

RESEARCH

Open Access



# Validity and reliability of the medial temporal lobe atrophy scale in a memory clinic population

Anna Molinder<sup>1,2\*</sup>, Doerthe Ziegelitz<sup>1</sup>, Stephan E. Maier<sup>1,3</sup> and Carl Eckerström<sup>4,5</sup>

## Abstract

**Background:** Visual rating of medial temporal lobe atrophy (MTA) is often performed in conjunction with dementia workup. Most prior studies involved patients with known or probable Alzheimer's disease (AD). This study investigated the validity and reliability of MTA in a memory clinic population.

**Methods:** MTA was rated in 752 MRI examinations, of which 105 were performed in cognitively healthy participants (CH), 184 in participants with subjective cognitive impairment, 249 in subjects with mild cognitive impairment, and 214 in patients with dementia, including AD, subcortical vascular dementia and mixed dementia. Hippocampal volumes, measured manually or using FreeSurfer, were available in the majority of cases. Intra- and interrater reliability was tested using Cohen's weighted kappa. Correlation between MTA and quantitative hippocampal measurements was ascertained with Spearman's rank correlation coefficient. Moreover, diagnostic ability of MTA was assessed with receiver operating characteristic (ROC) analysis and suitable, age-dependent MTA thresholds were determined.

**Results:** Rater agreement was moderate to substantial. MTA correlation with quantitative volumetric methods ranged from -0.20 ( $p < 0.05$ ) to -0.68 ( $p < 0.001$ ) depending on the quantitative method used. Both MTA and FreeSurfer are able to distinguish dementia subgroups from CH. Suggested age-dependent MTA thresholds are 1 for the age group below 75 years and 1.5 for the age group 75 years and older.

**Conclusions:** MTA can be considered a valid marker of medial temporal lobe atrophy and may thus be valuable in the assessment of patients with cognitive impairment, even in a heterogeneous patient population.

**Keywords:** Dementia, Medial temporal lobe atrophy (MTA), Alzheimer's disease, Mild cognitive impairment, Atrophy, Magnetic resonance imaging

## Background

The medial temporal lobe (MTL) is an early affected site for Alzheimer's disease (AD) related neurodegeneration [1]. Regional atrophy of the MTL structures detected with magnetic resonance imaging (MRI) is a recognized AD biomarker [2, 3]. However, MTL atrophy may be present in other types of dementia, e.g. in subcortical

vascular dementia (SVD) [4–8] and is independently associated with cognitive impairment in patients with cerebral vascular pathology [6, 9, 10]. MTL atrophy may also be present in patients with mild cognitive impairment (MCI) [11]. In this patient group and even in healthy individuals, MTL atrophy or increased atrophy rate indicates risk of future cognitive decline [12–15].

Assessment of MTL atrophy on MRI is often part of the standard evaluation of patients with cognitive decline. There are several (semi-)automated segmentation tools available for quantifying MTL volumes, but

\*Correspondence: anna.molinder@vgregion.se

<sup>2</sup> Neuroradiology, Sahlgrenska sjukhuset, Blå stråket 5, Gothenburg 413 46, Sweden

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the availability and usage of such tools vary across radiological departments. Furthermore, absolute hippocampal volumes will be biased by the quantitative measuring method used, since manual volumetry and the various automated software programs tend to delineate the anatomical structures differently [16]. In terms of easy clinical applicability, visual assessment of MTL atrophy is still superior to volumetric measuring methods. For visual assessment the medial temporal lobe atrophy scale (MTA) introduced by Scheltens et al. is widely used [17, 18]. In the original article, the MTA scale was able to differentiate between AD patients and controls, a finding that has been replicated in later studies [19–21]. Depending on methods used, comparisons between MTA and manual volumetry or automated methods have shown acceptable to good correlations [22–26]. Studies of MTA with regard to reliability, validity and diagnostic ability, however, have mostly focused on AD and its prodromal phases, fewer on SVD or mixed dementia.

The patient population admitted at memory clinics is characterized by rather diverse cognitive symptoms and underlying disorders, sometimes with mixed neurodegenerative and vascular pathology. Such a mixed clinical patient population, ranging from subjective cognitive impairment (SCI) and MCI to dementia including AD, SVD and mixed dementia, is the subject of the present report.

The overall aim of the study was to investigate the reliability and validity of the MTA scale, with regard to both quantitative hippocampal volumes and to clinical diagnoses, using a well-defined memory clinic patient cohort with different underlying disorders and different stages of cognitive impairment.

## Methods

### Study participants

#### *The Gothenburg MCI study*

The present study is part of the Gothenburg MCI study [27], a clinical longitudinal study focused on neurodegenerative, vascular and stress disorders prior to the development of dementia. The Gothenburg MCI study was approved by the local ethics committee (approval number: L091-99, 1999; T479-11, 2011), and is conducted in accordance with the Declaration of Helsinki of 1975 and 1983. Written informed consent is obtained from all participants in the Gothenburg MCI study.

The study participants for the Gothenburg MCI study were recruited at the Memory Clinic, where they were examined due to subjective or objective cognitive complaints. Inclusion criteria for the Gothenburg MCI study were: age between 50 to 79 years; mini mental state examination (MMSE) score > 18; duration of cognitive decline for 6 or more months. Exclusion criteria

consisted of somatic diseases that may cause cognitive impairment, e.g., brain tumors, subdural hemorrhage, encephalitis, unstable heart disease or hypothyroidism as well as severe psychiatric disorders, substance abuse or confusion caused by drugs. Controls were primarily recruited through senior citizen organizations. In a few cases, the controls were spouses to patients at the memory clinic. Additionally, twenty-three patients were reclassified as healthy controls when they upon examination had neither objective nor subjective signs of cognitive impairment. Inclusion and exclusion criteria were the same as for the patients with the exception that controls were not included if they had subjective or objective signs of cognitive disorders.

### Present study

Participants from the Gothenburg MCI study were included in the present study if they had undergone at least one MRI exam during the observation period, with a technically successful T<sub>1</sub>-weighted volume scan suitable for medial temporal lobe atrophy (MTA) evaluation. Between 1999 and 2014, 458 patients and 73 controls underwent both MRI and clinical examination, including a global deterioration scale (GDS) classification, as part of the Gothenburg MCI study. A total of 756 MRI scans were performed, i.e., some of the enrolled subjects underwent more than one MRI examination. Four of these scans, obtained in four patients who underwent only a single MRI examination, had to be excluded because of distortion artifacts or inadequate volume coverage for MTA assessment. Participants entering the study as patients ( $N=454$ ) received 655 MRI examinations and controls ( $N=73$ ) 97 MRI exams. Out of the total of 752 MRI examinations included in the study, 136 were performed with a 0.5 Tesla scanner and 616 were performed with a 1.5 Tesla scanner. Each MRI exam, whether performed at baseline or at follow up, was accompanied by a new clinical assessment including GDS classification. Follow up time ranged from 1 to 9 years.

For the purpose of this study, all included MRI exams were grouped according to the subject's GDS classification at the time of each MRI scan, regardless as to whether the participant had entered as patient of the Gothenburg MCI study or as presumed healthy control. The cognitively healthy cohort (CH) comprises 105 examinations.

### Clinical evaluation

At each clinic visit, participants were classified according to the GDS, based on anamnestic data and assessment of cognitive symptoms using the following clinical checklists: Stepwise Comparative Status Analysis (STEP); I-Flex, short form of the Executive Interview (EXIT);

Mini mental state examination (MMSE); and Clinical Dementia Rating (CDR). GDS 1 stands for cognitively intact, GDS 2 for SCI, GDS 3 for MCI and GDS 4 for mild dementia [28]. The CDR sum of boxes assessment was based on information from both the patient and an informant. The guidelines for the classification were as follows: For GDS 2 (SCI) participants should have MMSE  $\geq 28$ , CDR  $\leq 0.5$ , I-FLEX  $< 3$ , and no positive outcomes on variables 13-20 of STEP; GDS 3 (MCI) corresponds to MMSE  $\geq 26$ , CDR  $> 0.5$ , I-FLEX  $\leq 3$ , and one or fewer positive outcomes on variables 13-20 of STEP; and for GDS 4 (mild dementia) participants should have MMSE  $\leq 25$ , CDR  $> 1.0$ , STEP  $> 1$ , and I-FLEX  $> 3$ . When the guidelines were not applicable, a consensus decision among the physicians at the clinic was made to determine the appropriate GDS score.

The detailed diagnostic procedures and further details concerning the Gothenburg MCI study design have been presented in an earlier publication [27].

Study participants with GDS 4 (dementia) were further classified according to specific diagnoses, with AD (98 MRI exams) according to the NINCDS-ADRDA criteria [29], subcortical vascular dementia (25 MRI exams) according to the Erkinjuntti criteria [30] or mixed Alzheimer/vascular dementia (51 MRI exams). For mixed dementia, AD criteria had to be fulfilled as well as moderate/severe white matter changes (WMC) (Fazekas score  $\geq 2$ ) on MRI, or mild WMC in combination with a marked fronto-subcortical-dysexecutive syndrome. The clinician who set the dementia diagnoses had access to MRI images but was blinded to volumetric and visual rating data, as well as neuropsychological test results and cerebrospinal fluid (CSF) biomarker data.

Furthermore, a diagnostically heterogeneous group with GDS 4 was summarized as "Other dementias" and includes: Twenty-one examinations that were performed in participants with dementia *non ultra descripta*, ten with dementia of uncertain etiology, four with fronto-temporal dementia according to Neary et al. [31], two with mixed fronto-temporal dementia and vascular dementia, two with primary progressive aphasia according to Gorno-Tempini et al. [32] and one with Lewy body dementia according to McKeith et al. [33]. These dementia subgroups are not included in analyses concerning classification accuracy, due to their small group sizes. Average demographical and clinical data of respective groups at the time of MRI examination are presented in Table 1.

#### Image acquisition

The MRI protocol performed as part of the Gothenburg MCI study included a T<sub>1</sub>-weighted MPRAGE 3D volume scan used for MTA scoring and volumetric

**Table 1** Number of MRI studies and associated participant characteristics

	N	Age, years	Female (%)	Education, years	MMSE
CH	105	64.7 (6.1)	64	12.7 (3.2)	29.5 (0.6)
SCI	184	65.7 (7.1)	62	13.0 (3.7)	29.1 (1.1)†
MCI	249	67.0 (7.1)*	56	12.2 (3.7)	28.0 (1.6)†
Dementia	214	68.9 (6.8)†	55	11.3 (3.5)†	23.4 (4.5) †
AD	98	68.0 (6.6)†	62	11.1 (3.4)*	22.3 (5.2) †
SVD	25	69.6 (6.4)†	24†	11.2 (3.3)*	26.0 (1.9)†
Mixed	51	72.2 (5.8)†	69	11.2 (3.7)*	22.7 (4.1)†
Other	40	66.7 (7.3)	40*	11.6 (3.7)	25.0 (3.0)†

Data are given as mean (standard deviation (SD))

Data concerning education was missing in 32 cases, evenly spread in the SCI, MCI and dementia group. MMSE data was missing in 14 cases, all in the dementia subgroups except for the SVD group

CH Cognitively healthy, SCI Subjective cognitive impairment, MCI Mild cognitive impairment, AD Alzheimer's disease, SVD Subcortical vascular dementia, Mixed Mixed dementia, Other Other dementia, MMSE Mini mental state examination

\*Significant results compared to CH,  $p < 0.05$ . † Significant results compared to CH,  $p < 0.001$

measurements. Between years 1999 and 2004, MRIs were performed on a 0.5 Tesla MR scanner (Philips NT5, Eindhoven, The Netherlands). The following scan parameters were used: repetition time (TR) 30 ms; echo time (TE) 10 ms; slice thickness 1.5 mm; slice gap 0 mm; flip angle 40°; field-of-view (FOV) 220 × 220 mm<sup>2</sup>; acquisition pixel size 0.86 × 1.12 mm<sup>2</sup>; and reconstruction pixel size 0.86 × 0.86 mm<sup>2</sup>. Between years 2005 and 2014 participants were examined on a 1.5 Tesla MR scanner (Siemens Symphony, Siemens Medical Systems, Erlangen, Germany) (TR 1610 ms; TE 2.38 ms; slice thickness 1 mm; slice gap 1 mm; flip angle 15°; FOV 250 × 203 mm<sup>2</sup>; acquisition pixel size 1.0 × 1.0 mm<sup>2</sup>; reconstruction pixel size 0.49 × 0.49 mm<sup>2</sup>).

#### Image analysis

The T<sub>1</sub>-weighted 3D MPRAGE MRI data were used for volumetric measurements and visual ratings. All raters were blinded to clinical information.

#### Visual assessment

Visual rating of MTA was performed within the Osirix software version 5.8.2 (Pixmeo, Geneva, Switzerland) viewing platform. The 3D T<sub>1</sub>-weighted data sets were reformatted in a coronal view, angulated perpendicularly to a line connecting the anterior and posterior commissure (AC-PC-line). Slabs of 3 mm thickness were reconstructed from the original 3D T<sub>1</sub>-weighted volume to increase signal to noise-ratio. The visual MTA rating was done separately for the right and left medial temporal lobe (MTL) in accordance to the method described

by Scheltens et al. [17], i.e., it included the assessment of the hippocampal formation (hippocampus and para-hippocampal gyrus) and of the width of the surrounding cerebrospinal fluid (CSF) spaces, e.g. the temporal horn and the choroid fissure. The visual estimate of the volume of MTL structures results in subjective MTA scores ranging from 0 (no atrophy) to 4 (severe atrophy). In MTA 0, no CSF will be seen surrounding the hippocampus; in MTA 1, there is an increase of the width of the choroid fissure; in MTA 2–4, the temporal horn gradually enlarges and there is a gradual loss of height of the hippocampal formation (see Fig. 1).

MTA rating was performed by two raters, hereafter referred to as Rater 1 and Rater 2. Rater 1 received training by an experienced neuro-radiologist (Rater 2) including example rating and feedback for 100 data sets. Randomly selected subgroups were re-evaluated for both 0.5 Tesla MRI ( $n=30$ ) and 1.5 Tesla MRI ( $n=74$ ) by Rater 1 for intra-rater reliability calculations and by Rater 2 as second reader for inter-rater reliability calculations.

#### Volumetric assessment

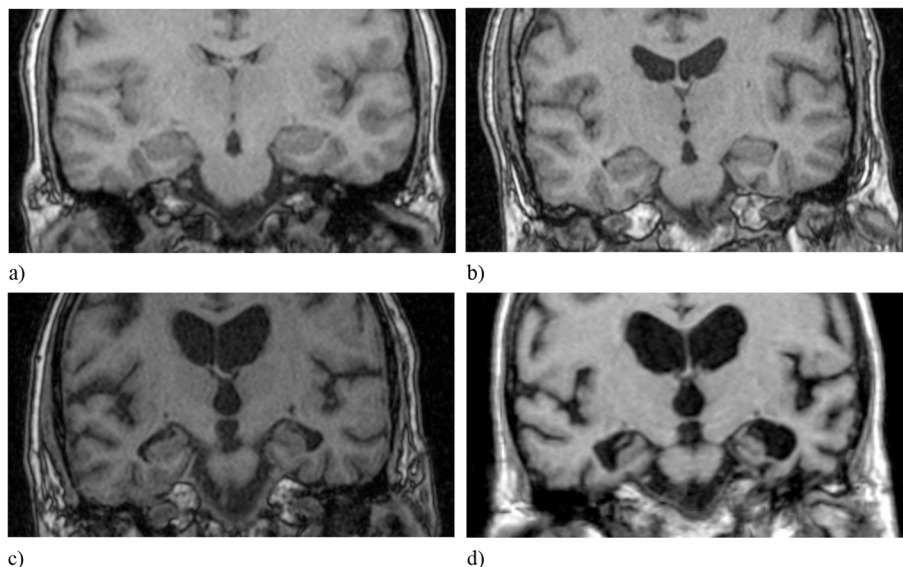
Volumetric evaluation, previously performed on the same material for different studies, comprised assessment of the hippocampal volumes of 134 0.5 Tesla examinations using manual hippocampal volumetric measurement [14] and of 560 1.5 Tesla MRI examinations using the semi-automated software suite FreeSurfer version 5.3.0 as previously described [34].

#### Statistical analysis

Demographical data were analyzed using independent-samples t-test for continuous data and  $\chi$  square for nominal data. Group comparisons were performed using Mann–Whitney U test for MTA scores and independent-samples t-test for hippocampal volumes. Intra- and inter-rater reliability of MTA assessments was determined with Cohen's weighted kappa statistics, which takes the ordered nature of the MTA scale into account. Linear correlation between ordinal MTA data and continuous hippocampal volume data was measured with a Spearman rank correlations test ( $\rho$ ). In order to examine the group classification ability of mean MTA and hippocampal volumes with respect to specific dementia diagnoses, receiver operating characteristic (ROC) analysis was performed. Lastly, different MTA cut-off values were evaluated for the differentiation of participants with specific dementia diagnoses from cognitively healthy participants. Analyses were made separately for two age groups, in order to adjust for normal age-dependent hippocampal atrophy. Sensitivity and specificity for MTA cut-off points were calculated using cross tabulation. Statistical analyses were conducted in IBM SPSS, version 26 (IBM Corp., Armonk, N.Y., USA).

#### Results

Participants with MCI or AD, SVD or mixed dementia, as shown in Table 1, were older than the cognitively healthy group. Fewer years of education were evident in the AD, SVD and mixed dementia groups than in CH. Compared



**Fig. 1** Coronal T<sub>1</sub>-weighted slices at the level of the hippocampus body of four different study participants with different levels of medial temporal lobe atrophy (MTA). **a** MTA 0 bilaterally. **b** MTA 1 bilaterally. **c** MTA 2 right side, MTA 3 left side. **d** MTA 4 bilaterally. Images (a–c) were acquired with a 1.5 Tesla scanner, whereas image (d) was acquired with a 0.5 Tesla scanner

to CH, mean MMSE scores were significantly lower in all other groups.

A box-and-whiskers plot of FreeSurfer hippocampal volume distributions identified 20 extreme outliers ( $>3 \times$  interquartile range (IQR)). In these cases, segmentations were of poor quality and reported volumes discrepant to visual assessment. Extreme outliers were hence deemed invalid and the volumes were excluded from further analyses.

**Reliability**

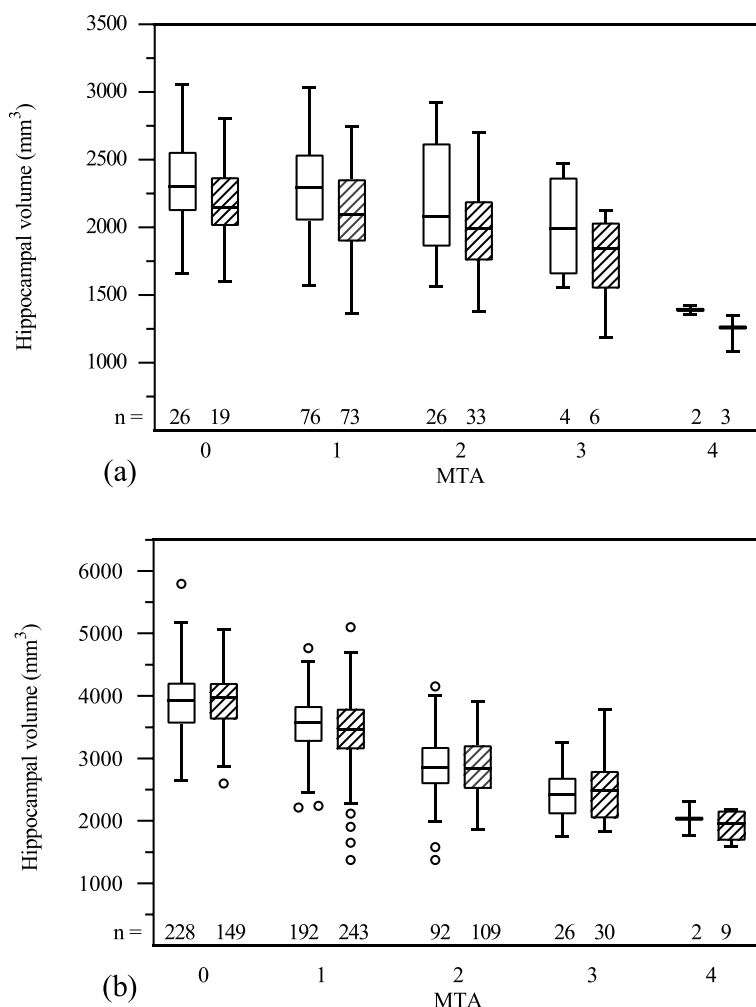
For the 0.5 Tesla MRI exams, intra-rater weighted kappa values were 0.78 on both right and left sides. For the 1.5 Tesla exams, intra-rater weighted kappa was 0.71 on the right side and 0.80 on the left side. Inter-rater agreement for the 0.5 Tesla exams was 0.59 and 0.65 and for the 1.5 Tesla exams 0.53 and 0.67, on right and left side respectively.

**Correlation with quantitative hippocampal volumes**

Figure 2a and b illustrate hippocampal volumes in relation to MTA scores. The linear correlation between manually determined hippocampal volumes and MTA score was weak with a Spearman’s correlation coefficient of  $-0.20$  ( $p < 0.05$ ) on the right side and  $-0.31$  ( $p < 0.001$ ) on the left side. The linear relationship between FreeSurfer volume estimates and MTA score was moderate, with a Spearman’s correlation coefficient of  $-0.64$  ( $p < 0.001$ ) on the right side and  $-0.68$  ( $p < 0.001$ ) on the left side.

**Group differences**

Mean MTA score was significantly higher and FreeSurfer volume significantly smaller in participants with SCI, MCI or any of the dementia subtypes than in cognitively healthy (CH) subjects (Table 2). Meanwhile, for manually determined hippocampal volumes a significant reduction



**Fig. 2** Tukey boxplot for hippocampal volume vs MTA determined with **a** manual volumetry and **b** FreeSurfer volumetry. Line inside box indicates median. Whiskers indicate  $\pm 1.5$  IQR (interquartile range). White boxes: right side. Hatched boxes: left side

**Table 2** Bilateral mean MTA score and hippocampal volume in patient groups

	N	Mean MTA score	N	Manual volumetry (mm <sup>3</sup> )	N	FreeSurfer volumetry (mm <sup>3</sup> )
CH	105	0.53 (0.46)	34	2201 (310)	57	3874 (475)
SCI	184	0.71 (0.63)*	6	2283 (393)	158	3699 (466)*
MCI	249	0.95 (0.76)†	63	2211 (360)	163	3477 (565)†
Dementia						
AD	98	1.71 (0.89)†	13	1954 (310)*	76	2871 (552)†
SVD	25	1.22 (0.68)†	6	2352 (199)	15	3477 (376)*
Mixed	51	1.64 (0.89)†	9	1936 (310)*	37	2809 (547)†

Data are given as mean of right and left sides (SD)

In participants with dementia other than AD, SVD or Mixed, manual volumetry was performed in three MRI exams and FreeSurfer volumetry in 34 MRI exams

AD Alzheimer’s disease, CH Cognitively healthy, MCI Mild cognitive impairment, Mixed Mixed dementia, MTA Medial temporal lobe atrophy, SCI Subjective cognitive impairment, SVD Subcortical vascular dementia

\* Significant results compared to CH,  $p < 0.05$ . † Significant results compared to CH,  $p < 0.001$

compared to CH was only observed in AD and mixed dementia patients.

**Discrimination ability**

The ability of mean MTA score and hippocampal volume to distinguish between patients with dementia subtypes and CH participants is reported in Table 3. Both MTA and FreeSurfer showed good discriminatory ability between AD and CH as well as between mixed dementia and CH. SVD was separated from CH to a fair degree by MTA and FreeSurfer, and not at all using manual volumetry.

**MTA cut-off values**

Table 4 provides age-range specific sensitivity and specificity percentages for different MTA score thresholds for the discrimination of investigated dementia entities from CH. For the age group below 75 years, at an MTA score threshold of 1, all three dementia subtypes were recognized with a sensitivity of over 80% (specificity 67.7%). In the age group  $\geq 75$  years, all CH ( $n=6$ ) were rated

MTA  $\geq 1$ . In this age group, most acceptable sensitivity and specificity resulted with a higher MTA threshold of 1.5. The SVD group at or above 75 years age is considered too small ( $n=6$ ) to provide reliable threshold values.

**Discussion**

Our objective was to examine reliability and validity of MTA in a memory clinic patient population. Intra and inter-rater agreement as a measure of reliability was found to be substantial to moderate. Validity of MTA was tested both with respect to correlation between MTA and quantitative hippocampal volumes and with respect

**Table 3** Discrimination ability of mean MTA score and hippocampal volumes between dementia subtypes and CH

	MTA	Manual volumetry	FreeSurfer volumetry
Dementia			
AD	0.87 (0.82–0.92)†	0.75 (0.60–0.91)*	0.91 (0.87–0.96)†
SVD	0.79 (0.68–0.90)†	0.36 (0.17–0.56)	0.75 (0.61–0.89)*
Mixed	0.86 (0.80–0.93)†	0.73 (0.56–0.91)*	0.94 (0.88–0.99)†

Data are given as Area under the curve, AUC (95% confidence interval (CI)), based on receiver operating characteristic analysis (ROC)

AD Alzheimer’s disease, CH Cognitively healthy, Mixed Mixed dementia, MTA Medial temporal lobe atrophy, SVD Subcortical vascular dementia

\*  $p < 0.05$ , †  $p < 0.001$

**Table 4** Sensitivity and specificity (%) vs CH for different MTA score thresholds

MTA cut-off	Sensitivity			Specificity
	AD	SVD	Mixed	
<i>≤ 74 years</i>				
N	77	19	26	99
0.5	96.1	89.5	92.3	36.4
1.0	81.8	84.2	80.8	67.7
1.5	66.2	47.4	53.8	97.0
2.0	48.1	26.3	42.3	100
<i>≥ 75 years</i>				
N	21	6	25	6
0.5	95.2	83.3	100	0
1.0	90.5	66.7	88.0	16.7
1.5	76.2	50.0	72.0	66.7
2.0	61.9	16.7	48.0	100

MTA cut-off indicates lowest pathological mean MTA score (average right + left) Dementia subtypes other than AD, vascular dementia and mixed dementia not shown ( $n \leq 74$  years) = 33, ( $n \geq 75$  years) = 7)

AD Alzheimer’s disease, CH Cognitively healthy, Mixed Mixed dementia, MTA Medial temporal lobe atrophy, SVD Subcortical vascular dementia

to the ability of MTA to discriminate between dementia groups and CH. The MTA score correlated significantly with hippocampal volumes, and could readily separate AD and mixed dementia from the cognitively healthy group.

Intra-rater agreement was substantial, as interpreted according to Landis and Koch [35]. There was moderate to substantial inter-rater agreement, without any obvious difference between 0.5 Tesla and 1.5 Tesla images. Rater 2 showed a tendency to give higher scores than Rater 1, but out of a total of 208 ratings, comprising right and left side ratings of 104 MRI examinations, only two ratings differed more than one score point between the two raters. Inter-rater variability of the MTA scale has also been investigated in previous studies, with agreement varying from fair to good, with kappa values ranging from 0.28 to 0.51, up to a substantial agreement with a weighted kappa 0.84 [36–39]. A decrease in agreement over time for radiologists not working together has been shown [37]. In our case, Rater 1 was a radiology resident and Rater 2 an experienced neuro-radiologist working in a different department. The level of expertise of the raters might influence the rating, although while one study that compared expert with non-expert readers observed improved performance with extended practise in non-expert readers [40], another study found no difference in inter-rater agreement due to level of experience [36].

Validity was assessed in two ways: a) as correlation between MTA and quantitative hippocampal volumes and b) as the ability of the MTA score to discriminate among patient groups. The correlation between FreeSurfer hippocampal volumes and MTA was moderate, but a weaker correlation, yet still statistically significant, was observed for manual volumetry. Our results, based on a heterogeneous study population, are in line with previous studies, with similar modest correlations between manual volumetry and MTA [22–24], and higher correlations in studies using (semi-)automated methods, such as FreeSurfer or NeuroQuant [25, 26]. Despite such findings, good agreement of hippocampal volumes has been reported between FreeSurfer and manual volumetry [41, 42], although different definitions of anatomical boundaries lead to a bias with larger FreeSurfer volumes than manually determined volumes [43].

Both MTA score and FreeSurfer volumes permitted good discrimination between the AD group and CH group, with AUC values comparable to previous studies [19, 26, 44, 45]. Based on MTA score and FreeSurfer volumes, good discrimination between mixed dementia patients and CH group was also attained. As can be expected, considering the underlying neurodegeneration, the mixed dementia group showed increased MTA scores and decreased hippocampal volumes to almost the

same extent as the AD group. Patients with SVD had also higher MTA scores and smaller FreeSurfer hippocampal volumes than the CH group, supporting previous reports of concurrent hippocampal atrophy in SVD [4–6, 8]. Although FreeSurfer volumes of patients with MCI and SVD were almost indistinguishable, MTA scores were higher in the SVD group ( $p < 0.05$ ). This finding may reflect that the MTA score not only assesses hippocampal volume but also the surrounding CSF spaces, which might be indicative of subcortical and global brain atrophy [24, 46], rather than isolated hippocampal atrophy. Whereas subcortical atrophy may be a feature of SVD, the MCI group is heterogeneous and contains participants who remain cognitively stable.

MTA cut-off values that differentiate patients with AD from controls have previously been suggested by different research groups, and range from  $\geq 1$  to  $\geq 2.5$  depending on patient age [44, 47, 48]. In the present material, recommended threshold values are  $MTA \geq 1$  in the age group below 75 years and  $MTA \geq 1.5$  in participants 75 years or older. In contrast with previous studies, we tested the various cut-off values in SVD and mixed dementia groups as well as in AD, and found similar sensitivity for SVD in the younger age group and mixed dementia as for AD.

We have selected cut-off values that prioritize sensitivity over specificity levels. Higher cut-off levels, of  $MTA \geq 1.5$  and  $MTA \geq 2$ , respectively, could be justified to avoid false positive tests, but at the cost of a lower detection rate. With the proposed thresholds, 31 out of 149 examinations of participants with confirmed AD or mixed dementia would have been classified as having no MTL atrophy. FreeSurfer hippocampal volumes were available in 23 of these “misclassified” examinations. Comparison of their mean FreeSurfer volumes showed a significantly larger volume in the misclassified group, with 3479 (SD 417)  $\text{mm}^3$  vs 2690 (SD 457)  $\text{mm}^3$  in the correctly classified group ( $p < 0.001$ ), suggesting that the MTA scores reflect actual hippocampal size and as previously reported [49] there may indeed be a subset of AD patients without pronounced hippocampal atrophy. The variation of proposed cut-off values in studies may naturally also be affected by the subjective nature of the MTA scale. A smaller study reported different optimal cut-off values set by the two raters [50], even though inter-rater correlations were high. The accuracy of the MTA cut-off increased when the average between the two raters were used. Consensus decision of several raters was applied in the original study of MTA [17], which, however, is seldom practicable in routine clinical work.

The present study suggests that MTA is a reliable and valid marker of MTL atrophy even in a heterogeneous patient population. MTL atrophy is not specific to AD

and our findings indicate that MTA is sensitive to atrophy also in patients with SVD and mixed dementia. As MTA is associated with cognitive dysfunction in patients with cerebral vascular disease as well as in AD, MTA is an important piece of information that should be reported and should be regarded along with other radiological findings in patients with cognitive impairment.

Limitations of our study include the transition between two MRI scanners operating at different field strength, reflecting the reality in many radiology departments, where the installed MRI systems often consist of scanners from different manufacturers and of different field strengths. For the purposes of this study, MTA ratings from 0.5 Tesla and 1.5 Tesla MRI exams were not distinguished in the statistical analysis. Eventual influence of field strength on the correlation assessment between MTA ratings and volumetric methods was not accessible, since manual volumetry was performed only on 0.5 Tesla scans and FreeSurfer volumetry only on 1.5 Tesla scans. To best of our knowledge, no previous studies have compared MTA performance at different field strengths. One study [51], however, reported substantial to excellent agreement between 1.5 Tesla MRI and 64-detector row computed tomography (CT) images, a modality which offers clearly less image contrast than 0.5 Tesla MRI. Another limitation is the small group sizes in the older age group. This was particularly notable when testing MTA cut-off points, where specificity values should be interpreted with caution. Few examinations were assigned the highest MTA score, possibly affecting linear correlations.

## Conclusions

In conclusion, our findings suggest that the MTA scale is a reliable and valid marker of medial temporal lobe atrophy and of use in the assessment of patients with cognitive impairment, even in a heterogeneous clinical patient population.

## Abbreviations

AC-PC: Anterior and posterior commissure; AD: Alzheimer's disease; AUC: Area under the curve; CDR: Clinical Dementia Rating; CH: Cognitively healthy; CI: Confidence interval; CSF: Cerebrospinal fluid; CT: Computed tomography; EXIT: Executive Interview; FOV: Field-of-view; GDS: Global deterioration scale; IQR: Interquartile range; MCI: Mild cognitive impairment; Mixed: Mixed dementia; MMSE: Mini mental state examination; MRI: Magnetic resonance imaging; MTA: Medial temporal lobe atrophy; MTL: Medial temporal lobe; ROC: Receiver operating characteristic; SCI: Subjective cognitive impairment; SD: Standard deviation; STEP: Stepwise Comparative Status Analysis; SVD: Subcortical vascular dementia; TE: Echo time; TR: Repetition time; WMC: White matter changes.

## Acknowledgements

We would like to thank Lena Cavallin, MD, PhD, for training, co-rating and insights, Marie Johansson and Eva Bringman for administrating the study, Erik Olsson and Niklas Klasson for help with the MRI analyses and the staff at the memory clinic in Mölndal for data collection.

## Authors' contributions

AM assessed the radiological material, drafted the manuscript, performed the statistical analysis and interpreted data. DZ and SM interpreted data, revised the manuscript and supervised. CE designed the study, interpreted data, revised the manuscript and supervised. All authors read and approved the final manuscript.

## Funding

Open access funding provided by University of Gothenburg. This work was financed by grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF agreement [Grant number ALFGBG-784831 and ALFGBG-727661]. Additional support was received from the Sahlgrenska University Hospital, the Swedish Research Council, Swedish Brain Power, the Swedish Dementia Foundation, the Swedish Alzheimer Foundation, Stiftelsen Psykiatriska forskningsfonden, and Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse. The funding sources were not involved in the drafting of this manuscript.

## Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to privacy of study participants but are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

The Gothenburg MCI study was approved by the regional ethical review board (Regionala etikprövningsnämnden) in Gothenburg (approval number: L091-99, 1999; T479-11, 2011), and is conducted in accordance with the Declaration of Helsinki of 1975 and 1983. Written informed consent is obtained from all participants in the Gothenburg MCI study.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Radiology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. <sup>2</sup>Neuroradiology, Sahlgrenska sjukhuset, Blå stråket 5, Gothenburg 413 46, Sweden. <sup>3</sup>Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>4</sup>Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. <sup>5</sup>Department of Immunology and Transfusion Medicine, Region Västra Götaland, Sahlgrenska University Hospital, Gothenburg, Sweden.

Received: 2 February 2021 Accepted: 12 July 2021

Published online: 24 July 2021

## References

1. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 1991;82(4):239–59.
2. Bobinski M, de Leon MJ, Wegiel J, Desanti S, Convit A, Saint Louis LA, et al. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. *Neuroscience.* 2000;95(3):721–5.
3. Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, et al. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology.* 1997;49(3):786–94.
4. Laakso MP, Partanen K, Riekkinen P, Lehtovirta M, Helkala EL, Hallikainen M, et al. Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia: an MRI study. *Neurology.* 1996;46(3):678–81.
5. Du AT, Schuff N, Laakso MP, Zhu XP, Jagust WJ, Yaffe K, et al. Effects of subcortical ischemic vascular dementia and AD on entorhinal cortex and hippocampus. *Neurology.* 2002;58(11):1635–41.



6. Bastos-Leite AJ, van der Flier WM, van Straaten EC, Staekenborg SS, Scheltens P, Barkhof F. The contribution of medial temporal lobe atrophy and vascular pathology to cognitive impairment in vascular dementia. *Stroke*. 2007;38(12):3182–5.
7. Eckerstrom C, Olsson E, Klasson N, Bjerke M, Gothlin M, Jonsson M, et al. High white matter lesion load is associated with hippocampal atrophy in mild cognitive impairment. *Dement Geriatr Cogn Disord*. 2011;31(2):132–8.
8. van de Pol L, Gertz HJ, Scheltens P, Wolf H. Hippocampal atrophy in subcortical vascular dementia. *Neurodegener Dis*. 2011;8(6):465–9.
9. Logue MW, Posner H, Green RC, Moline M, Cupples LA, Lunetta KL, et al. Magnetic resonance imaging-measured atrophy and its relationship to cognitive functioning in vascular dementia and Alzheimer's disease patients. *Alzheimers Dement*. 2011;7(5):493–500.
10. Arba F, Quinn T, Hankey GJ, Ali M, Lees KR, Inzitari D, et al. Cerebral small vessel disease, medial temporal lobe atrophy and cognitive status in patients with ischaemic stroke and transient ischaemic attack. *Eur J Neurol*. 2017;24(2):276–82.
11. Du AT, Schuff N, Amend D, Laakso MP, Hsu YY, Jagust WJ, et al. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2001;71(4):441–7.
12. Rusinek H, De Santi S, Frid D, Tsui WH, Tarshish CY, Convit A, et al. Regional brain atrophy rate predicts future cognitive decline: 6-year longitudinal MR imaging study of normal aging. *Radiology*. 2003;229(3):691–6.
13. Devanand DP, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, et al. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology*. 2007;68(11):828–36.
14. Eckerstrom C, Olsson E, Borga M, Ekholm S, Ribbelin S, Rolstad S, et al. Small baseline volume of left hippocampus is associated with subsequent conversion of MCI into dementia: the Goteborg MCI study. *J Neurol Sci*. 2008;272(1–2):48–59.
15. Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr Alzheimer Res*. 2009;6(4):347–61.
16. Brinkmann BH, Guragain H, Kenney-Jung D, Mandrekar J, Watson RE, Welker KM, et al. Segmentation errors and intertest reliability in automated and manually traced hippocampal volumes. *Ann Clin Transl Neurol*. 2019;6(9):1807–14.
17. Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, 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):967–72.
18. 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):1225–33.
19. Bresciani L, Rossi R, Testa C, Geroldi C, Galluzzi S, Laakso MP, et al. Visual assessment of medial temporal atrophy on MR films in Alzheimer's disease: comparison with volumetry. *Aging Clin Exp Res*. 2005;17(1):8–13.
20. Schoonenboom NS, van der Flier WM, Blankenstein MA, Bouwman FH, Van Kamp GJ, Barkhof F, et al. CSF and MRI markers independently contribute to the diagnosis of Alzheimer's disease. *Neurobiol Aging*. 2008;29(5):669–75.
21. Westman E, Cavallin L, Muehlboeck JS, Zhang Y, Mecocci P, Vellas B, et al. Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer's disease. *PLoS One*. 2011;6(7):e22506.
22. Wahlund LO, Julin P, Lindqvist J, Scheltens P. Visual assessment of medial temporal lobe atrophy in demented and healthy control subjects: correlation with volumetry. *Psychiatry Res*. 1999;90(3):193–9.
23. Cavallin L, Bronge L, Zhang Y, Oksengard AR, Wahlund LO, Fratiglioni L, et al. Comparison between visual assessment of MTA and hippocampal volumes in an elderly, non-demented population. *Acta Radiol*. 2012;53(5):573–9.
24. Clerx L, van Rossum IA, Burns L, Knol DL, Scheltens P, Verhey F, et al. Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. *Neurobiol Aging*. 2013;34(8):2003–13.
25. Fischbach-Boulanger C, Fitsiori A, Noblet V, Baloglu S, Oesterle H, Draghici S, et al. T1- or T2-weighted magnetic resonance imaging: what is the best choice to evaluate atrophy of the hippocampus? *Eur J Neurol*. 2018;25(5):775–81.
26. Persson K, Barca ML, Cavallin L, Braekhus A, Knapskog AB, Selbaek G, 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.
27. Wallin A, Nordlund A, Jonsson M, Lind K, Edman Å, Göthlin M, et al. The Gothenburg MCI study: design and distribution of Alzheimer's disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. *J Cereb Blood Flow Metab*. 2016;36(1):114–31.
28. Auer S, Reisberg B. The GDS/FAST staging system. *Int Psychogeriatr*. 1997;9(Suppl 1):167–71.
29. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939–44.
30. Erkinjuntti T, Inzitari D, Pantoni L, Wallin A, Scheltens P, Rockwood K, et al. Research criteria for subcortical vascular dementia in clinical trials. *J Neural Transm Suppl*. 2000;59:23–30.
31. Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology*. 1998;51(6):1546–54.
32. Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. *Neurology*. 2011;76(11):1006–14.
33. McKeith IG, Perry EK, Perry RH. Report of the second dementia with Lewy body international workshop: diagnosis and treatment. Consortium on Dementia with Lewy Bodies. *Neurology*. 1999;53(5):902–5.
34. Eckerstrom C, Klasson N, Olsson E, Selnes P, Rolstad S, Wallin A. Similar pattern of atrophy in early- and late-onset Alzheimer's disease. *Alzheimers Dement (Amst)*. 2018;10:253–9.
35. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–74.
36. Scheltens P, Launer LJ, Barkhof F, Weinstein HC, van Gool WA. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. *J Neurol*. 1995;242(9):557–60.
37. Cavallin L, Loken K, Engedal K, Oksengard AR, Wahlund LO, Bronge L, et al. Overtime reliability of medial temporal lobe atrophy rating in a clinical setting. *Acta Radiol*. 2012;53(3):318–23.
38. Velickaite V, Ferreira D, Cavallin L, Lind L, Ahlstrom H, Kilander L, et al. Medial temporal lobe atrophy ratings in a large 75-year-old population-based cohort: gender-corrected and education-corrected normative data. *Eur Radiol*. 2018;28(4):1739–47.
39. Martensson G, Hakansson C, Pereira JB, Palmqvist S, Hansson O, van Westen D, et al. Medial temporal atrophy in preclinical dementia: visual and automated assessment during six year follow-up. *NeuroImage Clin*. 2020;27:102310.
40. Boutet C, Chupin M, Colliot O, Sarazin M, Mutlu G, Drier A, et al. Is radiological evaluation as good as computer-based volumetry to assess hippocampal atrophy in Alzheimer's disease? *Neuroradiology*. 2012;54(12):1321–30.
41. Morey RA, Petty CM, Xu Y, Hayes JP, Wagner HR 2nd, Lewis DV, et al. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage*. 2009;45(3):855–66.
42. Clerx L, Gronenschild EH, Echavarri C, Verhey F, Aalten P, Jacobs HI. Can FreeSurfer compete with manual volumetric measurements in Alzheimer's disease? *Curr Alzheimer Res*. 2015;12(4):358–67.
43. Wenger E, Martensson J, Noack H, Bodammer NC, Kuhn S, Schaefer S, et al. Comparing manual and automatic segmentation of hippocampal volumes: reliability and validity issues in younger and older brains. *Hum Brain Mapp*. 2014;35(8):4236–48.
44. Ferreira D, Cavallin L, Larsson EM, Muehlboeck JS, Mecocci P, Vellas B, 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):277–90.
45. Koikkalainen JR, Rhodius-Meester HFM, Frederiksen KS, Bruun M, Hasselbalch SG, Baroni M, et al. Automatically computed rating scales from MRI for patients with cognitive disorders. *Eur Radiol*. 2019;29(9):4937–47.

46. Knoop AJ, van der Graaf Y, Appelman AP, Gerritsen L, Mali WP, Geerlings MI. Visual rating of the hippocampus in non-demented elders: does it measure hippocampal atrophy or other indices of brain atrophy? The SMART-MR study. *Hippocampus*. 2009;19(11):1115–22.
47. Pereira JB, Cavallin L, Spulber G, Aguilar C, Mecocci P, Vellas B, et al. Influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cut-offs. *J Intern Med*. 2014;275(3):317–30.
48. Rhodius-Meester HFM, Benedictus MR, Wattjes MP, Barkhof F, Scheltens P, Muller M, 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.
49. Burnham SC, Bourgeat P, Dore V, Savage G, Brown B, Laws S, et al. Clinical and cognitive trajectories in cognitively healthy elderly individuals with suspected non-Alzheimer's disease pathophysiology (SNAP) or Alzheimer's disease pathology: a longitudinal study. *Lancet Neurol*. 2016;15(10):1044–53.
50. Vanhoenacker AS, Sneyers B, De Keyser F, Heye S, Demaerel P. Evaluation and clinical correlation of practical cut-offs for visual rating scales of atrophy: normal aging versus mild cognitive impairment and Alzheimer's disease. *Acta Neurol Belg*. 2017;117(3):661–9.
51. Wattjes MP, Henneman WJ, van der Flier WM, de Vries O, Traber F, Geurts JJ, et al. Diagnostic imaging of patients in a memory clinic: comparison of MR imaging and 64-detector row CT. *Radiology*. 2009;253(1):174–83.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

