



Hippocampal atrophy has limited usefulness as a diagnostic biomarker on the early onset Alzheimer's disease patients: A comparison between visual and quantitative assessment



Neus Falgàs^a, Raquel Sánchez-Valle^a, Núria Bargalló^{c,d}, Mircea Balasa^{a,b},
 Guadalupe Fernández-Villullas^a, Beatriz Bosch^a, Jaume Olives^a, Adrià Tort-Merino^a,
 Anna Antonell^a, Cristina Muñoz-García^a, María León^a, Oriol Grau^a, Magdalena Castellví^a,
 Nina Coll-Padrós^a, Lorena Rami^a, Alberto Redolfi^e, Albert Lladó^{a,*}

^a Alzheimer's disease and other cognitive disorders Unit, Neurology department, IDIBAPS, Hospital Clínic, Barcelona, Spain

^b Atlantic Fellow for Equity in Brain Health, Global Brain Health Institute, Trinity College Dublin, Ireland

^c Imaging Diagnostic Center, Hospital Clínic, Barcelona, Spain

^d Magnetic Resonance Image Core Facility, IDIBAPS, Spain

^e Laboratory of Alzheimer's Neuroimaging and Epidemiology - LANE, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

ARTICLE INFO

Keywords:

Alzheimer's disease
 Frontotemporal dementia
 Atrophy
 Magnetic resonance imaging

ABSTRACT

NIA-AA diagnostic criteria include volumetric or visual rating measures of hippocampal atrophy (HA) as a diagnostic biomarker of Alzheimer's disease (AD). We aimed to determine its utility as a diagnostic biomarker for early onset Alzheimer's disease (EOAD) by assessing Medial Temporal Atrophy (MTA) and hippocampal volume (HV) determination. MTA score and HV quantified by FreeSurfer were assessed in 140 (aged ≤ 65) subjects with biomarker supported diagnosis: 38 amnesic (A-EOAD), 20 non-amnesic (NA-EOAD), 30 late onset AD (LOAD), 20 fronto-temporal dementia (FTD) and 32 healthy controls (HC). The results showed that the proportion of $MTA \geq 1.5$ was higher on LOAD and FTD than EOAD and HC but none of the MTA thresholds (≥ 1 , ≥ 1.5 and ≥ 2) showed acceptable diagnostic accuracy. LOAD had lower HV than the other groups. A-EOAD HV was lower than NA-EOAD and HC but equal to FTD. The 6258 mm³ cut-off showed good diagnostic accuracy between A-EOAD and HC. Both tools showed a moderate inverse correlation. In conclusion, MTA has a limited diagnostic utility as an EOAD biomarker as it does not discriminate AD from FTD or HC in initial symptomatic stages. HV may discriminate A-EOAD from HC but not from FTD.

1. Introduction

Up to 10% of subjects with neurodegenerative dementias have an early onset disease, defined as a clinical onset below 65 years (Garre-Olmo et al., 2010). The most frequent cause of early-onset dementia is Alzheimer's disease (AD), usually presenting with progressive anterograde episodic memory impairment. However, non-amnesic clinical presentations, are also commonly seen in early onset AD (EOAD) and

frequently overlap with other neurodegenerative dementias such as frontotemporal dementia (FTD) leading to greater rates of misdiagnosis and diagnostic delay (Balasa et al., 2011; Mendez, 2006). In this context the use of specific biomarkers is crucial in achieving an accurate diagnosis.

Hippocampal atrophy (HA) is a well-recognized feature of AD considered a neurodegeneration biomarker in current NIA-AA diagnostic criteria (Apostolova et al., 2006; Sarazin et al., 2010; Albert

Abbreviations: AD, Alzheimer's disease; EOAD, Early Onset Alzheimer's disease; A-EOAD, Amnesic Early Onset Alzheimer's disease; NA-EOAD, Non-Amnesic Early Onset Alzheimer's disease; LOAD, Late-onset Alzheimer's disease; FTD, Fronto-temporal dementia; bvFTD, behavioral variant of Fronto-temporal dementia; nfvPPA, non-fluent variant of Primary Progressive Aphasia; svPPA, semantic variant of Primary Progressive Aphasia; HC, healthy controls; HA, Hippocampal atrophy; HV, Hippocampal volume; MTA, Medial temporal atrophy; MMSE, Mini Mental State Examination; FAQ, Pfeiffer Functional Activities Questionnaire; MCI, mild cognitive impairment.

* Corresponding author at: Alzheimer's disease and other cognitive disorders unit, Neurology Service, Hospital Clínic Barcelona, C/Villarroel, 170, 08036 Barcelona, Spain.

E-mail address: allado@clinic.cat (A. Lladó).

<https://doi.org/10.1016/j.nicl.2019.101927>

Received 13 September 2018; Received in revised form 11 June 2019; Accepted 30 June 2019

Available online 05 July 2019

2213-1582/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

et al., 2011; McKhann et al., 2011). Medial Temporal Atrophy (MTA) visual rating scale is a popular HA evaluation instrument, although its utility in EOAD is still unclear (Scheltens Ph et al., 1992). Since quantitative volumetric analysis is time consuming, semiautomatic methods have been developed. Data regarding its reliability in early symptomatic stages is lacking (Cuingnet et al., 2011).

In this study we sought to elucidate the differences in HA between LOAD, EOAD and FTD, and compare visual and semiautomatic quantitative assessment diagnostic accuracy. We hypothesized that quantitative analysis would be more useful, especially in amnesic EOAD.

2. Methods

2.1. Subjects

The study was approved by the Hospital Clinic Barcelona Ethics Committee and all participants gave written informed consent. One hundred forty subjects evaluated at the Alzheimer's disease and other cognitive disorders Unit at Hospital Clínic de Barcelona were enrolled on this retrospective cross-sectional study. All subjects underwent a complete neurological and neuropsychological evaluation, 3 T brain MRI and a spinal tap for the determination of AD CSF biomarkers. All subjects scored ≥ 20 on the Mini Mental State Examination (MMSE), had a clinical onset before 65. Patients were classified into 3 groups:

- 1) EOAD group ($n = 58$): 28 of them with mild cognitive impairment (MCI) (Pfeiffer Functional Activities Questionnaire (FAQ) ≤ 6) and 30 with mild dementia (Pfeiffer et al., 1982). All subjects had a typical AD CSF biomarker profile. So, all EOAD fulfilling the NIA-AA diagnostic criteria for MCI due to AD or AD dementia (Albert et al., 2011; McKhann et al., 2011). Based on their clinical presentation, EOAD participants were further classified into two subgroups: amnesic (A-EOAD, 38 subjects) and non-amnesic variant (NA-EOAD, 20 subjects).
- 2) LOAD group ($n = 30$): All subjects had a typical AD CSF biomarker profile. Nineteen of them fulfilled the NIA-AA diagnostic criteria for MCI due to AD and eleven for AD dementia (Albert et al., 2011; McKhann et al., 2011).
- 3) FTD group ($n = 20$): Six behavioral variant of FTD (bvFTD), seven non-fluent variant for primary progressive aphasia (nfvPPA) and seven semantic variant of primary progressive aphasia (svPPA) (Rascovsky et al., 2011; Gorno-Tempini et al., 2011). A *C9orf72* expansion was present in two cases of bvFTD and three nfvPPA subjects had *GRN* mutation. All FTD subjects had normal AD CSF biomarkers levels.
- 4) Healthy controls ($n = 32$): with no cognitive complaints, cognitive performance within normative range and normal AD CSF biomarkers. They were classified as young HC (aged under 65) ($n = 16$) and older ones (aged over 65) ($n = 16$)

2.2. Genetic and CSF biomarkers determination

APOE genotype was determined through the analysis of rs429358 and rs7412 by Sanger sequencing.

All subjects underwent a spinal tap. Levels of amyloid β ($A\beta_{42}$), total-tau (t-tau), and phosphorylated-tau (p-tau) were measured using commercial sandwich ELISA kits (Fujirebio, Gent, Belgium). The CSF cut-off values determined by our laboratory were used to classify subjects according to NIA-AA criteria (Balasa et al., 2014).

2.3. Brain MRI imaging

For each subject, high-resolution sagittal T1-weighted MRI images were acquired in a 3Tesla scan (Siemens Magnetom Trio, Erlangen, Germany) at the Magnetic Resonance Image Core Facility, using proprietary three-dimensional magnetization-prepared rapid-acquisition

gradient echo: MPRAGE sequences (TR = 2300 ms; TE = 2,98 ms; acquisition matrix 256×256 , voxel size $1 \times 1 \times 1$).

2.4. Visual assessment

We used T1-coronal images on midbrain to score right and left hippocampus; we scrolled through the whole hippocampus and then selected the main assessment slice in the middle of the hippocampal body, in front of the pons. It was rated according to the five-point scale developed and validated by Scheltens et al. and then averaged to obtain a single MTA value for each subject (Scheltens Ph et al., 1992; Harper et al., 2016). An expert radiologist specialized in dementia neuroimaging (N.B.) and a trained neurologist (N.F.) blinded to diagnosis and clinical data, visually assessed MRI images.

2.5. Quantitative assessment

T1 sequences were analyzed with FreeSurfer v5.3 in neuGRID platform to obtain the HV with the pipeline nG + FreeSurfer + 5.3.0 + Diagnostic + v05.xml. This pipeline relies on Freesurfer-ReconAll 5.3.0 in cross sectional mode (Fischl et al., 2004; Reuter et al., 2012; Cover et al., 2016). Freesurfer segments the hippocampi of 3D T1-weighted structural brain MRI scans. The pipeline involves intensity non-uniformity correction, affine transformation to MNI template, intensity normalization, skull stripping, removal of non-brain tissue, linear and non-linear transformations to a probabilistic brain atlas and labelling of subcortical structures. The right label per each single voxel is determined using spatial localization priors (Fischl et al., 2002). The hippocampal volumetric values in mm^3 are contrasted against a normative population of 238 healthy controls of ADNI. All automated hippocampal segmentations were performed on 64-bit Linux machines using the neuGRID web-portal (www.neugrid4you.eu). NeuGRID allows to efficiently analyze large amount of imaging data with > 5000 CPU cores and 20 TB of physical space (Redolfi et al., 2013; Redolfi et al., 2015). The neuGRID platform provided a final report directly to the physician reporting the volumetric information, the percentile assessments, as well as pictures of the segmented hippocampi allowing a direct evaluation of the FreeSurfer segmentation results (see Supplementary material Fig. S1). Additionally, we analyzed HV using FreeSurfer's standard approach via neuGRID, i.e., native-space volume normalized by total intracranial volume. This represented a validation between atlas space values derived with an affine transformation to the native space hippocampal volume.

2.6. Statistical analysis

Statistical analysis was conducted using Stata/IC 14.2. Categorical data was analyzed by χ^2 test and quantitative data by ANOVA and Student's *t*-test with Bonferroni *post-hoc* procedure. *APOE* $\epsilon 4$ status was dichotomized as carrier/non-carrier and variables were compared in $\epsilon 4$ carrier and non-carrier groups using ANOVA. Inter-rater reliability of left, right and averaged MTA score was determined by Kappa index and intraclass correlation coefficient (ICC) by a two-way random, absolute for single-measures [ICC (2,1)] and average measures ICCs [ICC(2,k)]. The quantitative assessment's accuracy was estimated using receiver operator characteristics (ROC). Best thresholds were selected maximizing sensitivity and specificity. Pearson correlation coefficients evaluated the association between visual and quantitative assessments.

3. Results

3.1. Demographics

LOAD and old HC were older than other groups as expected. Groups showed no differences in gender (Table 1). AD and FTD groups showed lower MMSE scores ($p < .01$) compared to HC. The presence of *APOE*

Table 1
Demographic and clinical data for each diagnostic group.

	A-EOAD (n = 38)	NA-EOAD (n = 20)	FTD (n = 20)	Young HC (n = 16)	LOAD (n = 30)	Older HC (n = 16)
Gender (% Female)	57.9	60	45	62.5	60	46.7
Age Mean	61.3 ± 5	59.7 ± 5.7	61.1 ± 4.4	57.5 ± 3.3	71.3 ± 5.2 ^a	74.7 ± 3.9 ^a
AAO Mean	58.3 ± 4.7	57.2 ± 5.5	58.1 ± 4.4	–	72.2 ± 4.9 ^b	–
Time to diagnosis	3 ± 1.5	2.5 ± 1.3	3 ± 2.2	–	2.4 ± 2.2	–
MMSE score	24.3 ± 2.6	25.1 ± 3	25.8 ± 2.8	29 ± 1.2 ^c	24.2 ± 3	28.4 ± 0.7 ^c
APOE ε4 (%)	65.8 ^d	35	20	18.8	55.2 ^f	6.7
CSF Aβ42 (pg/mL)	397.6 ± 108 ^e	366.28 ± 124 ^e	851.1 ± 305	893.2 ± 247	358.2 ± 105 ^e	733 ± 185
CSF p-tau (pg/mL)	108.6 ± 36 ^e	112.1 ± 44 ^e	48.6 ± 18	51.4 ± 9	95 ± 47 ^e	59.3 ± 17
CSF t-tau (pg/mL)	748.3 ± 426 ^e	701.9 ± 382 ^e	327.6 ± 158	224.3 ± 52	677 ± 439 ^e	299.1 ± 116

Data are presented as means ± standard deviation. A-EOAD, Amnesic Early Onset Alzheimer's disease; NA-EOD, Non-Amnesic Early Onset Alzheimer's disease; FTD, Frontotemporal Dementia; HC, Healthy Controls; AAO, Age At Onset.

^a Statistically significant ($p < .01$) differences compared to A-EOAD, NA-EOAD, FTD, Young HC.

^b Statistically significant ($p < .01$) differences compared to A-EOAD, NA-EOAD, FTD.

^c Statistically significant ($p < .01$) differences compared to A-EOAD, NA-EOAD, FTD and LOAD.

^d Statistically significant ($p < .05$) differences compared to NA-EOAD, FTD, LOAD and young and older HC.

^e Statistically significant ($p < .05$) differences compared to FTD, young and older HC.

^f Statistically significant ($p < .05$) differences compared to older HC.

ε4 allele was highest among A-EOAD (65.8%) than in FTD and HC groups ($p < .05$), including significant differences between A-EOAD and NA-EOAD ($p = .038$).

3.2. Visual assessment

3.2.1. Descriptive data

MTA scores showed substantial agreement between both raters: 88% of total agreement ($k = 0.83$) on the left, 88% ($k = 0.84$) on the right and 83% ($k = 0.78$) on the averaged MTA). Both single and average ICC values were > 0.9 which is considered an excellent correlation between both raters (see Supplementary material Table S1). Most discrepancies were found on 0 and 1 scoring and were solved by consensus.

No significant differences were observed between right and left Scheltens' score, neither between the whole groups FTD (nfvPPA, svPPA and bvFTD) on MTA. The comparison of MTA scores between MCI due to AD and mild AD dementia subjects did not show differences, neither considering APOE ε4 status.

The distribution of the visual MTA assessment is displayed in Figs. 1 and 2. Proportion of participants with MTA score ≥ 1.5 was higher in LOAD (77%, $p < .01$) and FTD (70%, $p < .01$) than A-EOAD (34.3%), NA-EOAD (30%) and HC groups (6,3%). FTD was the only group with MTA ≥ 3.5 (15%). A-EOAD showed differences compared to HC ($p = .032$) but not to NA-EOAD or FTD.

3.2.2. Diagnostic performance

We evaluated diagnostic sensitivity and specificity for different thresholds of the MTA score among AD and FTD groups (Table 2a). MTA performed better at discriminating LOAD and FTD from HC than EOAD from HC. However, no threshold reached 80% performance on both sensitivity and specificity. In addition, MTA did not reach 60% diagnostic accuracy at discriminating between AD groups from FTD (Table 2b).

3.3. Quantitative assessment

3.3.1. Descriptive data

There were no differences in right and left HV. Thus, averaged HV were used for subsequent analyses. No significant differences were found on mean HV values among FTD clinical variants (nfvPPA, svPPA and bvFTD) neither between MCI due to AD and mild dementia due to AD. HV considering APOE ε4 status did not show differences between groups.

The distribution of the HV by different diagnostic groups is shown in

Fig. 2. We observed significant differences on the average HV between groups ($p < .001$). The post hoc analysis showed that HV LOAD ($4890 \pm 866 \text{ mm}^3$) was smaller than the other groups ($p < .05$). A-EOAD HV ($5568 \pm 663 \text{ mm}^3$) was smaller than NA-EOAD ($6260 \pm 872 \text{ mm}^3$) ($p = .019$) and HC ($6766 \pm 711 \text{ mm}^3$) ($p < .001$), but we did not find differences between A-EOAD and FTD groups ($5671.1 \text{ mm}^3 \pm 1109 \text{ mm}^3$). No differences between NA-EOAD and HC or FTD were found. FTD had smaller HV than HC ($p = .001$). Furthermore, the distribution and means of the HV by different diagnostic groups using total intracranial volume normalization is shown as Supplementary material Table S2 and Fig. S2.

3.3.2. Diagnostic performance

ROC curves were used to determine HV threshold with a better diagnostic performance. The 5838 mm^3 threshold distinguished between LOAD and HC with very good sensitivity (100%) and specificity (87%). HV threshold of 6259 mm^3 showed good sensitivity (88%) and specificity (87%) to distinguish A-EOAD from HC (Fig. 3.1a). For identifying NA-EOAD versus HC, the best threshold was 6745 mm^3 , although it showed suboptimal sensitivity (69%) and specificity (70%) (Fig. 3.1b). The best HV cut-off for FTD versus HC was 6365.5 mm^3 (sensitivity 82% and specificity 75%) (Fig. 3.1c). We also compared AD groups to FTD (Fig. 3.2a and 3.2b). The best cut-off for comparing A-EOAD and FTD was 5763.5 mm^3 and for comparing NA-EOAD versus FTD was 5845 mm^3 , although both showed sensitivity and specificity below 70%. The 5124 mm^3 threshold distinguished LOAD from FTD with 75% sensitivity and 63% specificity. Moreover, diagnostic performance of HV using total intracranial volume normalization is shown in Supplementary Material Fig. S3.

3.4. Correlation between quantitative and visual assessment

There was a significant moderate inverse correlation ($r = -0.537$, $p < .001$) between MTA score and the quantitative HV evaluation.

4. Discussion

In this study we compared a visual scale and a semiautomatic tool for measuring HA in patients with EOAD, LOAD and FTD in order to determine the diagnostic accuracy of both tools in early symptomatic stages. In our cohort, MTA scale had little utility as a diagnostic biomarker and the quantitative measurement of HV was useful only in distinguishing LOAD and A-EOAD from HC.

To our knowledge this is the first study to compare both visual and volumetric assessment in a well-characterized biomarker-confirmed

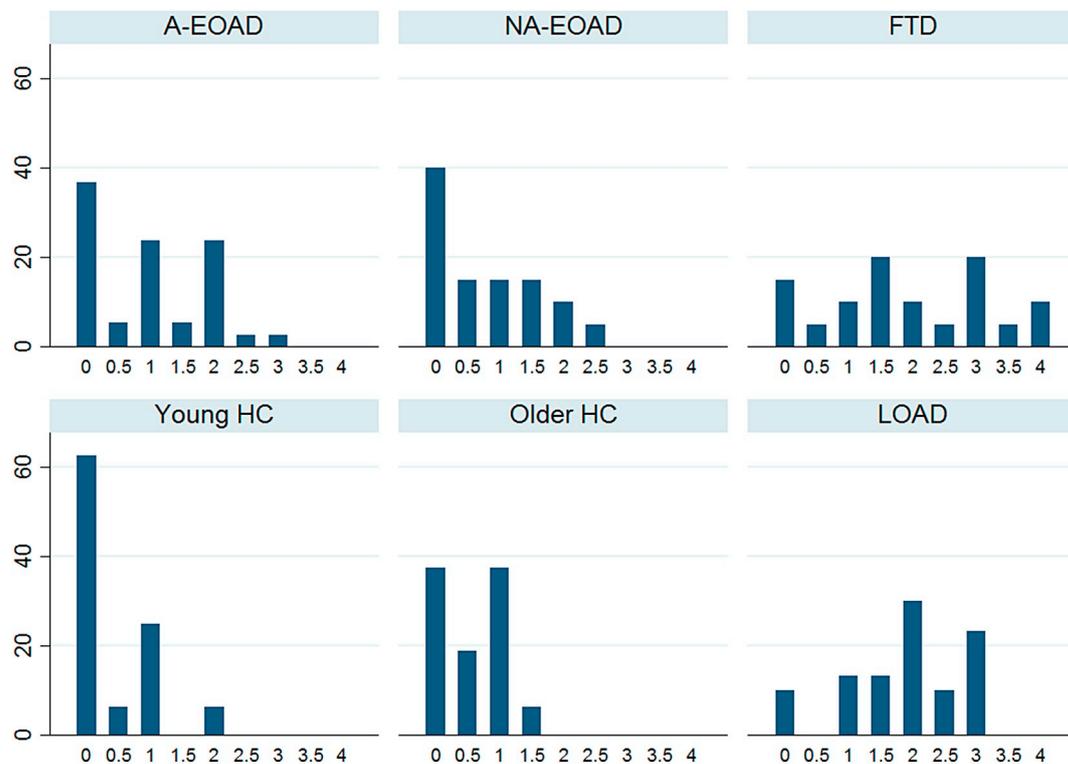


Fig. 1. Distribution of MTA scoring for diagnostic groups and HC (percentage, %).

cohort of LOAD, EOAD and FTD patients.

Since its development, Scheltens' HA visual rating scale has been validated in numerous studies as a good predictor of progression from MCI to AD dementia and to discriminate AD from HC and has also been proved to correlate with volumetric methods (Heo et al., 2013; Wahlund et al., 2000). The MTA score has been proposed as the best marker of HA and the $MTA \geq 1.5$ cut-off has been recommended to be used for AD diagnosis under the age of 75 (Van de Pol and Scheltens, 2014). In the case of volumetric assessment, it is not clear yet which HV cut off should be used.

In our study, MTA was not able to accurately discriminate HC from AD or FTD patients, since specificity and sensitivity did not reach the 80% threshold for any of the groups (Consensus report of the Working Group on: "Molecular and Biochemical Markers of Alzheimer's Disease", 1998). The quantitative assessment only achieved a good diagnostic performance in discriminating LOAD and A-EOAD from HC with little discriminative capacity between the other comparisons.

Most previous studies point to the fact that both visual scales and quantitative analysis are able to discriminate between AD and HC, despite relevant variability in terms of diagnostic performance and they mostly focus on LOAD patients (Heo et al., 2013; Cavado et al., 2014). In reference to LOAD, our results are in line with previous publications in terms of sensitivity albeit greater specificity in our cohort. (Harper et al., 2016) This could be explained by the characteristics of our older HC group in whom preclinical AD had been carefully ruled out through the use of CSF biomarkers and an extensive neuropsychological evaluation. In both visual scales and quantitative analysis, the diagnostic accuracy is better in LOAD than EOAD, as well as it is better in A-EOAD than in NA-EOAD. It makes sense, since LOAD patients mostly present amnesic clinical phenotype and have more HA whereas EOAD patients frequently have a non-amnesic clinical profile related to less HA (Balasa et al., 2011; Koedam et al., 2010; Phillips et al., 2018). Poor MTA diagnostic accuracy results on EOAD may be due to the fact we have only included patients in early symptomatic stages (i.e, MCI and mild dementia, $MMSE \geq 20$). Is well known that patients in advanced stages of the disease have more HA, thus, in consequence, assessing

MTA at these stages could increase its discriminative power; but on the daily clinical practice it is in the early stages of the disease when the differential diagnosis is more complex and it is precisely where biomarkers should be more useful and make the difference. On the other hand, most studies using Scheltens' MTA scale do not have CSF results and the AD diagnosis is based only on clinical criteria (Cuingnet et al., 2011; Heo et al., 2013; Wahlund et al., 2000; Van de Pol and Scheltens, 2014; Duara et al., 2013; Ferreira et al., 2015; Varon et al., 2015; Pereira et al., 2014; Shen et al., 2011; Ridha et al., 2007). We consider this fact to be a significant caveat since trying to establish AD diagnosis without biological confirmation of the disease can lead to higher rates of misdiagnosis (Beach et al., 2012). Claus et al., 2017 reported good diagnostic accuracy (83.3% sensitivity and 86.4% specificity) for $MTA \geq 1$ at discriminating 18 EOAD patients from subjective cognitive impairment subjects (Claus et al., 2017). Their data does not fit with our results, which could be explained by several methodological differences such as: the use of computerized tomography (CT) scan, the lack of biomarker supported diagnosis, the wide range of clinical severity stages included in the sample and its small size.

Volumetric assessment is expected to provide an added value in AD diagnosis (Bosco et al., 2017). Several techniques have been developed to achieve the most accurate measurement and many comparisons of both tools have been published (Cuingnet et al., 2011; Heo et al., 2013; Wahlund et al., 2000; Van de Pol and Scheltens, 2014; Duara et al., 2013; Ferreira et al., 2015; Varon et al., 2015; Pereira et al., 2014; Shen et al., 2011). However, comparing them is difficult because of the heterogeneity of study samples and imaging techniques used. Cuingnet et al. compared the diagnostic performance of different quantitative methods, including volumetric measurement of HV of clinically diagnosed MCI and mild AD patients aged 55–90 (Cuingnet et al., 2011). HV evaluation was as sensitive as other methods but however less specific: 63% sensitivity and 80% specificity on distinguishing HC from AD patients and 73% sensitivity and 74% specificity between HC and MCI. By contrast, our data show acceptable diagnostic performance of the volumetric analysis to distinguish A-EOAD from HC. That difference can be explained also by methodological differences such as no

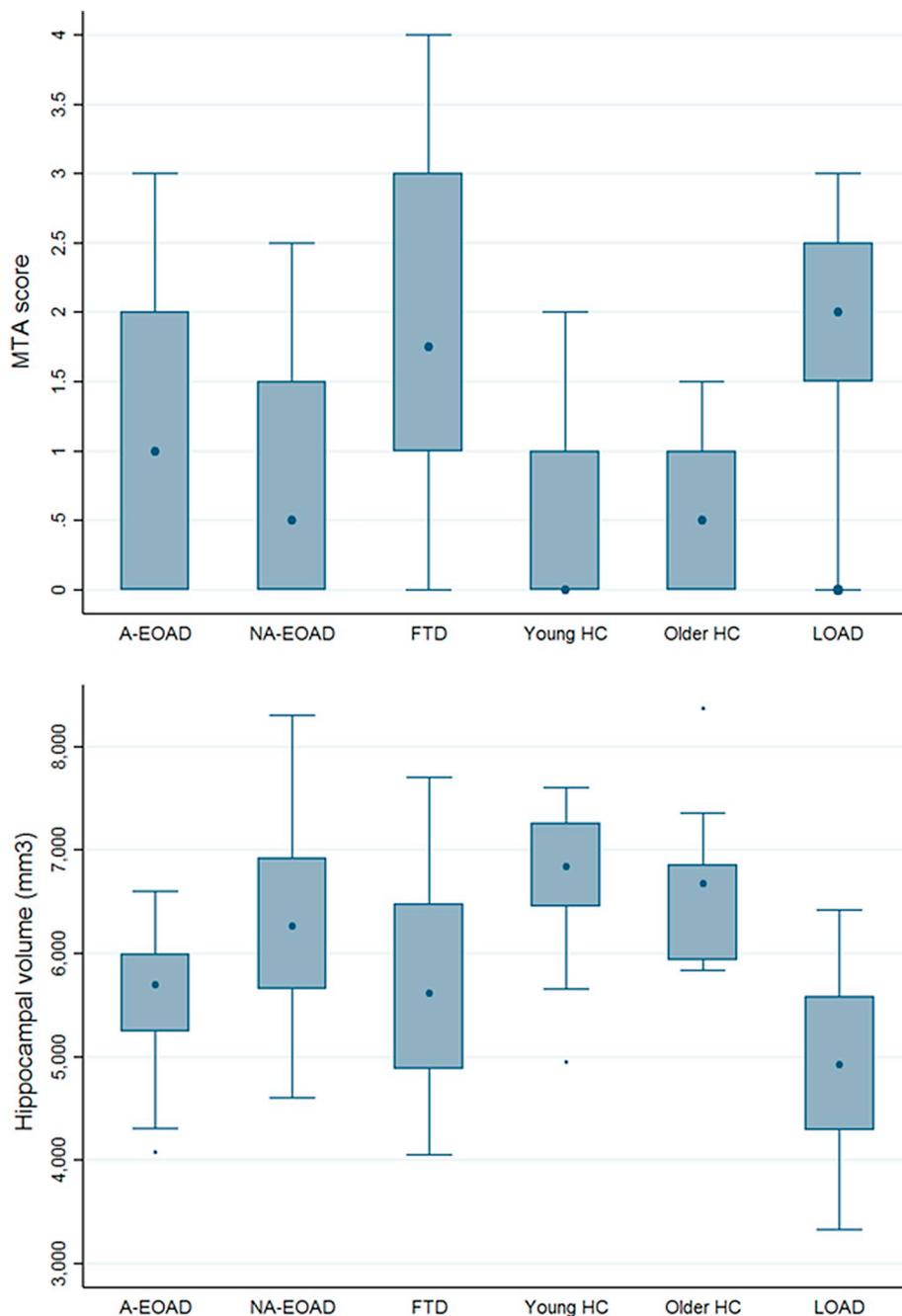


Fig. 2. Box plots of the MTA and HV distribution depending on the diagnosis.

biomarker supported diagnosis.

When comparing both LOAD and EOAD with FTD patients, visual and quantitative HA assessment does not appear to be a good diagnostic biomarker. This data fits well with previous works that show similar HA on both disorders (Van de Pol et al., 2006; Hornberger et al., 2012). The relevant MTA and its wide distribution found on FTD could be related to the heterogeneity and the fast disease progression of the FTD itself.

The clinical overlap between NA-EOAD and FTD sometimes leads to differential diagnosis difficulties in clinical practice and highlights the crucial importance of using disease-specific biomarkers (Beach et al., 2012). Unfortunately, in our cohort, neither visual nor quantitative

assessment performed optimally at differentiating them.

In line with previous studies mostly performed in late-onset cognitive impairment, we observed a moderate correlation between MTA visual rating and volumetric techniques (Dhikav et al., 2017). The moderate correlation could be explained because MTA score is based mostly on a single slice while HV is a three-dimensional measure. These data suggest that both techniques may perform in a different way with regard to HV measurements, therefore caution must be taken before considering them equivalent.

Furthermore, variability through different HV quantification methods has been previously described (Buckner et al., 2004). Both

Table 2
Diagnostic performance for each diagnostic group.

a) Diagnostic accuracy for MTA visual rating of diagnostic groups vs HC.				
	A-EOAD (n = 38)	NA-EOAD (n = 20)	LOAD (n = 30)	FTD (n = 20)
MTA ≥ 1	Se 58% Sp 69%	Se 45% Sp 69%	Se 90% Sp 56%	Se 80% Sp 69%
MTA ≥ 1,5	Se 34% Sp 93%	Se 30% Sp 93%	Se 77% Sp 94%	Se 70% Sp 94%
MTA ≥ 2	Se 29% Sp 94%	Se 15% Sp 94%	Se 63% Sp 100%	Se 50% Sp 94%
AUC	0.67 (0.54–0.81)	0.63 (0.46–0.80)	0.88 (0.79–0.98)	0.85 (0.72–0.97)

b) Diagnostic accuracy for MTA visual rating of AD group vs FTD group.			
	A-EOAD (n = 38)	NA-EOAD (n = 20)	LOAD (n = 30)
MTA ≥ 1	Se 58% Sp 20%	Se 45% Sp 20%	Se 80% Sp 10%
MTA ≥ 1,5	Se 34% Sp 30%	Se 30% Sp 30%	Se 70% Sp 23%
MTA ≥ 2	Se 29% Sp 50%	Se 15% Sp 50%	Se 50% Sp 37%
AUC	0.28 (0.14–0.43)	0.24 (0.09–0.31)	0.53 (0.35–0.68)

A-EOAD, Amnesic Early Onset Alzheimer's Disease; NA-EOAD, Non-Amnesic Early Onset Alzheimer's Disease; FTD, Frontotemporal Dementia; HC = Healthy Controls; MTA, Medial temporal Atrophy score; Se, Sensitivity; Sp, Specificity.

approaches implemented in neuGRID, i.e.: HV in template space and total intracranial volume normalization methods, provided similar results. It suggests that HV diagnostic performance is consistent across the two analysis methods. Moreover, in the normalization method, the HV is not biased by extensive cortical atrophy.

Our results would support the idea that visual MTA rating may play a limited role in the clinical diagnosis of EOAD, in particular in early clinical stages. With regards to volumetric assessment, it would only show an advantage compared to visual rating in discrimination between LOAD and A-EOAD from HC. Current criteria for the clinical diagnosis of AD includes HA as a neurodegeneration biomarker (Albert et al., 2011; McKhann et al., 2011). In light of our results showing low discriminative capacity of HA between groups, this feature should be interpreted with caution in early-onset cognitive impairment especially in patients with non-amnesic presentations.

The main limitation of our study is the relatively small sample size, although, as far as we are aware this is the larger cohort study focused

on EOAD patients reported until now. On the other hand, we have only used one semiautomatic volumetric analysis and it is possible that other volumetric techniques for evaluating HA may show different diagnostic performances.

5. Conclusion

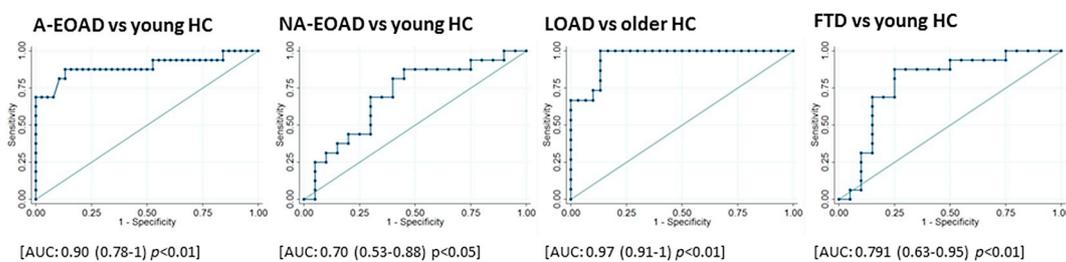
In conclusion, the utility of HA as a diagnostic biomarker in our cohort of EOAD patients, is limited. Further studies in larger and well-characterized cohorts are needed to determine the diagnostic utility of HA as a biomarker in the early stages for EOAD patients.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2019.101927>.

Disclosure statement

Authors state that there are no conflicts of interest to disclose.

3a) ROC curves of HV of diagnostic groups compared to HC



3b) ROC curves of HV of AD groups compared to FTD

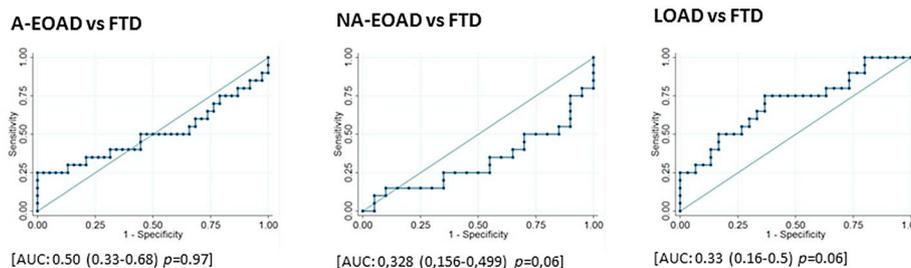


Fig. 3. Diagnostic performance of HV quantitative assessment.

Acknowledgements

The authors thank patients, their relatives and healthy controls for their participation in the research.

Funding

This work was supported by Spanish Ministry of Economy and Competitiveness-Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea, “Una manera de hacer Europa” [PI14/00282 to Dr. A. Lladó], PERIS 2016–2020 Departament de Salut de la Generalitat de Catalunya [SLT002/16/00408 to Dr. Sanchez-Valle] and Fundació Marató de TV3, Barcelona, Spain [Grant 20143810 to Dr. Sanchez-Valle] and CERCA Programme/Generalitat de Catalunya. Dr. Neus Falgàs received funding from Hospital Clinic Barcelona [Ajut Josep Font]. Dr. Anna Antonell received funding from Departament de Salut de la Generalitat de Catalunya [PERIS 2016–2020 SLT002/16/00329].

References

- Albert, M.S., DeKosky, S.T., Dickson, D., et al., 2011. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 270–279.
- Apostolova, L.G., Dinov, I.D., Dutton, R.A., et al., 2006. 3D comparison of hippocampal atrophy in amnesic mild cognitive impairment and Alzheimer's disease. *Brain* 129, 2867–2873 (Pt 11). Nov.
- Balasa, M., Gelpi, E., Antonell, A., et al., 2011. Neurological tissue Bank/University of Barcelona/Hospital Clínic NTB/UB/HC collaborative group. Clinical features and APOE genotype of pathologically proven early-onset Alzheimer disease. *Neurol.* 76 (20), 1720–1725 May 17.
- Balasa, M., Sanchez-Valle, R., Antonell, A., et al., 2014. Usefulness of biomarkers in the diagnosis and prognosis of early-onset cognitive impairment. *J. of Alzheimer's Dis: JAD* 40, 919–927.
- Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W., 2012. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* 71 (4), 266–273 Apr.
- Bosco, P., Redolfi, A., Bocchetta, M., et al., 2017. The impact of automated hippocampal volumetry on diagnostic confidence in patients with suspected Alzheimer's disease: a European Alzheimer's disease consortium study. *Alzheimers Dement.* 13 (9), 1013–1023 Sep.
- Buckner, R.L., Head, D., Parker, J., et al., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage.* 23 (2), 724–738 Oct.
- Cavedo, E., Pievani, M., Boccardi, M., et al., 2014. Medial temporal atrophy in early and late-onset Alzheimer's disease. *Neurobiol. Aging* 35 (9), 2004–2012 Sep.
- Claus, J.J., Staekenborg, S.S., Holl, D.C., et al., 2017. Practical use of visual medial temporal lobe atrophy cut-off scores in Alzheimer's disease: validation in a large memory clinic population. *Eur. Radiol.* 27 (8), 3147–3155 Aug.
- Consensus report of the Working Group on: "Molecular and Biochemical Markers of Alzheimer's Disease", 1998. The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. *Neurobiol. Aging* 19 (2), 109–116 Mar-Apr. (Review. Erratum in: *Neurobiol Aging* 1998 May-Jun;19(3):285).
- Cover, K.S., van Schijndel, R.A., Versteeg, A., et al., 2016. Alzheimer's disease neuroimaging initiative, neuGRID. Reproducibility of hippocampal atrophy rates measured with manual, FreeSurfer, AdaBoost, FSL/FIRST and the MAPS-HBSI methods in Alzheimer's disease. *Psychiatry Res.* 252, 26–35 Jun 30.
- Cuingnet, R., Gerardin, E., Tessieras, J., et al., 2011. Alzheimer's disease neuroimaging initiative. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 56 (2), 766–781 May 15.
- Dhikav, V., Duraiswamy, S., Anand, K.S., 2017. Correlation between hippocampal volumes and medial temporal lobe atrophy in patients with Alzheimer's disease. *Ann. Indian Acad. Neurol.* 20 (1), 29–35 Jan-Mar.
- Duara, R., Loewenstein, D.A., Shen, Q., et al., 2013. 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.* 18 (5), 47 Sep.
- Ferreira, D., Cavallin, L., Larsson, E.M., et al., 2015. AddNeuroMed consortium and the Alzheimer's disease neuroimaging initiative. 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.* 278 (3), 277–290 Sep.
- Fischl, B., Salat, D.H., Busa, E., et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33 (3), 341–355 Jan 31.
- Fischl, B., Salat, D.H., van der Kouwe, A.J., et al., 2004. Sequence-independent segmentation of magnetic resonance images. *Neuroimage* 23 (Suppl. 1), S69–S84.
- Garre-Olmo, J., Genís Batlle, D., del Mar Fernández, M., et al., 2010. Registry of dementia of Girona study group (ReDeGi study group). Incidence and subtypes of early-onset dementia in a geographically defined general population. *Neurol.* 75 (14), 1249–1255 Oct 5.
- Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., et al., 2011. Classification of primary progressive aphasia and its variants. *Neurol.* 76, 1006–1014.
- Harper, L., Fumagalli, G.G., Barkhof, F., et al., 2016. MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain* 139, 1211–1225 (Pt 4). Apr.
- Heo, J.H., Kim, M.K., Lee, J.H., Lee, J.H., 2013. Usefulness of medial temporal lobe atrophy visual rating scale in detecting Alzheimer's disease: preliminary study. *Ann. Indian Acad. Neurol.* 16 (3), 384–387 Jul.
- Hornberger, M., Wong, S., Tan, R., et al., 2012. In vivo and post-mortem memory circuit integrity in frontotemporal dementia and Alzheimer's disease. *Brain* 135, 3015–3025 Pt 10. Oct.
- Koedam, E.L., Lauffer, V., van der Vlies, A.E., van der Flier, W.M., Scheltens, P., Pijnenburg, Y.A., 2010. Early-versus late-onset Alzheimer's disease: more than age alone. *J. Alzheimers Dis.* 19 (4), 1401–1408.
- McKhann, G.M., Knopman, D.S., Chertkow, H., et al., 2011. 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.* 7, 263–269.
- Mendez, M.F., 2006. The accurate diagnosis of early-onset dementia. *Int. J. Psychiatry Med.* 36 (4), 401–412 (Review).
- Pereira, J.B., Cavallin, L., Spulber, G., et al., 2014. AddNeuroMed consortium and for the Alzheimer's disease neuroimaging initiative. Influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cut-offs. *J. Intern. Med.* 275 (3), 317–330 Mar.
- Pfeffer, R.L., Kurosaki, T.T., Harrah, C.H., Chance, J.M., Filos, S., 1982. Measurement of functional activities in older adults in the community. *J. Gerontol.* 37, 323–329.
- Phillips, J.S., Da Re, F., Dratch, L., et al., 2018. Neocortical origin and progression of gray matter atrophy in nonamnesic Alzheimer's disease. *Neurobiol. Aging* 63, 75–87 Mar.
- Rascovsky, K., Hodges, J.R., Knopman, D., et al., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 134, 2456–2477.
- Redolfi, A., Bosco, P., Manset, D., Frisoni, G.B., neuGRID consortium, 2013. Brain investigation and brain conceptualization. *Funct. Neurol.* 28 (3), 175–190 Jul-Sep.
- Redolfi, A., Manset, D., Barkhof, F., et al., 2015. neuGRID consortium, for the Alzheimer's disease neuroimaging initiative. Head-to-head comparison of two popular cortical thickness extraction algorithms: a cross-sectional and longitudinal study. *PLoS One* 10 (3) Mar 17.
- Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61 (4), 1402–1418 Jul 16.
- Ridha, B.H., Barnes, J., van de Pol, L.A., et al., 2007. Application of automated medial temporal lobe atrophy scale to Alzheimer disease. *Arch. Neurol.* 64 (6), 849–854 Jun.
- Sarazin, M., Chauviré, V., Gerardin, E., et al., 2010. The amnesic syndrome of hippocampal type in Alzheimer's disease: an MRI study. *J. Alzheimers Dis.* 22 (1), 285–294.
- Scheltens Ph, Leys D., Barkhof, F., et al., 1992. Atrophy of medial temporal lobes on MRI in “probable” Alzheimer's disease and normal aging: diagnostic value and neuropsychological correlates. *J. Neurol. Neurosurg. Psychiatry* 55, 967–972.
- Shen, Q., Loewenstein, D.A., Potter, E., et al., 2011. Volumetric and visual rating of magnetic resonance imaging scans in the diagnosis of amnesic mild cognitive impairment and Alzheimer's disease. *Alzheimers Dement.* 7 (4), e101–e108 Jul.
- Van de Pol, L.A., Scheltens, P., 2014. 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.* 275 (3), 331–333 Mar.
- Van de Pol, L.A., Hensel, A., van der Flier, W.M., et al., 2006. Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 77 (4), 439–442 Apr.
- Varon, D., Barker, W., Loewenstein, D., et al., 2015. Alzheimer's disease neuroimaging initiative. Visual rating and volumetric measurement of medial temporal atrophy in the Alzheimer's disease neuroimaging initiative (ADNI) cohort: baseline diagnosis and the prediction of MCI outcome. *Int J Geriatr Psychiatry.* 30 (2), 192–200 Feb.
- Wahlund, L.O., Julin, P., Johansson, S.E., Scheltens, P., 2000. Visual rating and volumetry of the medial temporal lobe on magnetic resonance imaging in dementia: a comparative study. *J. Neurol. Neurosurg. Psychiatry* 69 (5), 630–635 Nov.