Original Article

Journal of INTERNAL MEDICINE

doi: 10.1111/joim.12358

Practical cut-offs for visual rating scales of medial temporal, frontal and posterior atrophy in Alzheimer's disease and mild cognitive impairment

D. Ferreira[1], L. Cavallin[2,3], E.-M. Larsson[4], J.-S. Muehlboeck[1], P. Mecocci[5], B. Vellas[6], M. Tsolaki[7], I. Kłoszewska[8], H. Soininen[9], S. Lovestone[10], A. Simmons[11,12,13], L.-O. Wahlund[1], E. Westman[1] & for the AddNeuroMed consortium and the Alzheimer's Disease Neuroimaging Initiative*

From the [1]Department of Neurobiology, Care Sciences and Society, Center for Alzheimer Research, Division of Clinical Geriatrics; [2]Department of Clinical Science, Intervention and Technology, Division of Medical Imaging and Technology, Karolinska Institutet; [3]Department of Radiology, Karolinska University Hospital, Stockholm; [4]Department of Radiology, Oncology and Radiation Science, Uppsala University, Uppsala, Sweden; [5]Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy; [6]INSERM U 558, University of Toulouse, Toulouse, France; [7]3rd Department of Neurology, Aristoteleion Panepistimeion Thessalonikis, Thessaloniki, Greece; [8]Medical University of Lodz, Lodz, Poland; [9]University of Eastern Finland, University Hospital of Kuopio, Kuopio, Finland; [10]Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford; [11]Institute of Psychiatry, King's College London; [12]NIHR Biomedical Research Centre for Mental Health; and [13]NIHR Biomedical Research Unit for Dementia, London, UK

Abstract. Ferreira D, Cavallin L, Larsson E-M, Muehlboeck J-S, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Simmons A, Wahlund L-O, Westman E; for the AddNeuroMed consortium and the Alzheimer's Disease Neuroimaging Initiative (Karolinska Institutet, Stockholm; Karolinska Institutet, Stockholm; Karolinska University Hospital, Stockholm; Uppsala University, Uppsala, Sweden; Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy; INSERM U 558, University of Toulouse, Toulouse, France; Aristoteleion Panepistimeion Thessalonikis, Thessaloniki, Greece; Medical University of Lodz, Lodz, Poland; University of Eastern Finland, University Hospital of Kuopio, Kuopio, Finland; University of Oxford, Oxford, Institute of Psychiatry, King's College London, London; NIHR Biomedical Research Centre for Mental Health, London; NIHR Biomedical Research Unit for Dementia, London, UK). Practical cut-offs for visual rating scales of medial temporal, frontal and posterior atrophy in Alzheimer's disease and mild cognitive impairment. *J Intern Med* 2015; doi: 10.1111/joim.12358.

Background. Atrophy in the medial temporal lobe, frontal lobe and posterior cortex can be measured with visual rating scales such as the medial temporal atrophy (MTA), global cortical atrophy – frontal subscale (GCA-F) and posterior atrophy (PA) scales, respectively. However, practical cut-offs are urgently needed, especially now that different presentations of Alzheimer's disease (AD) are included in the revised diagnostic criteria.

Aims. The aim of this study was to generate a list of practical cut-offs for the MTA, GCA-F and PA scales, for both diagnosis of AD and determining prognosis in mild cognitive impairment (MCI), and to evaluate the influence of key demographic and clinical factors on these cut-offs.

Methods. AddNeuroMed and ADNI cohorts were combined giving a total of 1147 participants (322 patients with AD, 480 patients with MCI and 345 control subjects). The MTA, GCA-F and PA scales were applied and a broad range of cut-offs was evaluated.

Results. The MTA scale showed better diagnostic and predictive performances than the GCA-F and PA scales. Age, apolipoprotein E (*ApoE*) ε4 status and age at disease onset influenced all three scales. For the age ranges 45–64, 65–74, 75–84 and 85–94 years, the following cut-offs should be used. MTA: ≥1.5, ≥1.5, ≥2 and ≥2.5; GCA-F, ≥1, ≥1, ≥1 and ≥1; and PA, ≥1, ≥1, ≥1 and ≥1, respectively, with an adjustment for early-onset *ApoE* ε4 non-carrier AD patients (MTA: ≥2, ≥2, ≥3 and ≥3; and GCA-F: ≥1, ≥1, ≥2 and ≥2, respectively).

Conclusions. If successfully validated in clinical settings, the list of practical cut-offs proposed here might be useful in clinical practice. Their use might also (i) promote research on atrophy subtypes, (ii)

*Data used in the 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 the ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

increase the understanding of different presentations of AD, (iii) improve diagnosis and prognosis and (iv) aid population selection and enrichment for clinical trials.

Keywords: Alzheimer's disease, frontal atrophy, medial temporal atrophy, mild cognitive impairment, posterior atrophy, visual rating scales.

Introduction

Current diagnostic criteria for Alzheimer's disease (AD) and mild cognitive impairment (MCI) incorporate biomarkers of neurodegeneration to increase diagnostic certainty [1, 2]. Neurodegeneration can be studied with magnetic resonance imaging (MRI), and medial temporal atrophy (MTA) is a supportive feature of AD [1, 3]. Atrophy may also occur in other brain regions such as the frontal lobe and posterior cortex, associated with AD presentations demonstrating prominent executive, language or visuospatial impairment. These presentations are included in the current diagnostic criteria for AD. It is thus relevant to consider clinically suitable tools for the assessment of frontal and posterior atrophy, in addition to MTA.

The most widely used tool for assessing MTA is the visual rating scale developed by Scheltens *et al.* [4], which has been successfully validated for AD (for review see [5]). Posterior atrophy (PA) can be assessed with a recently published visual rating scale also validated for AD [6, 7]. However, no visual rating scale of frontal atrophy has been specifically validated for AD, and only adaptations are available for other neurological states [8, 9]. The global cortical atrophy scale (GCA) [10, 11] has previously been applied in AD [11–17] and has shown greater global atrophy in patients with AD than in those with MCI and control subjects [12]. The GCA scale includes a separate assessment of the frontal lobe (i.e. GCA-F), and two previous studies have evaluated frontal atrophy in AD based on the GCA-F scale [11, 14].

The MTA, PA and GCA-F scales are quick and easy to use. They can be performed on both MRI and computed tomography images, which are generally available today as part of the clinical routine. However, these three scales have not yet been widely used in clinical practice. One of the reasons for this is the lack of reliable and practical cut-offs to determine deviation from normality [5, 18]. For the MTA scale, two cut-offs have commonly been used. The original proposal by Scheltens and colleagues consisted of an age-adjusted cut-off of MTA score ≥2 in either of the two hemispheres considered abnormal below the age of 75 years, with an MTA score ≥3 required to be abnormal above 75 years of age [4]. This cut-off has shown low specificity (67%) [19]. A cut-off of ≥1.5 based on the average score of both hemispheres has also been proposed [20]. This cut-off has shown higher specificity (94%), but at the expense of lower sensitivity (46%). Further, a cut-off of ≥2 based on the average score of both hemispheres has been reported to have higher sensitivity in patients with AD older than 75 years [21], and a score of 3 might be considered normal in individuals above 80 years of age [22]. For PA, a cut-off of ≥2 has been proposed by Koedam *et al.* [7]. To our knowledge, no cut-offs have been specified for the GCA or the GCA-F subscale, and their diagnostic performance remains unknown.

In addition to age, other factors such as apolipoprotein E (*ApoE*) ε4 allele status and age at disease onset influence the MTA scale [21]. These two factors should be considered when developing clinically useful cut-offs. To our knowledge, no previous studies have addressed this issue for the PA and GCA-F scales. The aims of this study were to (i) investigate the diagnostic and predictive value of the PA and GCA-F scales both in comparison with and in combination with the MTA scale; (ii) evaluate the influence of key demographic and clinical factors on the three scales; (iii) compare a broad range of cut-offs for diagnosing AD and predicting progression from MCI to AD; and (iv) generate a list of practical cut-offs in order to facilitate the clinical use of the three scales.

Materials and methods

Subjects

A total of 1147 participants were included in this study (322 patients with AD, 480 patients with MCI and 345 healthy control subjects). MCI patients were classified as 'converters' (MCI-C) or as 'stable' (MCI-S) if they fulfilled the diagnostic criteria of AD ($n = 95$) or remained stable ($n = 385$) after 12 months of follow-up, respectively. Data were obtained from the AddNeuroMed and ADNI

studies. AddNeuroMed is part of the Innovative Medicines in Europe (InnoMed) European Union Sixth Framework programme and was designed to develop and validate novel surrogate markers in AD [23, 24]. ADNI was 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. The project was established to develop standardized imaging techniques and biomarker procedures in healthy control subjects and patients with AD or MCI. Data were obtained from the ADNI database (adni.loni.usc.edu; Principal Investigator Michael W. Weiner).

Participant recruitment and eligibility criteria were similar in the two cohorts [25–27]. Briefly, AD diagnosis was based on the NINCDS-ADRDA and DSM-IV criteria for probable AD, as well as a total Clinical Dementia Rating (CDR) score [28] of ≥0.5. MCI diagnosis required a mini-mental state examination (MMSE) score [29] of between 24 and 30, memory complaints, normal activities of daily living, a total CDR score of 0.5 and a Geriatric Depression Scale (GDS) score [30] of ≤5. The inclusion criteria for control subjects were an MMSE score of between 24 and 30, a total CDR score of 0 and a GDS score ≤5. For all three groups, exclusion criteria included significant neurological or psychiatric illness, significant unstable systemic illness or organ failure and history of alcohol or substance abuse or dependence. All clinical diagnoses were made without the use of MRI scans.

MRI and visual rating scales

MRI data acquisition in the AddNeuroMed study was designed to be compatible with the ADNI protocol [31]. A high-resolution sagittal three-dimensional T1-weighted MPRAGE sequence was applied in both studies (voxel size $1.1 \times 1.1 \times 1.2$ mm^3). MTA, GCA and PA scales were applied as described elsewhere [4, 7, 10] based only on T1-weighted images. The MTA scale scores the degree of atrophy from 0 to 4 in the hippocampus, parahippocampal gyrus, entorhinal cortex and the surrounding cerebrospinal fluid spaces. The PA scale scores the degree of atrophy from 0 to 3 in the posterior cingulate sulcus, precuneus, parieto-occipital sulcus and the parietal cortex. The GCA scale was applied only to the frontal lobe in order to provide a measurement of frontal atrophy (i.e. the GCA-F subscale). The anatomical boundaries of the frontal lobe were defined by the central sulcus, the frontal bone and the fissure of Sylvius. Scores also range from 0 to 3. In the three visual rating scales, a score of 0 denotes no atrophy, whereas scores of 1 to 3 indicate an increasing degree of atrophy. MTA analysis was based on coronal reconstructions, GCA-F on axial reconstructions and PA on reconstructions from all three planes. All visual ratings were performed by a single rater (LC), who was blind to the diagnosis and to demographic and clinical information. Intrarater reliability examined in 100 randomly selected participants provided weighted kappa values of 0.93 and 0.94 for MTA (left and right hemispheres, respectively), 0.70 for GCA-F and 0.72 for PA.

Cut-offs for the visual rating scales

To accomplish the third objective, we evaluated a broad spectrum of cut-offs designed to determine deviation from normality in MTA, GCA-F and PA. All these cut-offs are discussed in the Supplementary material. The cut-off indicates the first value considered abnormal (e.g. a cut-off of MTA ≥2 means that a value ≥2 on the MTA scale should be considered abnormal). First, we included several previously described cut-offs: MTA ≥1.5 [20], MTA ≥2 [21], MTA age-75 [4] and PA ≥2 [7]. Next, we tested for the first time the following complementary cut-offs: GCA-F ≥ 1, GCA-F ≥ 2, GCA-F ≥ 3, GCA-F age-75, PA ≥1 and PA age-75. Finally, we created a new age-adjusted cut-off based on age range in decades (age-decades-adjusted cut-off): (i) the performance of the cut-offs was calculated by stratifying the sample by decades [45–54 + 55–64, 65–74, 75–84 and 85–94 years; decades 45–54 and 55–64 years were combined because of the small sample size below the age of 54 years ($n = 2$)]; and (ii) cut-offs that provided the best diagnostic performance for each decade were combined to create a single cut-off (Table S1).

Demographic and clinical variables

Age, gender and years of education were included in this study as demographic variables. Disease severity was assessed using the CDR scale, including only the memory domain of CDR in the analysis. Cognitive performance was assessed with the MMSE and ADAS-Cog [32], although ADAS-Cog was replaced by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery [33] for healthy control subjects and patients with

MCI in the AddNeuroMed study. The CERAD 10-word recall task was rescored as described previously [21] in order to be compatible with the 10-word recall task (ADAS1) in ADAS-Cog. Age at disease onset, disease duration and *ApoE* ε4 status were also included in the study. *ApoE* genotype was evaluated as described previously [31, 34].

Statistical analysis

Because previous studies have shown that patterns of AD atrophy [35] and diagnostic and predictive performances of the MTA scale [21] are similar in the AddNeuroMed and ADNI cohorts, we combined the two data sets to provide higher power for the statistical analyses. One-way independent ANOVA and the Mann–Whitney *U*-test were used for continuous variables, and the Kruskal–Wallis and chi-squared tests were used for ordinal and dichotomous variables, respectively. Mixed ANOVA and two-way independent ANCOVA were used to analyse the interaction between two independent variables. *P*-values in all *post hoc* analyses were adjusted using Bonferroni correction for multiple comparisons. Multiple linear regression (backwards) was performed to analyse the influence of demographic and clinical variables on the three visual rating scales. Diagnostic performance was determined by comparing patients with AD and control subjects; for predictive performance, patients with MCI-C and patients with MCI-S were compared. Both diagnostic and predictive performances were evaluated by: (i) sensitivity and specificity values calculated from true- and false-positive/negative values; (ii) receiver operating characteristic curve analyses with their respective areas under the curve (AUCs) and 95% confidence intervals; and (iii) Cohen's Kappa test using the criteria of Landis and Koch to interpret the magnitude of the agreement ($\kappa < 0$, no agreement; $\kappa = 0$–0.20, slight agreement; $\kappa = 0.21$–0.40, fair agreement; $\kappa < 0.41$–0.60, moderate agreement; $\kappa = 0.61$–0.80, substantial agreement; $\kappa = 0.81$–1.0, almost perfect agreement) [36]. $P \leq 0.05$ was considered statistically significant. All analyses were performed using IBM SPSS Statistics for Mac, Version 20.0. (IBM Corp., Released 2011, Armonk, NY, USA) for Mac.

Results

Demographic and clinical variables and regional atrophy

Table 1 shows the demographic and clinical characteristics of the four study groups (AD, MCI-S, MCI-C and control). MTA scores were converted to a scale of 0 to 3 to allow comparison with GCA-F

Table 1 *Demographic and clinical variables*

	CTRL (n = 345)	MCI-S (n = 385)	MCI-C (n = 95)	AD (n = 322)	P
AddNeuroMed/ADNI, n	115/230	101/284	25/70	123/199	0.004
Age, years	74.91 (5.80)	75.08 (6.89)	74.25 (6.54)	75.60 (7.14)	0.210
Gender, % female	51.0	39.5	41.1	54.7	<0.001
Years of education (n = 1144)	14.27 (4.39)	14.01 (4.56)	13.73 (4.23)	12.15 (4.79)[b,c,d]	<0.001
Disease onset, % early onset[a]	–	–	–	16.3	–
Disease duration, years[a]	–	–	–	3.68 (2.53)	–
ApoE ε4, % carriers (n = 1121)	27.9	47.3	63.4	62.1	<0.001
MMSE (n = 1144)	28.92 (2.46)	27.13 (1.71)[b]	26.55 (1.81)[b]	21.53 (5.40)[b,c,d]	<0.001
CDR memory (n = 1103)	0	0.5[b]	0.5[b]	0.92 (0.43)[b,c,d]	<0.001
ADAS1 (n = 1089)	3.05 (1.31)	4.61 (1.37)[b]	5.35 (1.34)[b,c]	6.33 (1.46)[b,c,d]	<0.001

CTRL, control; MCI-S, mild cognitive impairment remaining stable at 12 months of follow-up; MCI-C, mild cognitive impairment fulfilling the diagnostic criteria of AD at 12 months of follow-up; AD, Alzheimer's disease; *ApoE* ε4, apolipoprotein E ε4 allele; MMSE, mini-mental state examination; CDR memory, Clinical Dementia Rating – memory domain; ADAS1, Alzheimer's Disease Assessment Scale, 10-word recall task.
Values are presented as mean (SD) unless otherwise stated.
Post hoc analyses were adjusted by Bonferroni correction for multiple comparisons.
[a]Only for the AD group, n = 320.
[b]Significantly different from CTRL.
[c]Significantly different from MCI-S.
[d]Significantly different from MCI-C.

and PA (Table 2 and Fig. 1). A mixed ANOVA was performed to analyse the interaction between visual rating scale (within-subject factor with three levels: MTA, GCA-F and PA) and diagnostic group (between-subject factor with four levels: control, MCI-S, MCI-C and AD). The results showed a signif-

Table 2 *Interaction between regional atrophy and diagnostic groups*

	CTRL (n = 345)	MCI-S (n = 385)	MCI-C (n = 95)	AD (n = 322)	P	Main effect (visual rating scale)
MTA, mean (SD)	0.86 (0.50)[a,b]	1.17 (0.62)[a,b,c]	1.44 (0.61)[b,c,d]	1.74 (0.70) [b,c,d,e]	<0.001	1.30 (0.02)[a,b]
GCA-F, mean (SD)	0.39 (0.55)	0.52 (0.63)[c]	0.74 (0.72)[c,d]	0.80 (0.72)[c,d]	<0.001	0.61 (0.02)
PA, mean (SD)	0.68 (0.76)[a]	0.76 (0.73)[a]	0.88 (0.74)	0.88 (0.83)[c]	0.004	0.80 (0.03)[a]
P	<0.001	<0.001	<0.001	<0.001	–	<0.001
Main effect (diagnostic group), mean (SE)	0.64 (0.03)	0.82 (0.02)[c]	1.02 (0.05)[c,d]	1.14 (0.03)[c,d]	<0.001	–

MTA, medial temporal atrophy; GCA-F, frontal lobe atrophy (from the global cortical atrophy scale); PA, posterior cortical atrophy. MCI-S, mild cognitive impairment remaining stable at 12 months of follow-up; MCI-C, mild cognitive impairment fulfilling the diagnostic criteria of AD at 12 months of follow-up; AD, Alzheimer's disease.
MTA scores were converted to a scale of 0 to 3 to allow comparison with the GCA-F and PA scores. Conversion consisted of multiplying MTA scores by a factor of 0.75.
Post hoc analyses were adjusted by Bonferroni correction for multiple comparisons.
[a]Significantly different from GCA-F.
[b]Significantly different from PA.
[c]Significantly different from CTRL.
[d]Significantly different from MCI-S.
[e]Significantly different from MCI-C.

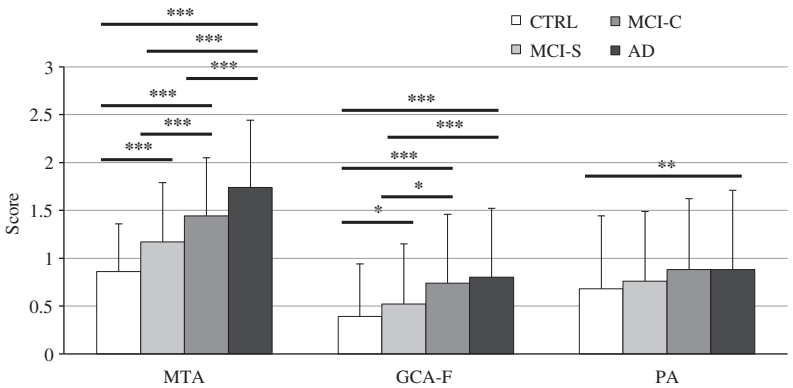

Fig. 1 *Regional atrophy across study groups.* Post hoc *analyses were adjusted by Bonferroni correction for multiple comparisons. MTA scores were converted to a scale of 0 to 3 to allow comparison with GCA-F and PA scores. Conversion consisted of multiplying MTA scores by a factor of 0.75. CTRL, control; MCI-S, mild cognitive impairment remaining stable at 12 months of follow-up; MCI-C, mild cognitive impairment fulfilling the diagnostic criteria of AD at 12 months of follow-up. AD, Alzheimer's disease; MTA, medial temporal atrophy; GCA-F, frontal lobe atrophy (from the global cortical atrophy scale); PA, posterior atrophy.* ***P < *0.001;* **P < *0.01;* *P < *0.05.*

icant interaction ($F_{(6,2042.653)} = 19.554$; $P < 0.001$), indicating a gradient of global atrophy with AD and MCI-C groups displaying the same degree of atrophy but greater in comparison with the MCI-S and control groups (AD = MCI-C > MCI-S > control) ($F_{(3,1143)} = 66.492$; $P < 0.001$). This global degree of atrophy was modulated by the factor visual rating scale. In particular, MTA perfectly separated the four groups: AD > MCI-C > MCI-S > control ($U_{(3)} = 275.801$; $P < 0.001$). GCA-F reflected the gradient of global atrophy: AD = MCI-C > MCI-S > control ($U_{(3)} = 66.792$; $P < 0.001$). PA only showed statistical differences between the AD and control groups: AD > control ($U_{(3)} = 13.480$; $P = 0.004$). By contrast, there was a gradient of higher MTA scores, followed by PA and finally GCA-F: MTA > PA > GCA-F ($\chi^2_{(2)} = 534.401$; $P > 0.001$). This pattern was modulated by the factor diagnostic group. The main effect (MTA > PA > GCA-F) was also found in the control and MCI-S groups (control: $\chi^2_{(2)} = 108.347$; $P > 0.001$; MCI-S: $\chi^2_{(2)} = 161.043$; $P > 0.001$). However, no significant differences were found between PA and GCA-F in the AD and MCI-C groups (MTA > PA = GCA-F) (AD: $\chi^2_{(2)} = 262.672$; $P > 0.001$; MCI-C: $\chi^2_{(2)} = 45.216$; $P > 0.001$).

Diagnostic and predictive performances of the visual rating scales

Diagnostic performance was evaluated by comparing patients with AD and control subjects ($n = 667$). The MTA scale showed the best diagnostic performance (AUC = 83.8), followed by the GCA-F and PA scales (AUC = 65.3 and 56.7, respectively) (Table 3 and Fig. S1A). Predictive performance was evaluated by comparing the MCI-C and MCI-S groups ($n = 480$). The MTA scale again showed the best performance (AUC = 62.4), followed by the GCA-F and PA scales (AUC = 58.1 and 54.7, respectively) (Table 3 and Fig. S1B).

Influence of demographic and clinical factors on the visual rating scales

Several multiple linear regression models were fitted to evaluate the influence of age, gender, years of education, age at disease onset, disease duration and *ApoE* ε4 status on the three visual rating scales. First, separate models were generated for the AD group ($n = 313$) because age at disease onset and disease duration were only available for this group. A greater degree of MTA was found to be related to older age, male gender and longer disease duration; a greater degree of GCA-F and PA was related to older age and earlier disease onset (Table 4). Next, the same models were applied to the whole sample ($n = 1119$) but excluding age at disease onset and disease duration as predictors. Results showed that older age, male gender, fewer years of education and *ApoE* ε4 carrier status significantly predicted increased MTA and GCA-F, whereas only older age significantly predicted increased PA (Table 4).

Table 3 *Diagnostic and predictive performances of the visual rating scales*

Visual rating scales	AUC	95% CI	SE	*P*
Diagnostic performance (AD vs. CTRL)				
MTA	83.8	80.8–86.9	0.016	<0.001
GCA-F	65.3	61.2–69.5	0.021	<0.001
PA	56.7	52.3–61.0	0.022	0.003
Predictive performance (MCI-C vs. MCI-S)				
MTA	62.4	56.2–68.6	0.032	<0.001
GCA-F	58.1	51.6–64.7	0.034	0.034
PA	54.7	48.3–61.2	0.033	0.033

AUC, area under the curve; CI, confidence interval; AD, Alzheimer's disease; CTRL, control; MCI-C, mild cognitive impairment fulfilling the diagnostic criteria of AD at 12 months of follow-up; MCI-S, mild cognitive impairment remaining stable at 12 months of follow-up; MTA, medial temporal atrophy; GCA-F, frontal lobe atrophy (from the global cortical atrophy scale); PA, posterior atrophy.

Diagnostic and predictive performances of different cut-offs

We evaluated different cut-offs to determine which provided the best diagnostic performance for separating patients with AD and control subjects ($n = 667$; Table 5). All these cut-offs are discussed in the Supplementary material. Because regression analyses showed that age was the main predictor for the three visual rating scales, we first calculated the age-decades-adjusted cut-off as described in the Methods and shown in Table S1. This age-decades-adjusted cut-off was compared with a broad spectrum of cut-offs (Table 5).

The age-decades-adjusted cut-off resulted in the best performance for MTA (AUC = 78.5). With regard to GCA-F and PA, both the ≥1 and the age-decades cut-offs had the best diagnostic performance (GCA-F ≥ 1: AUC = 63.4; GCA-F age-decades: AUC = 63.3; PA ≥1: AUC = 55.2; PA age-decades: AUC = 55.0). However, the Cohen's kappa coefficient showed only slight agreement

Table 4 *Multiple regression models: influence of demographic and clinical factors on the visual rating scales*

Model	Criterion (Y)	R^2	F	P	Predictors (X)	ß	P
Model 1 (AD group, n = 313)	MTA	0.135	16.080	<0.001	Age	0.302	<0.001
					Disease duration	0.134	0.011
					Gender	−0.127	0.018
	GCA-F	0.117	20.627	<0.001	Age	0.438	<0.001
					Age at disease onset	−0.195	0.006
	PA	0.056	9.149	<0.001	Age	0.310	<0.001
					Age at disease onset	−0.168	0.023
Model 2 (whole sample, n = 1119)	MTA	0.154	50.544	<0.001	Age	0.310	<0.001
					Gender	−0.111	<0.001
					ApoE ε4	0.222	<0.001
					Years of education	−0.093	0.001
	GCA-F	0.120	37.901	<0.001	Age	0.280	<0.001
					Gender	−0.126	<0.001
					Years of education	−0.156	<0.001
					ApoE ε4	0.090	0.002
	PA	0.048	56.703	<0.001	Age	0.220	<0.001

AD, Alzheimer's disease; MTA, medial temporal atrophy; GCA-F, frontal lobe atrophy (from the global cortical atrophy scale); PA, posterior atrophy; *ApoE* ε4, apolipoprotein E ε4 allele.
Multiple linear regression (backwards). Model 1: age + gender (male 0, female 1) + years of education + disease onset (early onset 0, late onset 1) + disease duration + *ApoE* ε4 status (noncarriers 0, carriers 1); model 2: age + gender (male 0, female 1) + years of education + *ApoE* ε4 status (noncarriers 0, carriers 1).

for PA (see Table 5 and Fig. S1C–E). Using the cut-offs with the best performance, combining MTA with GCA-F or PA, or both, increased sensitivity up to 93.8% when at least one of the scales was required to be abnormal, and specificity to 91.0% when the three scales were required to be abnormal (Table S2 and Fig. S1F). However, this improvement occurred at the expense of lower specificity and sensitivity values, respectively.

We also evaluated the predictive performance of the different cut-offs for separating patients with MCI-C from those with MCI-S (n = 480; Table 5). The age-75-adjusted cut-off showed the best predictive performance for MTA (AUC = 62.7), and the age-decades-adjusted cut-off had the best predictive performance for both GCA-F and PA (AUC = 57.9 and 55.0, respectively). Nonetheless, the Cohen's kappa coefficient showed slight agreement for the three scales (see Table 5 and Fig. S1G–I). Combining MTA with GCA-F or PA, or both, increased sensitivity up to 89.5% when at least one of the scales was required to be abnormal, and specificity to 89.9% with improved Cohen's kappa (Table S2 and Fig. S1J). However, this improvement occurred at the expense of lower specificity and sensitivity values, respectively.

Influence of age at disease onset and ApoE ε4 status on the different cut-offs

Based on the results obtained in the regression analyses presented above, we decided to evaluate the influence of age at disease onset and *ApoE* ε4 status on the three visual rating scales. Several two-way independent ANCOVA tests were performed for each visual rating scale in the AD group (n = 313). Disease onset (<65 years vs. ≥65 years) and *ApoE* ε4 status (carriers vs. noncarriers) were entered as independent variables. Age was entered as a covariate (Table S3). For MTA, there was a significant interaction between age at disease onset and *ApoE* ε4 status ($F_{(1,308)}$ = 5.113; P = 0.024), showing that MTA scores were higher in *ApoE* ε4 carriers, but only amongst those with early-onset (<65 years) disease (Fig. S2). For GCA-F and PA, only age at disease onset was a significant variable (GCA-F: $F_{(1,308)}$ = 5.470; P = 0.020; PA: $F_{(1,308)}$ = 5.799; P = 0.017). GCA-F and PA were increased in patients with AD with early disease onset, regardless of *ApoE* ε4 status.

Table 5 *Diagnostic and predictive performances of different cut-offs*

Cut-offs	SN	SP	AUC	95% CI	*P*	κ	*P*
Diagnostic performance (AD vs. CTRL)							
MTA ≥1.5	84.5	68.1	76.3	72.6–80.0	<0.001	52.3	<0.001
MTA ≥2	73.6	80.6	77.1	73.4–80.8	<0.001	54.3	<0.001
MTA age-75	69.9	83.2	76.5	72.8–80.3	<0.001	53.3	<0.001
MTA age-decades	80.1	76.8	78.5	74.9–82.1	<0.001	56.8	<0.001
GCA-F ≥ 1	61.8	64.9	63.4	59.1–67.6	<0.001	26.7	<0.001
GCA-F ≥ 2	17.4	96.5	57.0	52.6–61.3	0.002	14.3	<0.001
GCA-F ≥ 3	0.3	100	50.2	45.8–54.5	0.945	0.3	0.300
GCA-F age-75	33.2	85.5	59.4	55.0–63.7	<0.001	19.1	<0.001
GCA-F age-decades	58.4	67.8	63.1	58.9–67.3	<0.001	26.3	<0.001
PA ≥1	61.8	48.7	55.2	50.9–59.6	0.019	10.1	0.006
PA ≥2	23.3	85.2	54.3	49.9–58.6	0.057	8.7	0.005
PA ≥3	2.8	98.6	50.7	46.3–55.1	0.764	1.4	0.226
PA age-75	39.4	67.0	53.2	48.8–57.6	0.153	6.5	0.086
PA age-decades	59.6	50.4	55.0	50.7–59.4	0.025	10.0	0.009
Predictive performance (MCI-C vs. MCI-S)							
MTA ≥1.5	75.8	43.6	59.7	53.6–65.8	0.003	11.0	0.001
MTA ≥2	54.7	58.4	56.6	50.1–63.0	0.047	9.0	0.021
MTA age-75	60.0	65.5	62.7	56.4–69.1	<0.001	18.5	<0.001
MTA age-decades	66.3	51.7	59.0	52.7–65.3	0.007	11.2	0.002
GCA-F ≥ 1	57.9	55.6	56.7	50.3–63.2	0.042	8.9	0.018
GCA-F ≥ 2	15.8	92.7	54.3	47.5–61.0	0.198	10.7	0.009
GCA-F ≥ 3	0	100	50.0	43.5–56.5	0.999	0	0.999
GCA-F age-75	33.7	80.8	57.2	50.5–63.9	0.029	13.9	0.002
GCA-F age-decades	55.8	60.0	57.9	51.5–64.3	0.017	10.9	0.005
PA ≥1	66.3	41.6	53.9	47.5–60.3	0.234	4.5	0.161
PA ≥2	22.1	82.9	52.5	45.9–59.1	0.454	5.1	0.261
PA ≥3	0	100	50.0	43.5–56.5	0.999	0	0.999
PA age-75	44.2	63.1	53.7	47.1–60.2	0.268	5.4	0.188
PA age-decades	66.3	43.6	55.0	48.6–61.3	0.133	5.7	0.078

SN, sensitivity; SP, specificity; AUC, area under the curve; CI, confidence interval; κ, Cohen's kappa; AD, Alzheimer's disease; CTRL, control; MCI-C, mild cognitive impairment fulfilling the diagnostic criteria of AD at 12 months of follow-up; MCI-S, mild cognitive impairment remaining stable at 12 months of follow-up; MTA, medial temporal atrophy; GCA-F, frontal lobe atrophy (from the global cortical atrophy scale); PA, posterior atrophy.

MTA ≥1.5, as described by Schoonenboom *et al.* [20], is based on the average score of both hemispheres and requires a minimum MTA score of 1 in one hemisphere and 2 in the other; MTA ≥2, as described by Pereira *et al.* [21], requires an average MTA score of 2; MTA age-75, as described by Scheltens *et al.* [4], requires an MTA score of 2 in either of the two hemispheres in individuals younger than 75 years, and an MTA score of 3 in individuals aged 75 years and above; GCA-F and PA age-75 require GCA-F and PA scores, respectively, of 1 in individuals younger than 75 years, and of 2 in individuals aged 75 years and above; GCA-F and PA ≥1, ≥2 and ≥3 simply require total GCA-F and PA scores of 1, 2 and 3, respectively; the different age-decades cut-offs consist of four values adjusted for the following corresponding age ranges: 45–64, 65–74, 75–84 and 85–94 years. In particular, MTA age-decades requires MTA scores of ≥1.5, ≥1.5, ≥2 and ≥2.5; GCA-F age-decades requires GCA-F scores of ≥1, ≥1, ≥1 and ≥2; PA age-decades requires PA scores of ≥1, ≥1, ≥1 and ≥2 (see also Table S1).

Therefore, we decided to evaluate the influence of age at disease onset and *ApoE* ε4 status on the cut-offs by stratifying the patients with AD into four groups according to these factors ($n = 313$). Table 6 shows the percentage of AD patients with abnormal scores (i.e. sensitivity) according to the cut-offs that previously showed better diagnostic performance (see Table 5). Noncarrier AD patients with early-onset disease had significantly less MTA compared with the other three groups, according to both age-decades-adjusted and MTA ≥2 cut-offs. Moreover, sensitivity was suboptimal in this group of early-onset non-carrier AD patients (47.1%). These patients with AD also showed less GCA-F compared with late-onset noncarrier AD patients, according to both the ≥1 and age-decades-adjusted cut-offs. Thus, we evaluated all the cut-offs for this subgroup of early-onset noncarrier AD patients. As shown in Table S4 and Fig. S1K–L, the diagnostic performance was reduced (based on AUC values) compared to that for the whole AD group (Table 5 and Fig. S1C–D). The age-75-adjusted cut-off showed the best performance for both MTA and GCA-F with specificity above 80%, but sensitivity below 50% and slight agreement according to Cohen's kappa. Finally, PA was increased in early-onset *ApoE* ε4 carrier AD patients compared with late-onset carriers, according to the age-75-adjusted cut-off (Table 6).

Table 6 *Influence of age at disease onset and* ApoE *ε4 status on the different cut-offs (AD group,* n = *313)*

	ApoE ε4 carriers	*ApoE* ε4 noncarriers	*P*
MTA age-decades			
Early-onset AD	81.8	47.1	0.011
Late-onset AD	83.3	80.2	0.519
P	0.832	0.003	
MTA ≥2			
Early-onset AD	66.7	35.3	0.034
Late-onset AD	75.9	78.2	0.668
P	0.267	<0.001	
GCA-F ≥ 1			
Early-onset AD	51.5	35.3	0.276
Late-onset AD	60.5	71.3	0.075
P	0.339	0.004	
GCA-F age-decades			
Early-onset AD	51.5	35.3	0.276
Late-onset AD	56.8	66.3	0.124
P	0.578	0.015	
PA ≥1			
Early-onset AD	57.6	47.1	0.480
Late-onset AD	64.2	59.4	0.435
P	0.472	0.341	
PA age-decades			
Early-onset AD	57.6	47.1	0.480
Late-onset AD	61.1	57.4	0.553
P	0.705	0.426	
PA age-75			
Early-onset AD	57.6	47.1	0.480
Late-onset AD	37.7	33.7	0.512
P	0.034	0.286	

Values represent the percentage of patients with AD with abnormal scores (i.e. sensitivity) according to the cut-offs that previously showed better diagnostic performance (see Table 4). The cut-offs are explained in Table 5.
ApoE ε4, apolipoprotein E ε4 allele; MTA, medial temporal atrophy; GCA-F, frontal lobe atrophy (from the global cortical atrophy scale); PA, posterior atrophy; early onset, age <65 years; late onset, age ≥65 years.

Discussion

The final aim of this study was to generate a list of practical cut-offs to facilitate the clinical use of three visual rating scales, that is MTA, GCA-F and PA. The results show that MTA, GCA-F and PA are useful tools for assessing regional atrophy in AD and MCI. However, the performance of these scales differs because of the gradient of atrophy across the AD continuum (AD > MCI-C > MCI-S > control), as well as across the rating scales (MTA > GCA-F and PA). In addition, age, *ApoE* ε4 status and age at disease onset significantly influence the performance of the scales.

To our knowledge, the MTA, GCA and PA scales have not been directly compared in terms of diagnostic performance in previous studies. The fact that MTA showed better performance than GCA-F and PA was expected given that regional atrophy in the medial temporal lobe is a main finding in AD and is typically greater than atrophy in the frontal lobe or posterior cortex [37–40]. This gradient of more MTA than GCA-F and PA was confirmed in the present study. Similar diagnostic performance for MTA (AUC = 89%) and PA (AUC = 60%) has previously been reported [41]. The fact that the three rating scales had better performance for AD diagnosis than for MCI prognosis is consistent with previous studies [21, 35] and was also expected because brain atrophy is

greater in AD than in MCI [37–39, 42, 43], as confirmed in this study.

Regional atrophy in the medial temporal lobe may coexist with atrophy in the frontal lobe or posterior cortex, either as a part of the same pathophysiological process (e.g. advanced AD with global atrophy) or as parallel diseases (e.g. AD and vascular disease). In this sense, the combination of two or three of the rating scales may improve diagnostic and predictive performance [7]. Combining MTA with GCA-F or PA, or both, did not increase AUC values for patients with AD compared with healthy control subjects. However, abnormal values in two or three of the scales helped to successfully rule out healthy control subjects (specificity around 90%). Likewise, an abnormal value for at least one of the scales increased the certainty of AD (sensitivity around 90%), although did not allow normality to be excluded (specificity below 55%). With regard to MCI prediction, combining MTA with GCA-F resulted in slightly increased AUC values. Of note, normal values in the three scales helped to rule out conversion from MCI to AD, at least during the next 12 months (specificity around 90%). Abnormal values in at least one of the scales increased the certainty of MCI due to AD (sensitivity around 90%). Previous studies have shown that the MTA scale is useful for distinguishing patients with AD from healthy control subjects and from patients with other dementias [4, 41, 44–46], as well as for predicting progression from MCI to AD [41, 44, 47–49]. The findings of the few studies that have examined PA suggest that this scale may be useful for distinguishing patients with AD from those with other dementias, particularly amongst younger patients (<65 years) in whom medial temporal atrophy seems to be less evident [7, 50]. Recently, PA was found to predict progression from MCI to AD in patients with late-onset MCI [48]. Our findings suggest that previously reported results could be improved by combining MTA with the PA and/or GCA-F scales. This highlights the importance of obtaining suitable cut-offs for the MTA, GCA-F and PA scales.

Several factors influenced the scores on the three visual rating scales as well as the derived cut-offs. Age was the main influencing factor. Consequently, the proposed age-decades-adjusted cut-off provided the best performance in several settings. The fact that ageing is associated with both global and regional brain atrophy compromises the clinical utility of the visual rating scales in patients of advanced age. Two recent studies comparing patients with AD and healthy control subjects showed that the diagnostic performance of the MTA scale is worse in individuals older than 75 years compared with individuals below this age [21, 44]. An age-corrected cut-off based on a threshold of 75 years showed either suboptimal sensitivity (68%) [44] or suboptimal specificity (63%) [21] for the older group. Therefore, a more specific age correction is needed. An age correction based on decades allowed us to optimize the performance of the MTA scale over smaller age intervals, which led to better results. Sensitivity was 80% and specificity was 77% when comparing patients with AD and healthy control subjects. Of note, this new age-decades-adjusted cut-off provided better diagnostic performance than other cut-offs in previous studies [4, 20, 21, 50]. Age also affected the GCA-F and PA scores. The performance of the age-decades-adjusted cut-off was better, compared with most of the other cut-offs, but ≥1 for GCA-F and PA showed slightly better performance. However, sensitivity and specificity values were low (<65%). Low sensitivity for the PA scale has also been reported previously [7, 50].

The two other influencing factors were age at disease onset and *ApoE* ε4 status. Previous studies have shown that whilst late-onset AD is frequently related to more severe atrophy in medial temporal lobe structures, early-onset AD seems to favour atrophy in the posterior and frontal cortex [50–52]. This is partly supported by our results. Patients with early-onset AD showed more PA and GCA-F. In addition, patients with late-onset AD demonstrated more MTA, but this result was no longer significant after adjusting for age. Moreover, it should be noted that these results were modulated by *ApoE* ε4 status, as discussed below. The *ApoE* ε4 allele, the principal genetic risk factor for sporadic AD, has been associated with atrophy in the medial temporal lobe [52, 53]. There is increasing evidence to indicate that *ApoE* ε4 may modulate disease phenotype in AD, with *ApoE* ε4 carriers exhibiting medial temporal and occipital foci of atrophy, and noncarriers showing more pronounced frontoparietal volume loss [54]. In line with this, *ApoE* ε4 carriers had more atrophy in the medial temporal lobe than noncarriers. Of interest, frontal atrophy was more pronounced in noncarriers, but only in those with late-onset AD. In addition, *ApoE* ε4 carriers had more medial temporal and posterior atrophy, but only in those with early-onset AD.

Greater posterior atrophy was observed only when using the PA age-75 cut-off. The fact that the PA ≥1 and PA age-decades cut-offs did not show any significant difference suggests that both can better control the influence of age at disease onset and *ApoE* ε4 on posterior atrophy and therefore should be used in preference to other cut-offs. Therefore, age at disease onset and *ApoE* ε4 status not only influence regional atrophy independently but also interact with each other.

There is enough evidence to suggest that age, *ApoE* ε4 status and age at disease onset should be taken into account when using the MTA, GCA-F and PA scales in clinical practice. For example, for a hypothetical 65-year-old patient with suspicion of AD who is an *ApoE* ε4 noncarrier and whose symptoms started at the age of 60, an MTA cut-off of ≥1.5 would be suggested initially; however, our results show that an MTA cut-off of ≥2 is more appropriate for this patient (see Table 7). Amongst the previously suggested MTA cut-offs, only the one proposed by Scheltens and colleagues takes into account age [4], but this cut-off does not include *ApoE* ε4 status and age at disease onset. We acknowledge, however, that *ApoE* ε4 genotyping is not widely established in clinical practice and involves ethical issues. Therefore, at present, this specification for early-onset *ApoE* ε4 non-carrier AD patients may be of more benefit for research and clinical trials.

Table 7 *Practical cut-offs for visual rating scales of medial temporal, frontal and posterior atrophy. The cut-off is the initial value considered abnormal (e.g. a cut-off of ≥2 means that a value greater than or equal to 2 should be considered abnormal). 'Heterogeneous group' refers to the group of AD patients typically attending a memory clinic, including patients with early- or late-onset disease, and both* ApoE *ε4 carriers and noncarriers. 'Early-onset* ApoE *ε4 noncarriers' refers to AD patients with a disease onset before 65 years of age and who do not carry the* ApoE *ε4 allele*

	MTA	GCA-F	PA
Heterogeneous group			
45–64 years	≥1.5	≥1	≥1
65–74 years	≥1.5	≥1	≥1
75–84 years	≥2	≥1	≥1
85–94 years	≥2.5	≥1	≥1
Early-onset *ApoE ε4* non-carriers			
45–64 years	≥2	≥1	–
65–74 years	≥2	≥1	–
75–84 years	≥3	≥2	–
85–94 years	≥3	≥2	–

MTA, medial temporal atrophy; GCA-F, frontal lobe atrophy (from the global cortical atrophy scale); PA, posterior atrophy; AD, Alzheimer's disease; *ApoE* ε4, apolipoprotein E ε4 allele.

Some limitations should be recognized. The diagnosis of AD and MCI in the AddNeuroMed and ADNI cohorts lacks pathological confirmation. In spite of the strict selection criteria applied in both studies, we cannot assume the total absence of other underlying abnormalities or mixed neurodegenerative processes. For this reason, it is important to note that performance greater than 80% is not easily achieved as the performance for clinical diagnosis of AD is already approximately 85% when validated by pathological confirmation [55, 56]. It is also worth mentioning that intrarater reliability was lower for GCA-F and PA than for MTA. Similar findings have been reported previously [7, 11] and may be explained by the fact that MTA includes a relatively small area, whereas GCA-F and PA require the inspection of much larger areas and an 'averaged' judgement to be taken. Moreover, raters usually have more experience with MTA than with GCA-F and PA. Nonetheless, according to the criteria of Landis and Koch [36], intrarater values obtained in this study indicate substantial agreement for GCA-F and PA and excellent agreement for MTA. The predictive performance for MCI was based on a follow-up period of 12 months. It is thus necessary to evaluate these scales over longer follow-up periods, which normally provide better performance results [57]. GCA-F has been used to assess frontal atrophy in two previous studies [11, 14], but the scale still needs to be quantitatively validated using volumetric MRI. Finally, the proposed list of practical cut-offs needs to be evaluated in clinical settings and compared with other dementias.

In conclusion, MTA, GCA-F and PA are useful scales for assessing regional brain atrophy and aiding AD diagnosis, and potentially for determining MCI prognosis. Age, *ApoE* ε4 status and age at disease onset significantly influenced the scores in the three visual rating scales as well as the derived cut-offs. By taking these factors into account, we have proposed a list of practical cut-offs for the MTA, GCA-F and PA scales (see Table 7). The provision of reliable and practical cut-offs for assessing atrophy in the medial temporal lobe,

frontal lobe and posterior cortex is of utmost importance especially now that different presentations of AD have been recognized in the revised diagnostic criteria for AD [1]. MTA is incorporated in the algorithm to assess hippocampal atrophy on MRI [1, 3] and has a well-established association with the amnestic presentation [4, 40, 44, 58–60]. GCA-F and PA may be of value to support nonamnestic presentations such as those evidencing executive impairment and language and visuospatial impairment, respectively. In addition, the GCA-F and PA scales could assist the differential diagnosis between AD and non-AD dementias [7]. If successfully validated in clinical settings, the list of practical cut-offs proposed here may eventually favour the adoption of the three visual rating scales in clinical practice. Some steps have already been taken towards this goal by suggesting that the use of the MTA scale appears to be justified in clinical routine [44, 61, 62]. Another favourable outcome of validating this list of practical cut-offs would be to foster studies on disease subtypes based on patterns of structural atrophy and related cognitive impairment. It is hoped that this will increase the understanding of different presentations of AD, improve diagnostic and predictive methods, aid population selection and enrichment for clinical trials and, as a central goal, improve clinical care.

Conflict of interest statement

No conflicts of interests to declare.

Acknowledgements

AddNeuromed is supported by InnoMed, an Integrated Project funded by the European Union Sixth Framework programme priority FP6-2004-LIFESCIHEALTH-5, Life Sciences, Genomics and Biotechnology for Health. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering and generous contributions from the following: the Alzheimer's Association, the Alzheimer's Drug Discovery Foundation, BioClinica Inc., Biogen Idec Inc., Bristol-Myers Squibb Co., Eisai Inc., Elan Pharmaceuticals Inc., Eli Lilly and Co., F. Hoffmann-La Roche Ltd and its affiliated company Genentech Inc., GE Healthcare, Innogenetics NV, 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 Corp., 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. Data collection and sharing in ADNI were funded by the National Institutes of Health (NIH) (grant U01 AG024904). ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

The authors thank Swedish Brain Power, the Strategic Research Programme in Neuroscience at Karolinska Institutet (StratNeuro), the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Health Research Council of the Academy of Finland, University of Eastern Finland UEFBRAIN and EVO funding from Kuopio University Hospital. AS and SL were supported by funds from the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health and the NIHR Biomedical Research Unit for Dementia at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Kings College London, UK.

References

1 McKhann GM, Knopman DS, Chertkow H *et al.* The diagnosis of dementia 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* 2011; **7:** 263–9.

2 Albert MS, DeKosky ST, Dickson D *et al.* 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* 2011; **7:** 270–9.

3 Dubois B, Feldman HH, Jacova C *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* 2007; **6:** 734–46.

4 Scheltens P, Leys D, Barkhof F *et al.* Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* 1992; **55:** 967–72.

5 Scheltens P, van de Pol L. Impact commentaries. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* 2012; **83:** 1038–40.

6 Möller C, van der Flier WM, Versteeg A *et al.* Quantitative regional validation of the visual rating scale for posterior cortical atrophy. *Eur Radiol* 2014; **24:** 397–404.

7 Koedam ELGE, Lehmann M, van der Flier WM *et al.* Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol* 2011; **21:** 2618–25.

8 Kipps CM, Davies RR, Mitchell J, Kril JJ, Halliday GM, Hodges JR. Clinical significance of lobar atrophy in frontotemporal dementia: application of an MRI visual rating scale. *Dement Geriatr Cogn Disord* 2007; **23:** 334–42.

9 Chen Y, Chen X, Xiao W, Mok VCT, Wong KS, Tang WK. Frontal lobe atrophy is associated with small vessel disease in ischemic stroke patients. *Clin Neurol Neurosurg* 2009; **111:** 852–7.

10 Pasquier F, Leys D, Weerts J, Mounier-Vehier F, Barkhof F, Scheltens P. Inter-and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. *Eur Neurol* 1996; **36:** 268–72.

11 Scheltens P, Pasquier F, Weerts J, Barkhof F, Leys D. Qualitative assessment of cerebral atrophy on MRI: inter- and intra-observer reproducibility in dementia and normal aging. *Eur Neurol* 1997; **37:** 95–9.

12 Henneman WJP, Vrenken H, Barnes J *et al.* Baseline CSF p-tau levels independently predict progression of hippocampal atrophy in Alzheimer disease. *Neurology* 2009; **73:** 935–40.

13 Van der Vlies AE, Goos JDC, Barkhof F, Scheltens P, van der Flier WM. Microbleeds do not affect rate of cognitive decline in Alzheimer disease. *Neurology* 2012; **79:** 763–9.

14 Doody RS, Azher SN, Haykal HA, Dunn JK, Liao T, Schneider L. Does APO epsilon4 correlate with MRI changes in Alzheimer's disease? *J Neurol Neurosurg Psychiatry* 2000; **69:** 668–71.

15 Biessels GJ, De Leeuw F-E, Lindeboom J, Barkhof F, Scheltens P. Increased cortical atrophy in patients with Alzheimer's disease and type 2 diabetes mellitus. *J Neurol Neurosurg Psychiatry* 2006; **77:** 304–7.

16 Goos JDC, Kester MI, Barkhof F *et al.* Patients with Alzheimer disease with multiple microbleeds: relation with cerebrospinal fluid biomarkers and cognition. *Stroke* 2009; **40:** 3455–60.

17 Kim Y-S, Lee K-M, Choi BH, Sohn E-H, Lee AY. Relation between the clock drawing test (CDT) and structural changes of brain in dementia. *Arch Gerontol Geriatr* 2009; **48:** 218–21.

18 Van de Pol LA, Scheltens P. Medial temporal lobe atrophy scores translated to clinical practice: editorial comment on "influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cut-offs". *J Intern Med* 2014; **275:** 331–3.

19 Barkhof F, Polvikoski TM, van Straaten ECW *et al.* The significance of medial temporal lobe atrophy: a postmortem MRI study in the very old. *Neurology* 2007; **69:** 1521–7.

20 Schoonenboom NSM, van der Flier WM, Blankenstein MA *et al.* CSF and MRI markers independently contribute to the diagnosis of Alzheimer's disease. *Neurobiol Aging* 2008; **29:** 669–75.

21 Pereira JB, Cavallin L, Spulber G *et al.* Influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cut-offs. *J Intern Med* 2014; **275:** 317–30.

22 Cavallin L, Bronge L, Zhang Y *et al.* Comparison between visual assessment of MTA and hippocampal volumes in an elderly, non-demented population. *Acta Radiol* 2012; **53:** 573–9.

23 Lovestone S, Francis P, Strandgaard K. Biomarkers for disease modification trials–the innovative medicines initiative and AddNeuroMed. *J Nutr Health Aging* 2007; **11:** 359–61.

24 Lovestone S, Francis P, Kloszewska I *et al.* AddNeuroMed–the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. *Ann N Y Acad Sci* 2009; **1180:** 36–46.

25 Spulber G, Simmons A, Muehlboeck J-S *et al.* An MRI-based index to measure the severity of Alzheimer's disease-like structural pattern in subjects with mild cognitive impairment. *J Intern Med* 2013; **273:** 396–409.

26 Simmons A, Westman E, Muehlboeck S *et al.* The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: experience from the first 24 months. *Int J Geriatr Psychiatry* 2011; **26:** 75–82.

27 Petersen RC, Aisen PS, Beckett LA *et al.* Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology* 2010; **74:** 201–9.

28 Morris JC. The clinical dementia rating (CDR): current vision and scoring rules Neurology. *Neurology* 1993; **43:** 2412–4.

29 Folstein MF, Folstein SE, McHugh PR. Mini-mental- state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; **12:** 189–98.

30 Sheikh J, Yesavage J. Geriatric depression scale (GDS): recent evidence and development of a shorter version. *Clin Gerontol* 1986; **5:** 165–73.

31 Jack CR, Bernstein MA, Fox NC *et al.* The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging* 2008; **27:** 685–91.

32 Mohs R, Knopman D, Petersen R *et al.* Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's disease assessment scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* 1997; **11:** S13–21.

33 Morris J, Heyman A, Mohs R *et al.* The Consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989; **39:** 1159–65.

34 Hixson J, Vernier D. Restriction isotyping of human apolipoprotein E by gene amplification and cleavage with HhaI. *J Lipid Res* 1990; **31:** 545–8.

35 Westman E, Simmons A, Muehlboeck J-S *et al.* AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. *NeuroImage* 2011; **58:** 818–28.

36 Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33:** 159–74.

37 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 U S A* 2002; **99:** 4703–7.

38 Fox NC, Crum WR, Scahill RI, Stevens JM, Janssen JC, Rossor MN. Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *Lancet* 2001; **358:** 8–11.

39 Whitwell J, Przybelski S, Weigand S *et al.* 3D maps from multiple MRI illustrate changing atrophy patterns as subjects progress from MCI to AD. *Brain* 2007; **130:** 1777–86.

40 Scheltens P, Fox N, Barkhof F, De Carli C. Structural magnetic resonance imaging in the practical assessment of dementia: beyond exclusion. *Lancet Neurol* 2002; **1:** 13–21.
41 O'Donovan J, Watson R, Colloby SJ *et al.* Does posterior cortical atrophy on MRI discriminate between Alzheimer's disease, dementia with Lewy bodies, and normal aging? *Int Psychogeriatr* 2013; **25:** 111–9.
42 Spulber G, Niskanen E, Macdonald S *et al.* Evolution of global and local grey matter atrophy on serial MRI scans during the progression from MCI to AD. *Curr Alzheimer Res* 2012; **9:** 516–24.
43 De Leon M, George A, Golomb J *et al.* Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. *Neurobiol Aging* 1997; **18:** 1–11.
44 Duara R, Loewenstein DA, Shen Q *et al.* The utility of age-specific cut-offs for visual rating of medial temporal atrophy in classifying Alzheimer's disease, MCI and cognitively normal elderly subjects. *Front Aging Neurosci* 2013; **5:** 47.
45 Wahlund L, Julin P, Lindqvist J, Scheltens P. Visual assessment of medical temporal lobe atrophy in demented and healthy control subjects: correlation with volumetry. *Psychiatry Res* 1999; **90:** 193–9.
46 Barber R, Golkar A, Scheltens P, Ballard C, McKeith I, O'Brien J. Medial temporal lobe atrophy on MRI in dementia with Lewy bodies. *Neurology* 1999; **52:** 1153–8.
47 DeCarli C, Frisoni GB, Clark CM *et al.* Qualitative estimates of medial temporal atrophy as a predictor of progression from mild cognitive impairment to dementia. *Arch Neurol* 2007; **64:** 108–15.
48 Lehmann M, Koedam EL, Barnes J *et al.* Visual ratings of atrophy in MCI: prediction of conversion and relationship with CSF biomarkers. *Neurobiol Aging* 2013; **34:** 73–82.
49 Korf ESC, Wahlund L-O, Visser PJ, Scheltens P. Medial temporal lobe atrophy on MRI predicts dementia in patients with mild cognitive impairment. *Neurology* 2004; **63:** 94–100.
50 Lehmann M, Koedam ELGE, Barnes J *et al.* Posterior cerebral atrophy in the absence of medial temporal lobe atrophy in pathologically-confirmed Alzheimer's disease. *Neurobiol Aging* 2012; **33:** 627.
51 Möller C, Vrenken H, Jiskoot L *et al.* Different patterns of gray matter atrophy in early- and late-onset Alzheimer's disease. *Neurobiol Aging* 2013; **34:** 2014–22.
52 Van der Flier WM, Pijnenburg YA, Fox NC, Scheltens P. Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ε4 allele. *Lancet Neurol* 2011; **10:** 280–8.
53 Cherbuin N, Leach LS, Christensen H, Anstey KJ. Neuroimaging and APOE genotype: a systematic qualitative review. *Dement Geriatr Cogn Disord* 2007; **24:** 348–62.
54 Pievani M, Rasser PE, Galluzzi S *et al.* Mapping the effect of APOE epsilon4 on gray matter loss in Alzheimer's disease in vivo. *NeuroImage* 2009; **45:** 1090–8.
55 Mendez M, Mastri A, Sung J, Frey W. Clinically diagnosed Alzheimer disease: neuropathologic findings in 650 cases. *Alzheimer Dis Assoc Disord* 1992; **6:** 35–43.
56 Victoroff J, Mack W, Lyness S, Chui H. Multicenter clinicopathological correlation in dementia. *Am J Psychiatry* 1995; **152:** 1476–84.
57 Ferreira D, Perestelo-Pérez L, Westman E, Wahlund L-O, Sarría A, Serrano-Aguilar P. Meta-review of CSF core biomarkers in Alzheimer's disease: the state-of-the-art after the new revised diagnostic criteria. *Front Aging Neurosci* 2014; **6:** 1–24.
58 Shim YS, Youn YC, Na DL *et al.* Effects of medial temporal atrophy and white matter hyperintensities on the cognitive functions in patients with Alzheimer's disease. *Eur Neurol* 2011; **66:** 75–82.
59 Van de Pol LA, Korf ESC, van der Flier WM *et al.* Magnetic resonance imaging predictors of cognition in mild cognitive impairment. *Arch Neurol* 2007; **64:** 1023–8.
60 Deweer B, Lehericy S, Pillon B *et al.* Memory disorders in probable Alzheimer's disease: the role of hippocampal atrophy as shown with MRI. *J Neurol Neurosurg Psychiatry* 1995; **58:** 590–7.
61 Boutet C, Chupin M, Colliot O *et al.* Is radiological evaluation as good as computer-based volumetry to assess hippocampal atrophy in Alzheimer's disease? *Neuroradiology* 2012; **54:** 1321–30.
62 Wahlund L-O, Westman E, van Westen D, Wallin A, Cavallin L, Larsson E-M. Strukturell hjärnavbildning kan förbättra diagnostiken vid demens. *Läkartidningen* 2013; **110:** CEY4.

Correspondence: Eric Westman, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Novum 5th floor, SE-141 86 Stockholm, Sweden.
(fax: +46 858585470; e-mail: eric.westman@ki.se).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Diagnostic and predictive performances of the visual rating scales (receiver operating characteristic curve analyses).

Figure S2. Interaction between age at disease onset and *ApoE* ε4 status.

Table S1. Performance of visual rating scale cut-offs: sample stratified by age in decades.

Table S2. Diagnostic and predictive performances of combination of optimal cut-offs.

Table S3. Interaction between age at disease onset and *ApoE* ε4 status.

Table S4. Diagnostic performance of different cut-offs in early-onset *ApoE* ε4 non-carrier AD patients. ■