



RESEARCH ARTICLE

Visual atrophy rating scales and amyloid PET status in an Alzheimer's disease clinical cohort

Flavia Loreto¹ , Anastassia Gontsarova², Gregory Scott^{1,3}, Neva Patel⁴, Zarni Win⁴, Christopher Carswell⁵, Richard Perry^{1,5} & Paresh Malhotra^{1,3,5} 

¹Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, UK

²Department of Neuroradiology, Imperial College Healthcare NHS Trust, London, UK

³UK Dementia Research Institute Care Research and Technology Centre, Imperial College London and the University of Surrey, London, UK

⁴Department of Nuclear Medicine, Imperial College Healthcare NHS Trust, London, UK

⁵Department of Neurology, Imperial College Healthcare NHS Trust, London, UK

Correspondence

Paresh Malhotra, Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, UK.

E-mail: p.malhotra@imperial.ac.uk

Tel: +44(0)20-331 33172;

Fax: +44(0)20-331 31783

Received: 31 October 2022; Revised: 8 February 2023; Accepted: 10 February 2023

Annals of Clinical and Translational Neurology 2023; 10(4): 619–631

doi: 10.1002/acn3.51749

Abstract

Objectives: Visual rating scales (VRS) are the quantification method closest to the approach used in routine clinical practice to assess brain atrophy. Previous studies have suggested that the medial temporal atrophy (MTA) rating scale is a reliable diagnostic marker for AD, equivalent to volumetric quantification, while others propose a higher diagnostic utility for the Posterior Atrophy (PA) scale in early-onset AD. **Methods:** Here, we reviewed 14 studies that assessed the diagnostic accuracy of PA and MTA, we explored the issue of cut-off heterogeneity, and assessed 9 rating scales in a group of patients with biomarker-confirmed diagnosis. A neuroradiologist blinded to all clinical information rated the MR images of 39 amyloid-positive and 38 amyloid-negative patients using 9 validated VRS assessing multiple brain regions. Automated volumetric analyses were performed on a subset of patients ($n = 48$) and on a group of cognitively normal individuals ($n = 28$). **Results:** No single VRS could differentiate amyloid-positive from amyloid-negative patients with other neurodegenerative conditions. 44% of amyloid-positive patients were deemed to have age-appropriate levels of MTA. In the amyloid-positive group, 18% had no abnormal MTA or PA scores. These findings were substantially affected by cut-off selection. Amyloid-positive and amyloid-negative patients had comparable hippocampal and parietal volumes, and MTA but not PA scores correlated with the respective volumetric measures. **Interpretation:** Consensus guidelines are needed before VRS can be recommended for use in the diagnostic workup of AD. Our data are suggestive of high intragroup variability and non-superiority of volumetric quantification of atrophy over visual assessment.

Introduction

All recent diagnostic criteria^{1,2} recommend the use of MRI as the first step after clinical evaluation of suspected Alzheimer's Disease (AD) to examine patterns of grey matter atrophy, exclude brain lesions, and assess vascular white matter disease and microbleeds. Despite an increasing availability of in vivo biomarkers of AD neuropathology, MRI remains one of the most widely used and accessible diagnostic examinations for AD.³

In current clinical practice, the interpretation of MR images relies on visual assessment, with most reports

consisting of a qualitative description of abnormal brain features. Several automated and semi-automated quantification methods have been developed with the goal of making MRI assessment less subjective and more sensitive to neuroanatomical changes. However, visual rating scales (VRS) represent a simpler, more accessible approach to obtain quantitative measures of atrophy. These involve visual assessment of images and rating the degree of atrophy as an integer score on a Likert scale. The clinical implementation of VRS is relatively widespread compared to that of segmentation tools which, despite significant progress, are scarcely used in clinical practice due to

limited resources, time constraints, and unclear reliability at the individual level.³⁻⁵ Moreover, most studies to date suggest diagnostic equivalence between MRI segmentation and VRS.^{4,6}

The most widely used and validated VRS for the assessment of the MTL is Scheltens and colleagues' medial temporal atrophy (MTA) scale.⁷ This scale's diagnostic accuracy has been extensively investigated in AD, but findings are mixed. A recent meta-analysis found pooled sensitivity and specificity, respectively, equal to 74% and 88% for its use in the differentiation of AD from healthy controls (HC) and concluded that MTA is a reliable diagnostic marker for AD.⁸ However, sensitivity and specificity varied widely across the reviewed studies, especially for the identification of AD from other forms of dementia, which is of relevance to clinical practice. In these cases, MTA sensitivity ranged between 50% and 89% and specificity between 11% and 94%.⁸ Heterogeneity among studies was attributed to differences in the MTA rating methods used.⁸ However, the use of different cut-offs for the definition of rating scale abnormalities is also a major issue and source of heterogeneity that very few studies seem to have acknowledged and attempted to address so far.⁹⁻¹²

In view of the association between age and atrophy, age-specific VRS cut-offs have been proposed with the aim of maximizing scale sensitivity and specificity by differentiating "age-appropriate" atrophy from pathological atrophy. However, conservative cut-offs increase specificity at the expense of sensitivity while more liberal cut-offs present the converse problem. Different research teams have attempted to identify age-adjusted cut-offs that provide the best balance between the two, but these have often come to different conclusions. For example, at least four different sets of age-specific cut-offs have been proposed for the MTA to date^{7,9,10,11,13} (see Table S1). Other researchers, instead, have established different cut-off scores according to clinical populations rather than age groups.^{14,15} Moreover, studies have differed in the way they analyzed the left and right hemisphere scores, with some using the sum, others the highest score, while the majority incorporated the mean of both sides (see Table S2). Despite representing a major source of heterogeneity, the cut-off issue seems to be incompletely acknowledged in the literature, to the point that some studies on VRS diagnostic accuracy have not reported the scores used to determine abnormality.^{16,17}

Evidence that 20–30% of AD patients, especially those with early onset (EOAD), show parietal-dominant patterns of atrophy^{18,19} led to the development of the posterior atrophy (PA) scale.²⁰ In Table S2, we have reviewed the data extracted from 14 peer-reviewed research articles found using PubMed as the main database (in July 2022). Studies

were selected if these assessed the diagnostic and/or prognostic role of Koedam et al.'s Posterior Atrophy (PA) scale,²⁰ Scheltens et al.'s MTA scale,⁷ and a variable range of other VRS, in AD patients. An exhaustive review of all available literature is beyond the scope of this study, but this overview provides insight into the role of PA and MTA, separately and in combination, in AD. Currently available evidence converges towards incremental diagnostic value of PA over MTA for the identification of EOAD from younger controls and other forms of dementia.^{10,14,15,19,21,22}

Besides the MTA and PA, other VRS for the assessment of orbitofrontal,²³⁻²⁵ fronto-insular (FI), anterior cingulate (AC), and entorhinal^{26,27} areas, have been developed for the identification of AD from other forms of dementia, such as frontotemporal dementia (FTD).^{15,28} However, their implementation in research and clinical settings is limited thus far. In addition to atrophy measures, small vessel disease, which is typically seen as areas of white matter hyperintensity (WMH) on fluid-attenuated inversion recovery (FLAIR) MR images, is often assessed using the Fazekas scales for the quantification of deep white matter hyperintensities (DWMH) and periventricular hyperintensity (PVH).²⁹

Few studies have directly compared the diagnostic performance of multiple scales in the same cohort.^{14,15,30} Here, we employed a comprehensive set of VRS for the assessment of MRI images by an expert neuroradiologist blinded to all clinical information. The aims of the study were to (i) examine the patterns of atrophy as measured by expert interpretation of MRI using validated VRS in a biomarker-confirmed clinical cohort, (ii) explore the clinical correlates of VRS scores in patients with Alzheimer's pathology, (iii) test the association between visual and automated quantification of regional atrophy, (iv) critically evaluate the impact of VRS cut-off selection on research findings.

Methods

Subjects

One-hundred consecutive patients seen at the Imperial Memory Clinic between 2013 and 2015 were reviewed and included in the present study if they met the following inclusion criteria: (i) they were referred to our Memory Clinic for the investigation of possible AD; (ii) the diagnostic workup included both MRI and amyloid PET; (iii) if performed externally, the MR images were made available to our Centre; (iv) decision to perform amyloid PET followed appropriate use criteria³¹ and was made by consensus within the multidisciplinary team.^{32,33} If patients had more than one MRI scan, we selected the

scan preceding amyloid PET. Where this preceded amyloid PET by more than 12 months, we selected the one following amyloid PET if this was available and performed within 6 months. A total of 77 patients, 39 amyloid-positive (A β -pos) and 38 amyloid-negative (A β -neg), met the inclusion criteria. In addition, the MR images of 28 cognitively normal (CN) individuals (mean age \pm SD = 71.88 \pm 5.99 years) scanned at ICHT for research purposes were acquired for volumetric analyses.

MRI acquisition

Most patients (90%) were scanned at our Centre using a 1.5 T Siemens MAGNETOM Avanto (repetition time = 900 ms; echo time = 3.37 ms; 160 slices/slab, voxel size = 1 \times 0.5 \times 0.5 mm). Patients scanned at a different centre before referral (n = 8) were not excluded to ensure consecutive selection. All patients had at least a T1-weighted image available plus a variable range of additional sequences, such as FLAIR and T2*/SWI, which were included for the assessment of WMH and cerebral microbleeds. All CN individuals were scanned at ICHT using a 3T Siemens MAGNETOM Verio (repetition time = 900 ms; echo time = 2.52 ms; 176 slices/slab, voxel size = 1 \times 1 \times 1 mm.) T1-weighted images were extracted to perform volumetric analyses.

MRI analysis

Visual rating protocol

Visual ratings of the complete dataset (n = 77) were performed by a single expert neuroradiologist (A.G.) blinded to all clinical and pathological information, except patient

age at MR scanning. Images were rated in the native space in keeping with standard clinical reads using Research PACS version 12.1.5.1046 (Philips Vue PACS). Cerebral atrophy across seven brain regions as well as deep and periventricular WMH were rated using 9 validated rating scales (Table 1). To ensure consistency, an electronic proforma was created to record visual rating scores and publicly available reference images were used at the training stage and throughout the study. For more information on the reference images used in this study see Enkirch et al (2018) for the ERICA scale,²⁶ Wardlaw and Ferguson's guide for the FAZKAS scale (https://www.ed.ac.uk/sites/default/files/atoms/files/white_matter_rating_qualitative_final.pptx), and Harper et al.'s supplementary material for the remaining scales.¹⁵ For all atrophy scales, left and right hemisphere scores were averaged to obtain a single score. Moreover, the rater recorded whether other neuropathologies (e.g., strategic infarct, normal pressure hydrocephalus, demyelination) were visible on MR images.

Volumetric analyses

To explore the relationship of MTA and PA scores with the respective volumes, volumetric analyses were performed on a subset of patients (n = 48) whose T1 images met FreeSurfer requirements, as well as on the CN group (n = 28). The *recon-all* function of FreeSurfer 6.0 was used to extract estimates of brain volumes (<https://surfer.nmr.mgh.harvard.edu/>)³⁴: hippocampal volumes (HV) were obtained from subcortical output files extracted through *aseg.stats*; parietal volumes were obtained from *aparc.stats* files using the Desikan atlas.³⁵ Within the Desikan atlas, the parietal lobe is segmented across 5

Table 1. Description of visual rating scales.

Scale	Brain regions/features	Plane-modality	Score range
Orbitofrontal (OF) ^{15,24}	Corpus callosum, width of the olfactory sulcus	Coronal-T1	0–3
Anterior cingulate (AC) ^{15,25}	Corpus callosum, width of the cingulate sulcus	Coronal-T1	0–3
Frontoinsula (FI) ^{15,24}	Width of the circular insular sulcus, anterior commissure	Coronal-T1	0–3
Anterior temporal (AT) ^{15,24}	Width of the anterior temporal sulci, aspects of the temporal pole	Coronal-T1	0–4
Medial temporal atrophy (MTA) ⁷	Width of the choroid fissure, width of the temporal horn, height of the hippocampus	Coronal-T1	0–4
Entorhinal cortex (ERICA) ²⁶	volume of the entorhinal cortex and parahippocampal gyrus, width of the collateral sulcus, width of cleft between the entorhinal cortex and the cerebellar tentorium	Coronal-T1	0–3
Posterior atrophy (PA) ²⁰	Width of the posterior cingulate and parieto-occipital sulci, volume of the parietal lobes, and precuneus	Coronal, axial, sagittal-T1	0–3
Fazekas ²⁹	areas of abnormal high signal intensity around the ventricles (periventricular, PVH) and in the deep white matter (DWMH)	Axial-FLAIR	0–3

regions in each hemisphere, namely the postcentral gyrus, the supramarginal gyrus, the superior and inferior parietal cortex, and the precuneus.³⁵ The “mean parietal area” was obtained by averaging these 5 regions, separately for the left and right hemispheres. Recon-all logs and images were inspected for major segmentation errors.

Amyloid PET image acquisition and clinical reads

All 77 patients included in this study were scanned using a Siemens Biograph 64 PET/CT scanner. The amyloid PET procedure and clinical interpretation were performed as described elsewhere.^{36,37} Briefly, images were acquired following post-injection interval of intravenous 18F-Florbetapir or 18F-Florbetaben and were visually read as positive or negative by an experienced nuclear medicine radiologist. Equivocal cases were independently read by two nuclear medicine experts and by a third reader when there was disagreement.

Clinical measures and diagnostic categorization

Clinical notes were retrospectively reviewed to determine the final clinical diagnosis, the syndromic stage at the time of MRI, the duration of cognitive impairment at the time of presentation to our Centre, and the age of onset of cognitive symptoms. For the A β -pos group, we also recorded whether clinical presentation was characterized by amnesic or non-amnesic symptoms. The A β -neg group was further divided into “progressive A β -neg” or “stable A β -neg” based on the most likely cause of cognitive impairment.³⁶ Patients were classified as “progressive” if they showed evidence of symptom progression over follow-ups that was suggestive of a nonAD neurodegenerative process or if they had concomitant neurological conditions affecting brain integrity. Patients were defined as “stable” if the course of cognitive symptoms and the results of clinical investigations were suggestive of a non-neurodegenerative cause of cognitive symptoms.

Score dichotomization

VRS scores were dichotomized into “normal” or “abnormal” using age-adjusted cut-offs, with “normal” scores indicating age-appropriate levels of atrophy. Where possible, we used validated VRS cut-offs based on published data.^{10,13} However, for the OF, AC, AT, FI, and ERICA, cut-offs had to be established based on agreement between the clinical and research teams due to the lack of existing recommendations (Table S3).

Statistical analysis

Demographic information was compared between groups using Student’s *t*-test for continuous data and χ^2 test for categorical data.

Visual assessment

Higher visual rating scores indicate worse levels of atrophy. In this study, and in line with previous work, the left (LH) and right (RH) hemispheres scores were averaged to obtain a single score. Missing values, due to unavailable MR sequences or low-quality images (Fig. 1), were not imputed. The distribution of VRS scores was tested through Shapiro–Wilk test, and group differences between the A β -pos, stable A β -neg, and progressive A β -neg groups in non-normally distributed data were analyzed using Kruskal–Wallis non-parametric test and Mann–Whitney *U* for post hoc analyses. Separate Mann–Whitney *U* tests were performed on the amyloid-positive and amyloid-negative groups to test for the association between amyloid PET status and VRS score, irrespective of clinical group. In the A β -pos group, we used Mann–Whitney *U* to test whether MTA and PA status (normal versus abnormal) was associated with age at the time of MRI, age of onset, duration (months) of cognitive impairment, while χ^2 tests were used to assess the association with sex and phenotype (amnesic versus non-amnesic). These two regions were selected here to explore possible predictors of more (i.e., MTA) or less (i.e., PA) typical patterns of atrophy in this clinical population. The association between categorical variables was tested through χ^2 tests. In all post hoc analyses, *p*-values were adjusted for multiple comparisons using Bonferroni correction.

Visual and quantitative assessment

The relationship between visual and quantitative measures of MTL and parietal volumes was tested separately for the left and right hemispheres ($n = 48:22$ A β -pos, 26 A β -neg). Partial correlation analyses, controlling for age, sex, and estimated total intracranial volume (eTIV), were used to investigate the association between HV and MTA scores and between mean parietal area and PA scores. In addition, multiple correlation analyses were run to test the association between left and right PA scores and each of the 5 lateralized Desikan parietal regions separately. The mean left and right HVs were compared between the A β -pos ($n = 22$), A β -neg ($n = 26$) and CN ($n = 28$) using an ANCOVA with age, sex, and eTIV as covariates. The left and right parietal areas were compared between the three groups using an ANCOVA with age, sex, and eTIV as covariates. Post hoc comparisons were performed on significant *p*-values.

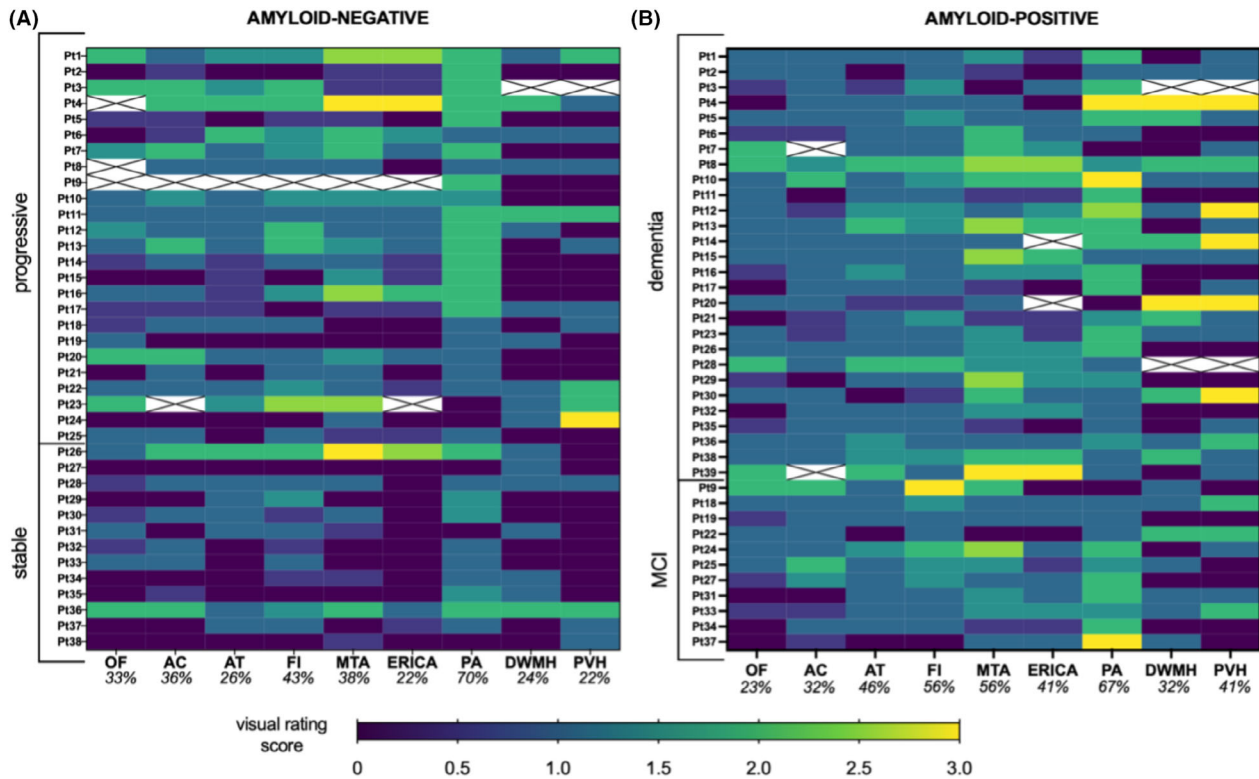


Figure 1. Heatmaps showing individual-level visual rating scores in the amyloid-negative (A) and amyloid-positive (B) groups. Crossed-out cells indicate missing visual rating score for that patient due to unavailable MR sequences. Percentages indicate the proportion of abnormal scores in each scale, dichotomized as per Table S3.

Results

Demographics

Clinical and demographic information is provided in Table 2. A β -pos and A β -neg patients did not significantly differ for age ($t_{(75)} = 0.58$, $p = 0.56$) or sex ($\chi^2_{(1)} = 0.11$, $p = 0.74$) (Table 2).

Group comparisons

All A β -pos patients received a clinical diagnosis of AD in line with National Institute on Ageing and Alzheimer's Association (NIA-AA).¹ A β -neg patients, instead, formed a heterogeneous group, which we subdivided into "stable A β -neg" ($n = 12$) and "progressive A β -neg" ($n = 26$) (Table 2), as described above. Nonparametric testing on the three groups (i.e., A β -pos, stable A β -neg, and progressive A β -neg) highlighted significant differences in the AC (Kruskal-Wallis $H_{(2)} = 6.08$, $p = 0.048$), AT (Kruskal-Wallis $H_{(2)} = 7.66$, $p = 0.02$), FI (Kruskal-Wallis $H_{(2)} = 6.24$, $p = 0.044$), MTA (Kruskal-Wallis $H_{(2)} = 13.1$, $p = 0.001$), ERICA (Kruskal-Wallis $H_{(2)} = 16.46$,

$p < 0.001$), PA (Kruskal-Wallis $H_{(2)} = 6.73$, $p = 0.03$) scales, with the PVH approaching significance (Kruskal-Wallis $H_{(2)} = 5.72$, $p = 0.057$). Bonferroni-corrected Mann-Whitney U tests for each pairwise comparison revealed significantly higher scores in the A β -pos group than the stable A β -neg group in the AT ($p = 0.01$), FI ($p = 0.02$), MTA ($p = 0.001$), and ERICA ($p < 0.001$), with PA differences approaching significance ($p = 0.054$). The progressive A β -neg group, instead, had significantly higher scores than the stable A β -neg group in the MTA ($p = 0.001$), ERICA ($p = 0.006$), and PA ($p = 0.045$) scales. None of the scales differentiated the A β -pos from the progressive A β -neg group (lowest p -value = 0.96, for the PVH scale). Heatmaps in Figure 1 illustrate the individual-level patterns of scores in the three groups. Mann-Whitney U tests revealed significantly higher AT ($U = 522.5$, $p = 0.025$), MTA ($U = 527.5$, $p = 0.04$), ERICA ($U = 419.5$, $p = 0.005$), and PVH ($U = 489.5$, $p = 0.023$) scores in the amyloid-positive group. In Figure S1 the mean rating scores of the current study population are plotted against those extracted from previous studies^{14,15} for direct comparison.

Table 2. Demographic and clinical information according to amyloid PET status.

	A β -pos (<i>n</i> = 39)	A β -neg (<i>n</i> = 38)
<i>Demographics</i>		
Mean age \pm SD (years)	67.19 \pm 7.84	66.14 \pm 8.14
Age range	49.4–81.1	53.8–87.9
Sex (%female)	41.02%	44.74%
<i>Clinical information</i>		
Duration of CI (months)	34.34 \pm 25.67 ¹	40.76 \pm 31.92 ²
Median interval between MRI and amyloid PET (days)	148	210
%Early age of onset (<65 years)	51%	45%
Clinical diagnosis, <i>n</i> (%)	<ul style="list-style-type: none"> • Amnesic AD, 28(72%) • Nonamnesic AD, 11(28%)³ 	<ul style="list-style-type: none"> • Stable MCI, 12(32%) • Other dementia, 10(26%)⁴ • Other neurological condition, 5(13%)⁵ • Progressive MCI, 11(29%)

AD, Alzheimer's disease; A β -neg, amyloid-negative; A β -pos, amyloid-positive; CI, cognitive impairment; DLB, dementia with Lewy bodies; FTD, frontotemporal dementia; MCI, mild cognitive impairment.

¹Available for 38/39 patients.

²Available for 33/38 patients.

³Nonamnesic AD: visuospatial *n* = 5, language *n* = 5, behavioral *n* = 1.

⁴Other dementia: dementia with Lewy bodies *n* = 3, frontotemporal dementia *n* = 7.

⁵Other neurological conditions: normal pressure hydrocephalus *n* = 2, temporal lobe epilepsy *n* = 2, multiple cavernomas *n* = 1.

Posterior and medial temporal atrophy: visual assessment

MTA and PA scores were dichotomized according to Rhodius-Meester et al.'s¹⁰ age-adjusted cut-offs (see Table S1). Findings are shown in Figure 2. A β -pos patients with normal MTA were comparable to those with abnormal MTA in all clinical and demographic variables except for age at the time of MRI, which was significantly lower in patients with abnormal MTA scores (normal: 67.1 \pm 6.71 years, abnormal: 62.17 \pm 8.91 years; *U* = 105 *p* = 0.02). A β -pos patients with abnormal PA score were significantly younger at the time of MRI scanning (normal: 71.8 \pm 5.61 years, abnormal: 64.9 \pm 7.87 years; *U* = 78 *p* = 0.006) as well as at the time of symptom onset (normal: 69.87 \pm 5.8, abnormal: 61.3 \pm 8; *U* = 58 *p* = 0.001). A β -pos patients with abnormal PA had also longer duration of cognitive symptoms at the time of their first memory clinic visit (normal: 23.1 \pm 17.4 months, abnormal: 40.2 \pm 27.6 months; *U* = 95 *p* = 0.038) and presented more frequently with non-amnesic symptoms compared to those with normal PA scores (% non-amnesic: normal 7.7%, abnormal 38.5%; $\chi^2_{(1)} = 4.05$, *p* = 0.04).

Posterior and medial temporal atrophy: quantitative assessment

There was a significant negative association between left and right MTA scores and the respective hippocampal volumes (HV) (LH: *r* = -0.41 *p* = 0.006; RH: *r* = -0.47

p = 0.001) after controlling for age, sex, and ϵ TIV. Mean left HV was 15.2% lower in subjects with left MTA scores ≥ 2 compared to those with normal MTA scores (3631mm³ vs 3079mm³). Mean right HV was 14.6% lower in subjects with right MTA scores ≥ 2 compared to those with normal MTA scores (3757 mm³ vs 3209 mm³). On the other hand, the left and right PA scores did not show a significant association with the mean parietal area (LH: *r* = -0.28 *p* = 0.06; RH: *r* = -0.21, *p* = 0.17) and, similarly, no significant correlations were found between PA scores and each parietal region separately.

The left ($F_{(2,70)} = 4.85$, *p* = 0.011) and right ($F_{(2,70)} = 4.51$, *p* = 0.014) inferior parietal cortices were the only parietal regions that significantly differed between groups, with smaller left inferior parietal cortex in the A β -pos group than both the A β -neg (LH: *p* = 0.034) and controls (LH: *p* = 0.019) groups, and smaller right inferior parietal cortex in the A β -pos group than controls (RH: *p* = 0.018) (Fig. 3A). There were significant differences in both the left ($F_{(2,70)} = 7.83$, *p* = 0.001) and right ($F_{(2,70)} = 5.33$, *p* = 0.007) hippocampi (Fig. 3B), which were due to smaller volumes in the A β -pos (LH: *p* = 0.002, RH: *p* = 0.02) and the A β -neg (LH: *p* = 0.004, RH: *p* = 0.02) groups compared to controls; A β -pos and A β -neg groups did not significantly differ (*p* = 1.00).

Alzheimer's disease group's characterization

The individual VRS score profiles varied widely within the A β -pos group (Fig. 1B), despite all patients having a

		A β -pos		A β -neg	
		MTA		MTA	
		normal	abnormal	normal	abnormal
PA	normal	18%	15%	27%	1%
	abnormal	26%	41%	35%	35%

Figure 2. Combination of MTA and PA status in the amyloid-positive and amyloid-negative groups.

clinical diagnosis of AD. Individual-level number of abnormal scores ranged between 0 and 7. Notably, only 7.7% of AD patients had all scale scores in the abnormal range while 13% had no abnormal scores, indicating “age-appropriate” atrophy across all cortical regions. Within the A β -pos patients, the MTA and PA scales were the ones with the highest median score (median = 1.5, IQR = 1 for both) (Fig. 4). Abnormal scores on either the MTA or PA correctly identified 82% A β -pos patients; however, abnormality in any of these two scales was also seen in 73% A β -neg (stable and progressive) patients. With reference to the remaining scales, the A β -pos group showed high proportions of abnormal scores in the FI (56%) and the AT (46%) scales, while the lowest proportion was seen in the OF scale (23%) (Fig. 1B).

The impact of cut-off selection

It is important to note that we had to establish operational cut-off values for the OF, AC, AT, FI, and ERICA scales due to the unavailability of validated age-adjusted norms. Changing the cut-off value from 1.5 to 2 for the 65–74 A β -pos age group (in line with PA cut-offs) for these scales would lead to a reduction of abnormal ERICA scores in the A β -pos group from 41% to 35%, FI scores from 56% to 44%, AT scores from 46% to 41%, and AC scores from 32% to 30%, while abnormal OF scores would remain unchanged. Moreover, even where validated cut-offs exist, these vary between studies. Replacing Rhodius-Meester et al.’s¹⁰

MTA and PA cut-off values with those suggested by Ferreira and colleagues¹¹ (Table S1) in the A β -pos group would lead to a reduction in MTA abnormality (from 56% to 49%) and a striking increase in PA abnormality (from 67% to 92%). This would corroborate prominent involvement of the posterior regions of the brain in our A β -pos group as well as higher sensitivity of the PA scale. However, a similarly high proportion of A β -neg patients would be classified as abnormal (86%) using these cut-offs.

Clinical correlates of Alzheimer’s disease

OF scores were significantly higher in the amnesic group ($n = 26$) ($U = 71$, $p = 0.005$), while PA ($U = 92.5$, $p = 0.045$) and PVH ($U = 84.5$, $p = 0.04$) scores were significantly higher in the non-amnesic group ($n = 11$) (Fig. S2). Mean MTA score was higher in the amnesic (mean \pm SD = 1.52 \pm 0.9) than the non-amnesic (mean \pm SD = 1.14 \pm 0.5) group, but this did not reach significance ($p = 0.197$). Syndromic stage was not associated with VRS scores (p between 0.11 and 0.99). About half (49%) of A β -pos patients had deep and/or periventricular WMHs, compared with 35% in the A β -neg group, and their presence was not associated with atrophy scores (A β -pos: p between 0.21 and 1, A β -neg: p between 0.062 and 0.775). Notably, 31% of Alzheimer’s disease patients had findings suggestive of other pathology including previous cerebral infarction, small vessel disease, and previous brain injury.

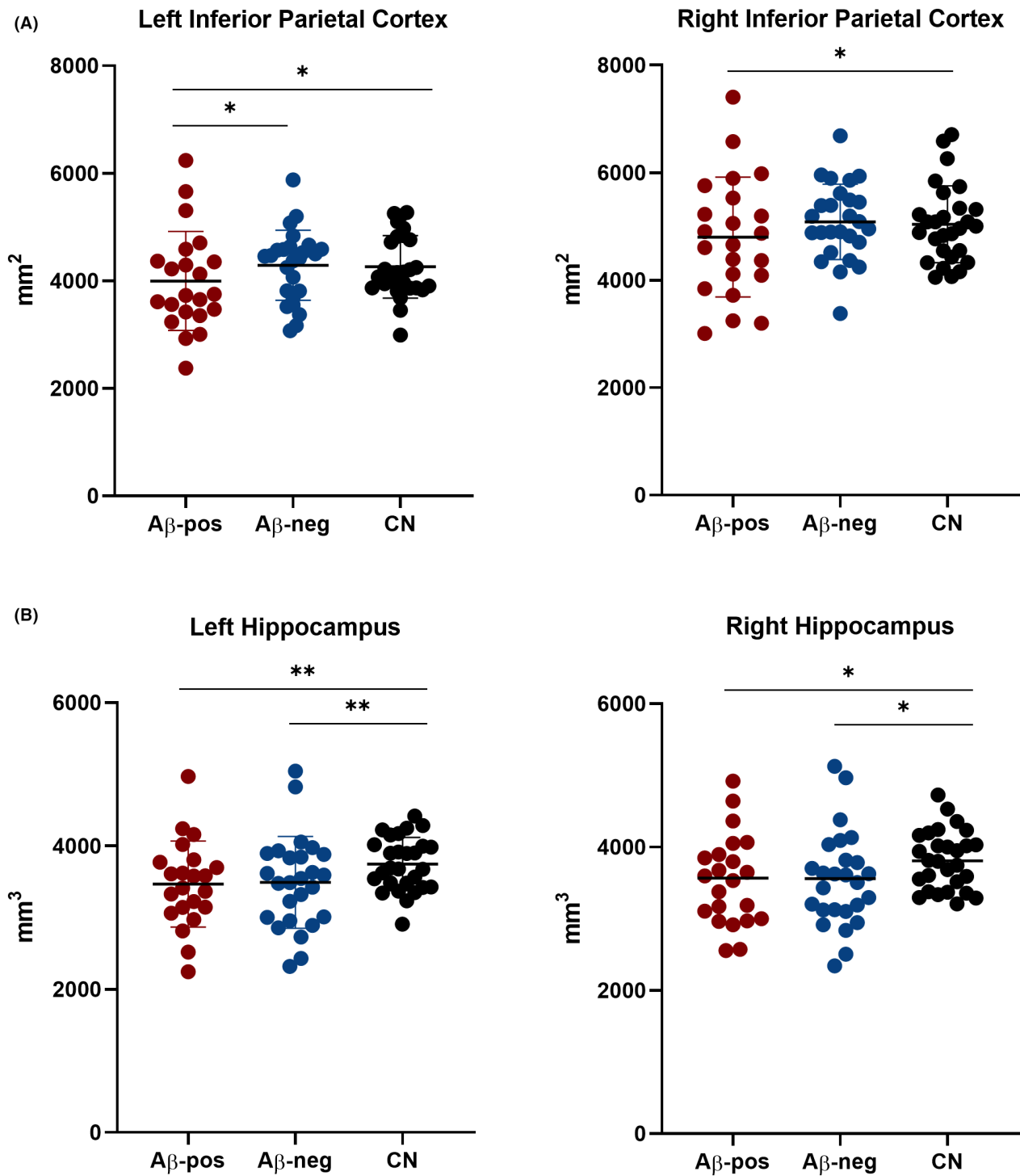


Figure 3. Quantification of the left and right parietal (A) and hippocampal (B) volumes in the amyloid-positive, amyloid-negative, and cognitively normal groups. ANCOVAs with “Group” as independent variable, and age, sex, and eTIV as covariates, revealed significant group differences for the hippocampal but not the parietal volumes. * $p < 0.05$; ** $p < 0.01$; ns, nonsignificant.

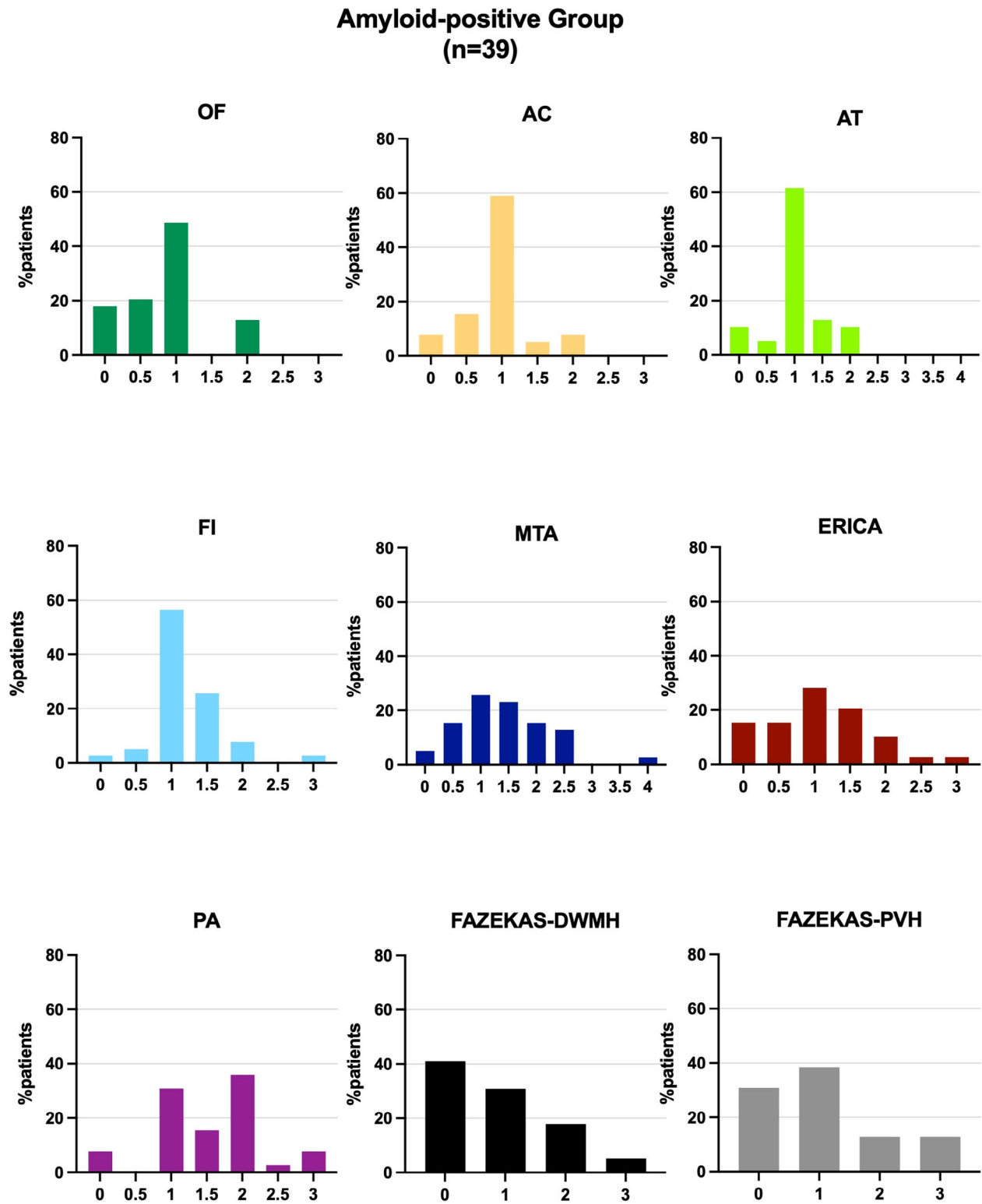


Figure 4. Distribution of visual rating scores for each rating scale. Plotted scores are the result of the average between the left and right hemisphere scores.

DISCUSSION

In this study, we examined atrophy on structural MRI, as measured by expert visual assessment, in a biomarker-confirmed clinical cohort of patients with suspected Alzheimer's disease.

The amyloid-positive group had worse scores than the stable amyloid-negative group in the MTA, PA, ERICA, AT, FI, but not the OF scale, whereas the progressive amyloid-negative group had worse MTA, ERICA, and PA mean scores than the stable amyloid-negative group. No single VRS differentiated amyloid-positive from progressive amyloid-negative patients. This is in line with previous studies finding lower diagnostic utility of a single brain region when comparing AD to other forms of dementia, rather than to controls,¹⁴ due to limited specificity.^{38,39} Furthermore, AD may frequently present with extra-temporal atrophy and relative sparing of the MTL.^{40,41} In our clinical cohort, 44% of amyloid-positive patients were deemed to have age-appropriate levels of MTL atrophy. Unexpectedly, patients with a normal MTA score were on average older than those with abnormal MTA. This is in contrast with most available literature showing that hippocampal-sparing forms of AD are associated with younger onset.⁴¹ There are two possible explanations for these findings: firstly, patients who are referred to amyloid-PET scanning tend to present with atypical clinical features leading to diagnostic uncertainty and consequent need for biomarker examination, and atypically young age and inconclusive findings on standard diagnostic investigations are among those features.^{31,33} It is, therefore, possible that patients who present with a typical Alzheimer-like pattern of neurodegeneration, with involvement of the MTL, may be referred for amyloid PET on the basis of atypically young age in order to increase diagnostic confidence. Secondly, individuals presenting with atypical symptoms may take longer to be referred to a specialist clinic,⁴² meaning that they may be older and at more advanced disease stages by the time of amyloid PET referral.

We reviewed 14 research studies that assessed the performance of the MTA and PA scales, converging towards a possible incremental diagnostic value of PA in EOAD. In our amyloid-positive group, patients with abnormal PA were indeed younger, had earlier age of onset and presented more often with non-amnesic symptoms and longer disease duration compared to those with normal PA scores. The proportion of patients with a combination of abnormal PA and normal MTA scores (26%) was comparable to that found by previous studies of similar patients, while the proportion of those with abnormal MTA and PA (41%) was higher.^{10,19} With reference to the remaining scales, the FI and the AT were the only

two atrophy scales with over 40% of amyloid-positive subjects having an abnormal score. The FI was also found to be promising in Harper et al.'s study, in which this was the best single scale, together with PA, for the identification of EOAD.¹⁵ The AT scale, instead, was found by previous studies to be better suited for the differentiation of other clinical groups, such as FTD and DLB.^{14,15} The lowest proportion of abnormal scores and the lack of significant group differences in the OF scale provide indirect support for the specificity of this scale for the detection of other forms of dementia, such as FTD.^{15,28} Notably, about half (49%) of amyloid-positive patients had evidence of periventricular and/or deep WMH, despite their relatively young age, and this proportion was higher than that seen in the amyloid-negative group (35%). Furthermore, the analysis of group differences according to amyloid PET status revealed significantly higher PVH scores in the amyloid-positive than in the amyloid-negative group. These findings are not in line with previous studies¹⁰ and would suggest WMH as a possible imaging feature of this specific clinical population.

The lack of association between hippocampal volumes and amyloid PET status is in keeping with previous studies suggesting non-superiority of volumetric quantification of the MTL over visual assessment.^{4,6} The extent of overlap of hippocampal volumes in patients with and without Alzheimer's pathology was striking and reinforces the need to rethink the role of this measure in AD diagnosis, especially in younger and atypical patients,²² who are frequently referred for amyloid PET.^{31,33} Therefore, it could be argued that the present findings do not apply to the wider population of typical AD. On the other hand, studying this clinical population furthers our understanding of the disease mechanisms driving heterogeneity in AD and helps reappraise the diagnostic value of standard diagnostic markers for less typical forms of AD.

The absence of an association between visual and volumetric measures of parietal lobe volume was against our predictions and could be attributed to the two methods measuring different features of the same region. As proposed by Fumagalli and colleagues, the visual rating score may reflect more the widening of a sulcus and the increase in CSF, while the volumetric analyses are more closely associated with both grey and white matter volume.⁴³

Our study's limitations include the lack of a measure of disease severity, such as the Mini-Mental State Examination. Another limitation is that visual rating scores were provided once and by a single rater. This means that we were not able to assess intra- and inter-rater reliability directly. However, it is important to note that most of these scales have been found to have good intra- and inter-rater reliabilities.¹⁵ Moreover, due to the

unavailability of visual scores in the healthy control group, we were unable to formally examine the sensitivity and specificity of these scales. Finally, the power to detect differences between the three groups was limited by the relatively small sample size of the amyloid-negative group.

In conclusion, this study has highlighted the key issues that require addressing before visual rating scales can be recommended for regular use in the diagnostic workup of AD. Consensus guidelines on the definition of VRS abnormality are necessary to determine the optimal use of these scales in clinical practice. Future studies should also validate abnormality cut-offs for each hemisphere, which would enable better detection of unilateral or asymmetric atrophy. We also showed the importance of biomarker confirmation of AD to assess the performance of standard diagnostic examinations in the atypical AD population. Atypical forms may go undetected in studies of clinically diagnosed AD because of the lack of the typical hallmarks of the disease, with the consequent risk of a selection bias towards those with typical atrophy patterns. The intragroup heterogeneity shown by this study points to the need to characterize AD neuroimaging profiles, especially in the presence of atypical phenotypes, for the potential identification of more sensitive and specific markers. While in the medium-long term AD diagnosis is increasingly likely to rely on biomarkers of amyloid and tau, the present work is highly relevant to current clinical practice, which still places a major focus on the visual assessment of MRI scans.

Funding Information

The work was funded by Alzheimer's Society (grant number P75464) and supported by the NIHR Biomedical Research Centre at Imperial College London. None of the funders was involved in the conduct of the study or preparation of the article.

Author Contributions

FL designed the study, collected and analysed the data and wrote the manuscript; AG designed the study and performed the visual ratings; NP and ZW contributed to data collection; RP conceived the idea and designed the study; PM conceived the idea, designed the study and wrote the manuscript; all authors reviewed and approved the manuscript.

Acknowledgments

The authors thank the Imperial College Healthcare NHS Trust staff and Tina Stoycheva for the support in collecting neuroimaging data.

Conflict of Interest Statement

ZW previously participated in the Eli Lilly PET advisory board and was an amyloid-PET read trainer. CC has taken part in an advisory panel for Roche Pharmaceuticals. RP previously sat on an advisory board for Eli Lilly and received support from GE for research imaging from 2014 to 2018. PM has given an educational talk at a meeting organized by GE. None of the authors currently have funding or support from any commercial organization involved in amyloid PET imaging. FL, AG, GS, and NP report no disclosures relevant to the manuscript.

Ethics Approval

Ethical approval for this study was obtained from the Camden and Kings Cross UK Research Ethics Committee (REC number 20/LO/0442) and the Cambridge East Research Ethics Committee (REC number 10/H0304/70).

Data Availability Statement

Data not provided in the article are available upon reasonable request.

References

- 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(3):263-269.
- Dubois B, Feldman HH, Jacova C, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet*. 2014;13:614-629.
- Caprioglio C, Garibotto V, Jessen F, et al. The clinical use of Alzheimer's biomarkers in patients with mild cognitive impairment: a European Alzheimer's disease consortium survey. *J Alzheimers Dis*. 2022;89(2):535-551.
- Ingala S, van Maurik IS, Altomare D, et al. Clinical applicability of quantitative atrophy measures on MRI in patients suspected of Alzheimer's disease. *Eur Radiol*. 2022 May 31;32:7789-7799.
- Martensson G, Hakansson C, Pereira JB, et al. Medial temporal atrophy in preclinical dementia: visual and automated assessment during six year follow-up. *Neuroimage Clin*. 2020;27:102310.
- Persson K, Barca ML, Cavallin L, et al. Comparison of automated volumetry of the hippocampus using NeuroQuant (R) and visual assessment of the medial temporal lobe in Alzheimer's disease. *Acta Radiol*. 2018;59(8):997-1001.
- 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(10):967-972.
8. Park HY, Park CR, Suh CH, Shim WH, Kim SJ. Diagnostic performance of the medial temporal lobe atrophy scale in patients with Alzheimer's disease: a systematic review and meta-analysis. *Eur Radiol*. 2021;31(12):9060-9072.
 9. 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(3):317-330.
 10. Rhodius-Meester HFM, Benedictus MR, Wattjes MP, et al. MRI visual ratings of brain atrophy and white matter Hyperintensities across the Spectrum of cognitive decline are differently affected by age and diagnosis. *Front Aging Neurosci*. 2017;9:117.
 11. Ferreira D, Cavallin L, Larsson EM, et al. 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;278(3):277-290.
 12. 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(3):331-333.
 13. Claus JJ, Staekenborg SS, Holl DC, et al. Practical use of visual medial temporal lobe atrophy cut-off scores in Alzheimer's disease: validation in a large memory clinic population. *Eur Radiol*. 2017;27(8):3147-3155.
 14. Falgas N, Balasa M, Bargallo N, et al. Diagnostic accuracy of MRI visual rating scales in the diagnosis of early onset cognitive impairment. *J Alzheimers Dis*. 2020;73(4):1575-1583.
 15. Harper L, Fumagalli GG, Barkhof F, et al. MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain*. 2016;139(Pt 4):1211-1225.
 16. Zhu H, Lu H, Wang F, et al. Characteristics of cortical atrophy and white matter lesions between dementia with Lewy bodies and Alzheimer's disease: a case-control study. *Front Neurol*. 2021;12:779344.
 17. 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(1):111-119.
 18. Noh Y, Jeon S, Lee JM, et al. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. *Neurology*. 2014;83:1936-1944.
 19. Lehmann M, Koedam EL, 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(3):627 e1-e12.
 20. Koedam EL, Lehmann M, van der Flier WM, et al. Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol*. 2011;21(12):2618-2625.
 21. Sheng C, Sun Y, Wang M, et al. Combining visual rating scales for medial temporal lobe atrophy and posterior atrophy to identify amnesic mild cognitive impairment from cognitively Normal older adults: evidence based on two cohorts. *J Alzheimers Dis*. 2020;77(1):323-337.
 22. Falgas N, Sanchez-Valle R, Bargallo N, et al. Hippocampal atrophy has limited usefulness as a diagnostic biomarker on the early onset Alzheimer's disease patients: a comparison between visual and quantitative assessment. *Neuroimage Clin*. 2019;23:101927.
 23. Harper L, Barkhof F, Fox NC, Schott JM. Using visual rating to diagnose dementia: a critical evaluation of MRI atrophy scales. *J Neurol Neurosurg Psychiatry*. 2015;86(11):1225-1233.
 24. Davies RR, Kipps CM, Mitchell J, Kril JJ, Halliday G, Hodges J. Progression in Frontotemporal Dementia. *Arch Neurol*. 2006;63:1627-1631.
 25. Davies RR, Schill VL, Graham A, Williams GB, Graham KS, Hodges JR. Development of an MRI rating scale for multiple brain regions: comparison with volumetrics and with voxel-based morphometry. *Neuroradiology*. 2009;51(8):491-503.
 26. Enkirch SJ, Traschutz A, Muller A, et al. The ERICA score: an MR imaging-based visual scoring system for the assessment of entorhinal cortex atrophy in Alzheimer disease. *Radiology*. 2018;288(1):226-333.
 27. Traschutz A, Enkirch SJ, Polomac N, et al. The entorhinal cortex atrophy score is diagnostic and prognostic in mild cognitive impairment. *J Alzheimers Dis*. 2020;75(1):99-108.
 28. Hornberger M, Savage S, Hsieh S, Mioshi E, Piguet O, Hodges JR. Orbitofrontal dysfunction discriminates behavioral variant frontotemporal dementia from Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2010;30(6):547-552.
 29. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and Normal aging. *AJR Am J Roentgenol*. 1987;149:351-356.
 30. Yuan Z, Pan C, Xiao T, et al. Multiple visual rating scales based on structural MRI and a novel prediction model combining visual rating scales and age stratification in the diagnosis of Alzheimer's disease in the Chinese population. *Front Neurol*. 2019;10:93.
 31. Johnson KA, Minoshima S, Bohnen NI, et al. Appropriate use criteria for amyloid PET: a report of the amyloid imaging task force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *Alzheimers Dement*. 2013;9(1):e-1-e-16.
 32. Carswell CJ, Win Z, Muckle K, et al. Clinical utility of amyloid PET imaging with (18)F-florbetapir: a retrospective study of 100 patients. *J Neurol Neurosurg Psychiatry*. 2018;89(3):294-299.

33. Kolanko MA, Win Z, Loreto F, et al. Amyloid PET imaging in clinical practice. *Pract Neurol*. 2020;20:451-462.
34. Fischl B. FreeSurfer. *Neuroimage*. 2012;62(2):774-781.
35. Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006 Jul 1;31(3):968-980.
36. Loreto F, Gunning S, Golemme M, et al. Evaluating cognitive profiles of patients undergoing clinical amyloid-PET imaging. *Brain Commun*. 2021;3(2):1-12.
37. Loreto F, Fitzgerald A, Golemme M, et al. Prevalence of depressive symptoms in a memory clinic cohort: a retrospective study. *J Alzheimers Dis*. 2022 Jun;23:1-9.
38. Barkhof F, Polvikoski TM, Van Straaten EC, et al. The significance of medial temporal lobe atrophy. A postmortem MRI study in the very old. *Neurology*. 2007;69(15):1521-1527.
39. de Souza LC, Chupin M, Bertoux M, et al. Is hippocampal volume a good marker to differentiate Alzheimer's disease from frontotemporal dementia? *J Alzheimers Dis*. 2013;36(1):57-66.
40. Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol*. 2011;10(9):785-796.
41. Whitwell JL, Dickson DW, Murray ME, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. *Lancet Neurol*. 2012;11(10):868-877.
42. Barnes J, Dickerson BC, Frost C, Jiskoot LC, Wolk D, van der Flier WM. Alzheimer's disease first symptoms are age dependent: evidence from the NACC dataset. *Alzheimers Dement*. 2015;11(11):1349-1357.
43. Fumagalli GG, Basilico P, Arighi A, et al. Parieto-occipital sulcus widening differentiates posterior cortical atrophy from typical Alzheimer disease. *Neuroimage Clin*. 2020;28:102453.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1