table 3). Age was not included because it was not significant as a univariate predictor of conversion ($\chi^2 = 0.03$, $p = 0.98$). In univariate models, all of the aforementioned variables were significant predictors of conversion to AD. In full models with either (i) VRS-ERC and VRS-HPC or (ii) VRS-ERC and volumetric ERC, when the statistical significance of each variable was adjusted for the other variables, only VRS-ERC was entered into the model. In two full models with VRS-ERC and either (i) CDR-SB or (ii) RAVLT, all variables were entered.

Conversion rates from MCI to AD by HPC-ERC, VRS-ERC, and RAVLT. In the entire sample of MCI subjects, the estimated conversion rate to AD was 75% (95% CI: 63–86%) by the year 5 visit (Table 4). For the four strata of VRS-ERC, the progression curves were significantly different (log-rank test $\chi^2 = 14.3$; 3 degrees of freedom [df], $p = 0.003$) (Figure 2). The conversion rates at 3.2 years of follow-up ranged from 23% (95% CI of 8–55%) among subjects in the lowest ERC atrophy strata to 87% (95% CI of 65–97%) among subjects with the highest atrophy. *Post hoc* testing revealed that subjects with minimal atrophy had a

Table 2 Cognitive classification at baseline: comparison of volumetric and visual rating measures of hippocampus and entorhinal cortex

	Visual rating system		Volumetric measures	
Region	AuROC	Sensitivity/specificity (cutoff)	AuROC	Sensitivity/specificity
Cognitively normal versus MCI				
Hippocampus (R + L)	0.66[a]	83%/26% (≥2)	0.78[b]	83%/54%
Entorhinal cortex (R + L)	0.73	84%/40% (≥2)	0.75	84%50/%
Cognitively normal versus AD				
Hippocampus (R + L)	0.84	83%/56% (≥3)	0.89	83%/78%
Entorhinal cortex (R + L)	0.87	85%/72% (≥3)	0.86	85%/74%

Superscripts refer to differences between visual rating system and volumetric measures for AuROC at $p = 0.05$. The sensitivity/specificity was chosen at the sensitivity closest to 80%. AuROC, area under the receiver operating curve; MCI, mild cognitive impairment; AD, Alzheimer's disease.

Table 3 Predictors of conversion from MCI to AD: cognitive measures and atrophy in the hippocampus and entorhinal cortex

Model Predictors	Univariate model χ^2	Univariate model $p > \chi^2$	Full model χ^2	Full model $p > \chi^2$
Hippocampal atrophy				
Volumetric	6.1	0.01	-	ns
VRS	4.3	0.04	-	ns
Entorhinal cortex atrophy				
Volumetric	7.7	0.006	-	ns
VRS	11.2	0.0008	4.6	0.03
VRS				
Entorhinal cortex	11.3	0.0008	8.9	0.003
Hippocampus	4.3	0.04	-	ns
VRS-ERC & CDR Sum of Boxes				
VRS of Entorhinal Cortex	11.3	0.0008	9.4	0.002
CDR sum of boxes	8.9	0.003	6.5	0.01
VRS-ERC & RAVLT				
VRS of Entorhinal Cortex	11.3	0.0008	5.8	0.02
RAVLT	16.6	<0.0001	11.3	0.0008

The table shows the results of proportional hazard models. VRS, Visual Rating System; ERC, entorhinal cortex; CDR, Clinical Dementia Rating Scale; ns, not significant; RAVLT, Rey Auditory Verbal Learning Test; ERC, entorhinal cortex.

slower progression to AD compared with those with mild atrophy (log-rank test $\chi^2 = 5.2$; 1 df, $p = 0.02$), moderate atrophy (log-rank test $\chi^2 = 8.3$; 3 df, $p = 0.004$), and severe atrophy (log-rank test $\chi^2 = 16.7$; 3 df, $p < 0.0001$). The progression curves for the four strata of VRS-HPC were not different (log-rank test $\chi^2 = 6.1$; 3 df, $p = 0.11$). However, an unplanned analysis showed that subjects with minimal HPC atrophy had a slower progression than all other subjects combined (log-rank test $\chi^2 = 5.9$; 1 df, $p = 0.02$). For RAVLT strata, the conversion rates at 3.2 years ranged from 18% for those who scored at least 38 on the RAVLT to 79% for those who scored less than 28.

Discussion

Consolidated visual rating scores for MTA have been used to predict the baseline diagnosis of subjects with memory loss (Devanand *et al.*, 2007; Duara *et al.*, 2008; Shen *et al.*, 2011; Lehmann *et al.*, 2013). This study extends those findings, by directly comparing standard volumetric and visual rating assessments of HPC and ERC. The volumetric method of assessing HPC was better than the VRS method in discriminating CN from MCI subjects, but volumetric and VRS measures of both HPC and ERC were comparable in discriminating CN from AD subjects at baseline. While VRS and volumetric measures for ERC and HPC were significant predictors of conversion from MCI to AD, VRS-ERC outperformed both VRS-HPC and volumetric ratings of ERC. Further, VRS-ERC added significantly to CDR-SB and RAVLT scores as predictors of conversion from MCI to AD.

These findings are consistent with a previous report in a clinic sample (Varon *et al.*, 2011), which showed that minimal atrophy of the ERC, as measured by

Table 4 Estimated conversion rates[a] from MCI to AD by rating of atrophy of the entorhinal cortex and hippocampus and by scores on a memory measure

Predictor	*N*	1.2 years	2.2 years	3.2 years	4.2 years	5.2 years
All subjects	89	20% (12–30)	47% (36–58)	62% (51–74)	69% (57–80)	75% (63–86)
Mean VRS-ERC Rating						
[0, 0.5]	14	0% (0–0)	7% (1–41)	23% (8–55)	31% (13–64%)	40% (13–64)
(0.5, 1]	23	17% (6–44)	48% (27–74)	63% (40–85)	72% (48–92)	b
(1, 2]	26	31% (17–52)	54% (35–75)	69% (47–89)	69% (47–89)	90% (65–99)
(2, 4]	26	21% (9–43)	65% (45–84)	87% (65–97)	b	b
Mean VRS-HPC Rating						
[0, 0.5]	15	7% (1–41)	21% (7–53%)	30% (12–72)	39% (18–70)	48% (25–77)
(0.5, 1]	30	22% (11–43)	58% (40–77)	67% (49–85)	74% (54–90)	74% (54–90)
(1, 2]	25	30% (16–53)	42% (24–65)	77% (54–94)	84% (62–97)	b
(2, 4]	19	11% (3–38%)	58% (34–82)	68% (42–90)	68% (42–90)	84% (53–99)
RAVLT score						
[38–51]	17	6% (1–37)	6% (1–37)	18% (4–58)	18% (4–58)	34% (11–77)
[32–38)	16	20% (7–50)	50% (28–78)	60% (35–86)	60% (35–86)	60% (35–86)
[28–31]	25	10% (2–33)	49% (30–73)	72% (51–89)	72% (51–89)	79% (58–94)
<28	34	33% (19–53)	64% (46–81)	79% (60–92)	95% (79–100)	b

[a]Conversion rates expressed as % (95% confidence interval). Mean visual rating system-entorhinal cortex and visual rating system-hippocampus scores are the average of right and left ratings. Rey Auditory Verbal Learning Test (RAVLT) thresholds were derived from the mean and standard deviation (SD) among the 50 cognitively normal subjects: 38 = mean – 1 SD; 32 = Mean – 1.68 SD; 28 = M – 2 SD.
[b]No more converters after this point in time.

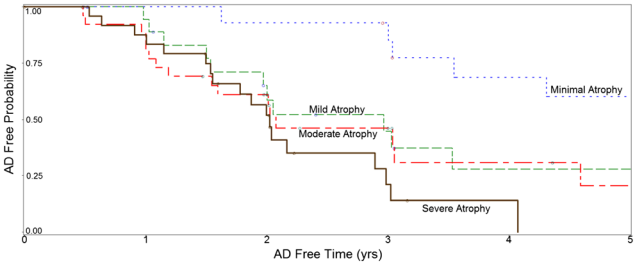

Figure 2 Dementia-free survival: progression from mild cognitive impairment (MCI) to Alzheimer's disease (AD) by visual rating system (VRS)-entorhinal cortex score. Probability of subjects with MCI not converting to AD by four levels of atrophy of the entorhinal cortex assessed by VRS: minimal or no atrophy (short dashed line), mild atrophy (medium dashed line), moderate atrophy (medium and short dashed line), and severe atrophy (solid line). Confidence intervals for the curves are in Table 4. The curves are different by log-rank test (χ^2=14.3; 2 df, p=0.003).

VRS, resulted in greater impairment and a faster rate of progression than minimal atrophy of the HPC (Varon *et al.*, 2011). At least two other reports using volumetric methods have also shown that ERC volume, compared with HPC volume, may be a better predictor of outcome in people with MCI (Killiany *et al.*, 2002; Bakkour *et al.*, 2009). These results are in agreement with the general consensus that the earliest neurodegenerative changes in AD occur in most typical cases in the transentorhinal cortex and then progress to other regions (Braak and Braak, 1991). HPC appears to be susceptible to different types of processes common in older subjects such as depression, steroid use, stroke, brain trauma, and seizures, which could make the atrophy seen in this structure less specific during early stages of the disease (Geuze *et al.*, 2005).

One of the goals of this investigation was to provide guidance to physicians in a clinical setting, by providing VRS cutoffs that predict higher or lower likelihood of conversion from MCI to AD. Among subjects who had low VRS scores for the ERC and HPC, more than half remained at an MCI level by 5 years, compared with only 13% of the subjects with high VRS scores (scores of 2–4) (Table 4). At the level of an individual patient, the ability of MTA measures to predict conversion to AD is modest—the sensitivity/specificity of VRS-ERC is 91%/46% at 2 years (data not shown in results). Even when multiple predictors from ADNI subjects were combined (atrophy on MRI, episodic memory scores, and cerebral spinal fluid markers), the sensitivity/specificity in predicting conversion from MCI to AD at 3 years was only 73%/71% (Liu *et al.*, 2013). Nonetheless, MTA changes noted in a patient's MRI should be considered by physicians, just as cognitive reserve, memory scores, diet, or medical issues are regarded as variables that may delay or accelerate symptoms in AD.

Some limitations of this study should be acknowledged. The MRI images used to do the volumetric measurements in ADNI were acquired in the sagittal plane; however, the VRS method requires coronal images to do the ratings. Thus, the images were reconstructed to provide the orientation required for the VRS. The reconstructed images have less resolution, which may limit the accuracy of some of the ratings. It would be advisable when using the VRS in the clinical setting to obtain an MRI with thin coronal slices using an MPRAGE sequence or similar, which will serve to do the visual ratings and if desired can also be used to acquire volumetric measurements. Another limitation is that FreeSurfer may not be as accurate as other volumetric methods (Eggert *et al.*, 2012).

The characteristics of this sample from the ADNI-1 cohort is an additional limitation in generalizing the results of the study. The majority of subjects in this sample were between 65 and 75 years of age, so the thresholds for atrophy from this study may not be optimal for younger or older subjects. Pereira *et al.* (2014) showed that using different atrophy cutoffs on the Scheltens MTA scale for subjects above or below 75 years of age resulted in better classification of subjects' cognitive status, compared with using a single threshold. Higher thresholds may be warranted among older subjects, because "normal aging" is associated with brain atrophy (Raji *et al.*, 2012). This sample is also highly selected in regard to education (mean of 4 years of college) and race (over 90% constituted primarily by Caucasians), and non-amnestic MCI subjects were excluded. Previous reports have shown that non-amnestic subjects resemble CN individuals on the VRS rating of HPC and ERC (Duara *et al.*, 2008). This should be kept in mind when evaluating subjects in the clinical setting where some patients who present with non-amnestic MCI may actually represent individuals with hippocampal sparing forms of AD (Murray *et al.*, 2011) and may not develop MTA until much later when compared with typical cases. Finally, our study did not include subjects with subjective cognitive impairment or early MCI because they were not part of ADNI-1. Those populations have now been included in ADNI-GO and ADNI-2; however, they were not used because of the limited longitudinal data available at the time this study was completed. Future studies would greatly benefit from including such populations, which more accurately resemble patients usually seen in the clinical setting.

MRI scans are frequently used in a dementia workup, but the assessment of regional neurodegeneration to characterize the likelihood of AD in patients with MCI is not commonly carried out, outside of specialized centers. Questions frequently posed by patients with MCI and their families in regard to etiology and likelihood of progression often go unanswered. The visual assessment of MTA on MRI could serve as a tool to assist clinicians with an initial approximation to an etiological characterization of MCI and early dementia, as outlined in the NIA-AA criteria (Albert *et al.*, 2011). Physicians who care for patients with cognitive disorders may be able to use methods such as VRS to provide better counseling for patients with MCI.

Conflict of interest

None declared.

Key point

- Visual rating of hippocampal and entorhinal atrophy on MRI has utility in predicting the progression of MCI.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). 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 for Neuro Imaging at the University of Southern California.

This work was supported by the State of Florida Alzheimer's Disease Initiative, Department of Elder Affairs, Tallahassee, Florida.

References

Albert MS, Dekosky ST, Dickson D, *et al.* 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**: 270–279.

Bakkour A, Morris JC, Dickerson BC. 2009. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. *Neurology* **72**: 1048–1055.

Berg L. 1988. Clinical Dementia Rating (CDR). *Psychopharmacol Bull* **24**: 637–639.

Boutet C, Chupin M, Colliot O, *et al.* 2012. Is radiological evaluation as good as computer-based volumetry to assess hippocampal atrophy in Alzheimer's disease? *Neuroradiology* **54**: 1321–1330.

Braak H, Braak E. 1991. Neuropathological stageing of Alzheimer's-related changes. *Acta Neuropathol (Berl)* **82**: 239–259.

Bresciani L, Rossi R, Testa C, *et al.* 2005. Visual assessment of medial temporal atrophy on MR films in Alzheimer's disease: comparison with volumetry. *Aging Clin Exp Res* **17**: 8–13.

Desikan RS, Cabral HJ, Settecase F, *et al.* 2010. Automated MRI measures predict progression to Alzheimer's disease. *Neurobiol Aging* **31**: 1364–1374.

Devanand DP, Pradhaban G, Liu X, *et al.* 2007. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology* **68**: 828–836.

Duara R, Loewenstein DA, Potter E, *et al.* 2008. Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer's disease. *Neurology* **71**: 1986–1992.

Eggert LD, Sommer J, Jansen A, Kircher T, Konrad C. 2012. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. *PLoS ONE* **7**: e45081.

Folstein MF, Folstein SE, McHugh PR. 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**: 189–198.

Geuze E, Vermetten E, Bremner JD. 2005. MR-based in vivo hippocampal volumetrics: 2. Findings in neuropsychiatric disorders. *Mol Psychiatry* **10**: 160–184.

Jack CR, Knopman DS, Jagust WJ, *et al.* 2013. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* **12**: 207–216.

Jauhiainen AM, Pihlajamäki M, Tervo S, *et al.* 2009. Discriminating accuracy of medial temporal lobe volumetry and fMRI in mild cognitive impairment. *Hippocampus* **192**: 166–175.

Killiany RJ, Hyman BT, Gomez-Isla T, *et al.* 2002. MRI measures of entorhinal cortex vs. hippocampus in preclinical AD. *Neurology* **58**: 1188–1196.

Lehmann M, Koedam EL, Barnes J, *et al.* 2013. Visual ratings of atrophy in MCI: prediction of conversion and relationship with CSF biomarkers. *Neurobiol Aging* **34**: 73–82.

Liu Y, Mattila J, Ruiz MAM, *et al.* 2013. Predicting AD conversion: comparison between prodromal AD guidelines and computer assisted PredictAD tool. *PLoS ONE* **8**: e55246.

Loewenstein DA, Acevedo A, Potter E, *et al.* 2009. Severity of medial temporal atrophy and amnestic mild cognitive impairment: selecting type and number of memory tests. *Am J Geriatr Psychiatry* **17**: 1050–1058.

Murray ME, Graff-Radford NR, Ross OA, *et al.* 2011. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* **10**: 785–796.

Pereira JB, Cavallin L, Spulber G, Aguilar C, *et al.* 2014. Influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cut-offs. *J Intern Med* **275**: 317–330.

Raji CA, Lopez OL, Kuller LH, *et al.* 2012. White matter lesions and brain gray matter volume in cognitively normal elders. *Neurobiol Aging* **33**: 834.e7–834.e16.

Rey A. 1964. L'examen clinique en psychologie'. Presses Univeritaires de France: Paris.

Scheltens PH, Leys D, Barkhof F, *et al*. 1992. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal aging: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* **55**: 967–972.

Scheltens P, Launer LJ, Barkhof F, *et al*. 1995. Inter-observer reliability of visual assessment of hippocampal atrophy on MRI. *J Neurol* **242**: 557–560.

Shen Q, Loewenstein DA, Potter E, *et al*. 2011. Volumetric and visual rating of MRI scans in the diagnosis of amnestic MCI and Alzheimer's disease. *Alzheimers Dement* **8**: 399–406.

Sperling RA, Aisen PS, Beckett LA, *et al*. 2011. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging—Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**: 280–292.

Urs R, Potter E, Barker W, *et al*. 2009. Visual rating system (VRS) for assessing magnetic resonance images (MRIs): a tool in the diagnosis of MCI and Alzheimer's disease. *J Comput Assist Tomogr* **33**: 73–78.

Varon D, Loewenstein DA, Potter E, *et al*. 2011. Minimal atrophy of the entorhinal cortex and hippocampus: progression of cognitive impairment. *Dement Geriatr Cogn Disord* **31**: 276–283.

Visser PJ, Verhey FRJ, Hofman PAM, Scheltens P, Jolles J. 2002. Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. *J Neurol Neurosurg Psychiatry* **72**: 491–497.

Wahlund L-O, Julin P, Johansson S-E, *et al*. 2000. Visual rating and volumetry of the medial temporal lobe on magnetic resonance imaging in dementia: a comparative study. *J Neurol Neurosurg Psychiatry* **69**: 630–635.

Wechsler D. 1987. The Wechsler Memory Scale—Revised. The Psychological Corporation: San Antonio, TX.

Westman E, Cavallin L, Muehlboeck J-S, *et al*. 2011. Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. *PLoS ONE* **6**: e22506.