



## Validating ASHS-T1 automated entorhinal and transentorhinal cortical segmentation in Alzheimer's disease

Yi-En Quek<sup>a,\*</sup>, Pierrick Bourgeat<sup>b</sup>, Yi Leng Fung<sup>a</sup>, Simon J. Vogrin<sup>c</sup>, Steven J. Collins<sup>c,d</sup>, Stephen C. Bowden<sup>a,c</sup>, for the Alzheimer's Disease Neuroimaging Initiative<sup>1</sup>

<sup>a</sup> Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria, Australia

<sup>b</sup> The Australian e-Health Research Centre, CSIRO Health and Biosecurity, Brisbane, Queensland, Australia

<sup>c</sup> Department of Clinical Neurosciences, St Vincent's Hospital Melbourne, Fitzroy, Victoria, Australia

<sup>d</sup> Department of Medicine (RMH), The University of Melbourne, Parkville, Victoria, Australia

### ARTICLE INFO

#### Keywords:

Entorhinal cortex  
Transentorhinal cortex  
Manual segmentation  
Automated segmentation  
Magnetic resonance imaging  
Alzheimer's disease

### ABSTRACT

The current study aimed to validate entorhinal and transentorhinal cortical volumes measured by the automated segmentation tool Automatic Segmentation of Hippocampal Subfields (ASHS-T1). The study sample comprised 34 healthy controls (HCs), 37 individuals with amnesic mild cognitive impairment (aMCI), and 29 individuals with Alzheimer's disease (AD) dementia from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Entorhinal and transentorhinal cortical volumes were assessed using ASHS-T1, manual segmentation, as well as a widely used automated segmentation tool, FreeSurfer v6.0.1. Mean differences, intraclass correlation coefficients, and Bland–Altman plots were computed. ASHS-T1 tended to underestimate entorhinal and transentorhinal cortical volumes relative to manual segmentation and FreeSurfer. There was variable consistency and low agreement between ASHS-T1 and manual segmentation volumes. There was low-to-moderate consistency and low agreement between ASHS-T1 and FreeSurfer volumes. There was a trend toward higher consistency and agreement for the entorhinal cortex in the aMCI and AD groups compared to the HC group. Despite the differences in volume measurements, ASHS-T1 was sensitive to entorhinal and transentorhinal cortical atrophy in both early and late disease stages. Based on the current study, ASHS-T1 appears to be a promising tool for automated entorhinal and transentorhinal cortical volume measurement in individuals with likely underlying AD.

### 1. Introduction

In Alzheimer's disease (AD), the entorhinal and transentorhinal cortices are typically among the earliest brain regions to show pathological changes. Tau pathology staging in postmortem AD brain tissue has shown that pathological tau deposits commence in the transentorhinal and entorhinal regions (Braak and Braak, 1991). Using structural magnetic resonance imaging (MRI), pathological tau accumulation in these regions has been observed to correlate with cerebral atrophy (Xie et al., 2018). Other MRI studies have also shown that the volumes of the entorhinal and transentorhinal cortices are reduced in patients with amnesic mild cognitive impairment (aMCI) relative to healthy controls (HCs) and also predict cognitive decline and conversion

from aMCI to AD dementia (Kulason et al., 2019; Penanen et al., 2004; Stoub et al., 2010; Tapiola et al., 2008; Tward et al., 2017; Venneri et al., 2011). These results suggest that volume loss in the entorhinal and transentorhinal cortices may be a useful biomarker for early, objective detection of AD.

There are two primary methods to assess the volumes of the entorhinal and transentorhinal cortices on MRI scans. Manual segmentation by an expert observer is widely considered to be the "gold standard", but it is time-consuming and labor-intensive and is thus impractical for routine use in the clinical setting (Bobinski et al., 1999). Automated segmentation requires minimal supervision and allow rapid volumetric calculation, and it may thus be a promising tool for routine clinical use. Few studies have validated automated segmentation to measure

\* Corresponding author at: Melbourne School of Psychological Sciences, Redmond Barry Building, The University of Melbourne, Parkville, VIC 3010, Australia.  
E-mail address: [yien@student.unimelb.edu.au](mailto:yien@student.unimelb.edu.au) (Y.-E. Quek).

<sup>1</sup> Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

<https://doi.org/10.1016/j.psyresns.2023.111707>

Received 3 April 2023; Received in revised form 25 July 2023; Accepted 9 August 2023

Available online 22 August 2023

0925-4927/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

entorhinal cortex volumes, and no study to date has validated automated segmentation to measure transentorhinal cortex volumes. The studies that have evaluated automated segmentation to measure entorhinal cortex volumes have reported very weak concordance with manual segmentation (Fung et al., 2019; Lehmann et al., 2010). These studies highlight that the entorhinal cortex is a challenging region for automated volumetric analysis.

A major challenge for the automated segmentation of the entorhinal cortex, and to some extent the transentorhinal cortex, relates to the proximity of the mesial temporal lobe (MTL) to the tentorium cerebelli. A large section of the entorhinal cortex and some parts of the transentorhinal cortex lie directly adjacent to the tentorium cerebelli. On T1-weighted MRI scans, the tentorium cerebelli has a similar intensity to that of gray matter (Penumetcha et al., 2011). Consequently, automated segmentation methods such as FreeSurfer and Advanced Normalization Tools (ANTs) that rely on intensity variations to distinguish between different tissue types often mislabel portions of the tentorium cerebelli as entorhinal or transentorhinal cortex (Xie et al., 2016). The over-segmentation of the entorhinal and transentorhinal cortices to include portions of the tentorium cerebelli would certainly result in mis-estimations of the volumes of these regions, which may confound research findings based on these measurements.

Recently, a new automated segmentation tool, Automatic Segmentation of Hippocampal Subfields (ASHS-T1), was developed to address the issue of the over-segmentation of the entorhinal and transentorhinal cortices (Xie et al., 2019; Yushkevich et al., 2015). ASHS-T1 uses a multi-atlas approach to label the MTL subregions and attempts to reduce mislabeling of the tentorium cerebelli as entorhinal or transentorhinal cortex by explicitly labeling the tentorium cerebelli (Xie et al., 2019). Xie et al. (2019) investigated the degree of dura mislabeling as cortex by ASHS-T1 and FreeSurfer against manually labeled tentorium cerebelli and showed that ASHS-T1 mislabeled only 6.5% of tentorium cerebelli voxels as gray matter, whereas FreeSurfer mislabeled 62.4% of tentorium cerebelli voxels as gray matter. The substantially reduced mislabeling of the tentorium cerebelli as entorhinal or transentorhinal cortex by ASHS-T1 compared to FreeSurfer suggests that ASHS-T1 should provide more accurate estimations of the volumes of these regions.

Whilst ASHS-T1 appears to be a more promising method to automatically estimate volumes of the entorhinal and transentorhinal cortices, it has yet to be comprehensively evaluated against the “gold standard” manual segmentation. The only study comparing ASHS-T1 against manual segmentation tested ASHS-T1 in the same sample that was used to construct the ASHS-T1 multi-atlas (Xie et al., 2019). Evaluating ASHS-T1 on the same sample used to construct its multi-atlas may produce more optimistic results than would be expected when evaluated in an independent sample (Baumann, 2003). Moreover, the sample used by Xie et al. (2019) was small, comprising only 15 HCs and 14 patients with aMCI, and did not include patients with AD dementia. Previous studies have suggested that automated methods may perform more poorly in brains with severe atrophy, such as that seen in patients with AD (Lehmann et al., 2010; Sánchez-Benavides et al., 2010). It remains uncertain whether the performance of ASHS-T1 is likewise affected by the severe atrophy present in the brains of patients with AD. Considering these limitations of the study by Xie et al. (2019), the aims of the current study were (1) to compare entorhinal and transentorhinal cortical volumes measured by ASHS-T1 against those measured by manual segmentation and FreeSurfer in HCs, individuals with aMCI, and individuals with AD dementia, and (2) to evaluate the clinical utility of ASHS-T1 entorhinal and transentorhinal cortical volumes in detecting atrophy in individuals with aMCI and individuals with AD dementia.

## 2. Methods

### 2.1. Participants

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). For up-to-date information, see [www.adni-info.org](http://www.adni-info.org).

The current study sample comprised 100 participants (34 HCs, 37 with aMCI, and 29 with AD dementia). Only participants with MRI scans acquired at 3.0 T were included in the current study. The ADNI diagnostic criteria are described in Petersen et al. (2010). The sample demographic characteristics are presented in Table 1.

### 2.2. MRI acquisition and preprocessing

The MRI scans were acquired across various scanners (General Electric, Philips, and Siemens) at multiple sites. The ADNI MRI protocols have been described previously (in Jack et al. (2008) and also at <http://adni.loni.usc.edu/methods/documents/mri-protocols/>).

Variations in MRI scan orientation can introduce substantial variability in MRI-based measurements (Bartzokis et al., 1998). Hence, all MRI scans were first segmented using FreeSurfer v6.0.1 and then aligned to a common orientation, perpendicular to the long axis of the FreeSurfer hippocampal segmentation, using a python script (<https://doi.org/10.25919/8kqn-d006>), to improve measurement reliability. All MRI scans were then re-sampled to  $0.3 \times 0.3 \times 1.0 \text{ mm}^3$  by cubic spline interpolation to enhance scan resolution.

### 2.3. MRI processing

#### 2.3.1. ASHS-T1

Entorhinal and transentorhinal cortices were automatically segmented using ASHS-T1 (Xie et al., 2019; Yushkevich et al., 2015). Due to differences in the MRI protocol between the ADNI data and the ASHS-T1 atlas set, the joint label fusion (JLF) output was used (Xie et al., 2019). Volume measurements were extracted from the ASHS-T1 segmentations. Intracranial volume (ICV) was also measured using ASHS-T1 (Xie et al., 2019).

#### 2.3.2. Manual segmentation

Entorhinal and transentorhinal cortices were manually segmented

**Table 1**  
Sample demographic characteristics.

Characteristic	HC (n = 34)	aMCI (n = 37)	AD (n = 29)	ANOVA or Chi-Square <sup>a</sup>	Post Hoc tests <sup>b</sup>
Age (years)	70.09 (5.68)	74.33 (7.37)	72.88 (9.17)	$p = 0.058$ , $\eta^2 = 0.06$	-
Gender (% female)	70.6	43.2	51.7	$p = 0.063$ , $V = 0.24$	-
Education (years)	17.00 (2.62)	16.51 (2.46)	15.21 (2.64)	$p = 0.021$ , $\eta^2 = 0.08$	HC > AD
MMSE	29.24 (0.89)	27.62 (2.44)	23.17 (3.07)	$p < 0.001$ , $\eta^2 = 0.55$	HC > aMCI, AD; aMCI > AD

Note: AD = Alzheimer’s disease; aMCI = amnesic mild cognitive impairment; HC = healthy control; MMSE = Mini-Mental State Examination.

<sup>a</sup> For continuous variables, one-way ANOVA was conducted. For categorical variables, Pearson’s chi-squared test was conducted.

<sup>b</sup> Pairwise comparisons are Bonferroni-adjusted post hoc analyses.

using ITK-SNAP Version 3.8.0 (Yushkevich et al., 2006). The segmentations were conducted according to the Berron et al. (2017) MTL sub-regions manual segmentation protocol. According to the Berron et al. (2017) protocol, segmentation of the entorhinal and transentorhinal cortices commences 4.4 mm anterior to the first slice in which the hippocampal head is visible and concludes 2.2 mm and 4.4 mm, respectively, posterior to the last slice in which the hippocampal head is visible. Because the current study protocol utilized MRI scans with a 1.0 mm distance between slices, segmentation of the entorhinal and transentorhinal cortices was commenced 5 slices anterior to the first slice in which the hippocampal head is visible to avoid volume under-estimation and was concluded 2 and 4 slices, respectively, posterior to the last slice in which the hippocampal head is visible. An example manual segmentation of the entorhinal and transentorhinal cortices is provided in the Supplementary Material.

All manual segmentations were completed by a single rater (YEQ) blind to participants' diagnosis. Manual segmentations on 10 randomly selected MRI scans were repeated to assess intra-rater reliability and conducted by an independent rater (YLF) to assess inter-rater reliability. The intra-class correlation coefficient (ICC) was used to evaluate intra- and inter-rater reliability. For the entorhinal cortex volume measurements, intra-rater reliability was ICC agreement = 0.89 (95% CI 0.54 to 0.96), and inter-rater reliability was ICC agreement = 0.93 (95% CI 0.91 to 0.99). For the transentorhinal cortex volume measurements, intra-rater reliability was ICC agreement = 0.87 (95% CI 0.43 to 0.96), and inter-rater reliability was ICC agreement = 0.72 (95% CI 0.42 to 0.88).

### 2.3.3. FreeSurfer

Entorhinal and transentorhinal cortices were automatically segmented using FreeSurfer v6.0.1 (Fischl, 2012). The cortical reconstruction process was performed with the "recon-all" pipeline. The cytoarchitecturally-defined labels of the entorhinal cortex and transentorhinal cortices were used (Augustinack et al., 2013; Fischl et al., 2009). Volume measurements were extracted from the "lh.BA\_exvivo.thresh.stats" and "rh.BA\_exvivo.thresh.stats" files.

## 2.4. Statistical analysis

To examine the differences in demographic characteristics among the diagnostic groups, the one-way ANOVA procedure (with age, education, and MMSE as the dependent variables) and Pearson's chi-squared test (with sex as the dependent variable) were conducted. Where a significant difference across groups was identified, post hoc pairwise comparisons with the Bonferroni adjustment were undertaken.

To compare the entorhinal cortex and transentorhinal cortex volumes estimated by ASHS-T1 with those estimated by manual segmentation and FreeSurfer in HCs, individuals with aMCI, and individuals with AD dementia, the dependent-samples *t*-test, ICC statistic, partial Pearson correlation analysis, and Bland–Altman method were used. The dependent-samples *t*-test was used to test the significance of the differences between the segmentation methods. The magnitude of the differences between the methods was expressed using Cohen's *d*. The ICC statistic was used to evaluate consistency and agreement between the segmentation methods. The ICC statistics were based on a single-measure, two-way mixed model, and both consistency and absolute agreement were reported to assess both the consistency in the volumes as well as the agreement in the absolute volumes. The partial Pearson correlation analysis was used to examine the strength of the linear relationship between the segmentation methods, controlling for ICV. To determine the significance of the difference between comparisons, the overlap of the 95% confidence intervals of the Cohen's *d* and ICC point estimates was examined. An overlap in two confidence intervals by no more than half the average margin of error was considered statistically significantly different (i.e.,  $p < 0.05$ ; Cumming and Finch, 2005). The Bland–Altman method was used to assess the agreement between the segmentation methods. Regression lines were included to assess

proportional bias. These analyses were also repeated to compare the entorhinal cortex and transentorhinal cortex volumes estimated by FreeSurfer to those estimated by manual segmentation and are presented in the Supplementary Material.

To evaluate the utility of ASHS-T1 in detecting entorhinal and transentorhinal atrophy in individuals with aMCI and individuals with AD dementia, the volumes of the entorhinal and transentorhinal cortices estimated by ASHS-T1 were compared across the diagnostic groups. The volumes of the entorhinal and transentorhinal cortices obtained by manual segmentation and by FreeSurfer were also compared across the diagnostic groups for comparison against ASHS-T1. For each volume measurement, one-way analysis of covariance (ANCOVA) with group as the independent variable and ICV as a covariate was used. Post hoc pairwise comparisons between groups were corrected for multiple testing using the Bonferroni adjustment. The magnitude of the differences in the volume measurements between the groups was expressed using Cohen's *d*.

Statistical analyses were conducted using IBM SPSS Statistics 27.0.1.0. Bland–Altman plots were generated using R version 4.1.1 (R Core Team, 2021) with the *BlandAltmanLeh* package.

## 3. Results

### 3.1. Demographic characteristics

As shown in Table 1, there were statistically significant differences among the HC, aMCI, and AD groups in years of education and MMSE score. In particular, the HC group had significantly more years of education than the AD group ( $p = 0.020$ ). The AD group scored significantly lower on the MMSE than both the aMCI and HC groups (both  $ps < 0.001$ ), and the aMCI group scored significantly lower on the MMSE than the HC group ( $p = 0.011$ ). The groups did not significantly differ in age and sex.

### 3.2. Comparing ASHS-T1 and manual segmentation

Mean differences, standard deviation of the differences, *p* values, and Cohen's *d* estimates between ASHS-T1 and manual segmentation volumes are shown in Table 2. ICC consistency and absolute agreement statistics between ASHS-T1 and manual segmentation volumes are shown in Table 3. Partial Pearson correlation coefficients between ASHS-T1 and manual segmentation volumes, controlling for ICV, are shown in Table 4. Scatter plots of ASHS-T1 and manual segmentation volumes are presented in Fig. 1. Bland–Altman plots of ASHS-T1 and manual segmentation volumes are presented in Fig. 2.

#### 3.2.1. Entorhinal cortex

The dependent-samples *t*-tests indicated that ASHS-T1 significantly underestimated entorhinal cortex volumes compared to manual segmentation for all three diagnostic groups, with the differences large in magnitude ( $d$  range =  $-2.36$  to  $-3.43$ ). There was a trend toward less underestimation of both left and right entorhinal cortex volumes with increasing disease severity (HC  $d = -3.43$  and  $-3.40$  vs. aMCI  $d = -3.00$  and  $-2.81$  vs. AD  $d = -2.36$  and  $-2.73$ ). Overall, ICC consistency (range = 0.43 to 0.79) was higher, but not significantly so, than ICC agreement (range = 0.06 to 0.27). There was a trend toward higher ICC consistency and agreement for both left and right entorhinal cortex volumes in the aMCI group (ICC consistency = 0.79 and 0.67, ICC agreement = 0.27 and 0.19) and AD group (ICC consistency = 0.69 and 0.67, ICC agreement = 0.25 and 0.19) compared to the HC group (ICC consistency = 0.43 and 0.56, ICC agreement = 0.06 and 0.09). Likewise, there was a trend toward higher partial correlation coefficients for both left and right entorhinal cortex volumes in the aMCI group ( $r_p = 0.83$  and 0.75) and AD group ( $r_p = 0.79$  and 0.75) compared to the HC group ( $r_p = 0.55$  and 0.63). The Bland–Altman plots showed very wide 95% limits of agreement for both the left entorhinal cortex ( $\pm 1.96 SD = -565$

**Table 2**

Mean differences, standard deviation of differences, *p* values, and Cohen's *d* between ASHS-T1 vs. manual segmentation and ASHS-T1 vs. FreeSurfer entorhinal and transentorhinal cortical volumes (mm<sup>3</sup>).

Method/Region	HC			aMCI			AD		
	Mean Difference (SD)	<i>p</i> Value	Cohen's <i>d</i> (95% CI)	Mean Difference (SD)	<i>p</i> Value	Cohen's <i>d</i> (95% CI)	Mean Difference (SD)	<i>p</i> Value	Cohen's <i>d</i> (95% CI)
<b>ASHS-T1 vs. Manual</b>									
Entorhinal Cortex									
Left	<b>-381.92</b> (111.21)	< 0.001	<b>-3.43</b> (-4.32, -2.54)	<b>-321.13</b> (107.12)	< 0.001	<b>-3.00</b> (-3.76, -2.23)	<b>-290.06</b> (122.88)	< 0.001	<b>-2.36</b> (-3.07, -1.64)
Right	<b>-420.35</b> (123.63)	< 0.001	<b>-3.40</b> (-4.28, -2.51)	<b>-339.20</b> (120.91)	< 0.001	<b>-2.81</b> (-3.52, -2.08)	<b>-332.68</b> (121.67)	< 0.001	<b>-2.73</b> (-3.53, -1.93)
Transentorhinal Cortex									
Left	<b>-91.37</b> (81.48)	< 0.001	<b>-1.12</b> (-1.55, -0.69)	<b>-121.50</b> (111.34)	< 0.001	<b>-1.09</b> (-1.50, -0.68)	<b>-113.57</b> (84.07)	< 0.001	<b>-1.35</b> (-1.85, -0.84)
Right	<b>-98.11</b> (105.67)	< 0.001	<b>-0.93</b> (-1.33, -0.52)	<b>-156.42</b> (114.19)	< 0.001	<b>-1.37</b> (-1.82, -0.91)	<b>-141.99</b> (100.51)	< 0.001	<b>-1.41</b> (-1.92, -0.98)
<b>ASHS-T1 vs. FreeSurfer</b>									
Entorhinal Cortex									
Left	<b>-379.43</b> (123.26)	< 0.001	<b>-3.08</b> (-3.89, -2.26)	<b>-318.42</b> (134.58)	< 0.001	<b>-2.37</b> (-2.99, -1.73)	<b>-290.97</b> (165.17)	< 0.001	<b>-1.76</b> (-2.34, -1.17)
Right	<b>-652.71</b> (172.88)	< 0.001	<b>-3.78</b> (-4.74, -2.80)	<b>-508.86</b> (177.90)	< 0.001	<b>-2.86</b> (-3.59, -2.12)	<b>-461.46</b> (195.32)	< 0.001	<b>-2.36</b> (-3.07, -1.64)
Transentorhinal Cortex									
Left	<b>-303.42</b> (310.73)	< 0.001	<b>-0.98</b> (-1.38, -0.56)	<b>-338.51</b> (346.38)	< 0.001	<b>-0.98</b> (-1.34, -0.57)	<b>-306.53</b> (275.97)	< 0.001	<b>-1.11</b> (-1.57, -0.64)
Right	<b>-31.70</b> (205.19)	0.374	<b>-0.16</b> (-0.49, 0.19)	<b>-10.19</b> (177.85)	0.729	<b>-0.06</b> (-0.38, -0.27)	<b>-82.02</b> (223.92)	0.058	<b>-0.37</b> (-0.74, 0.01)

Note: Bold type indicates statistical significance. AD = Alzheimer's disease; aMCI = amnesic mild cognitive impairment; ASHS-T1 = Automatic Segmentation of Hippocampal Subfields-T1; CI = confidence interval; HC = healthy control.

**Table 3**

Intraclass correlation coefficients between ASHS-T1 vs. manual segmentation and ASHS-T1 vs. FreeSurfer entorhinal and transentorhinal cortical volumes.

Method/Region	ICC Consistency (95% CI)			ICC Absolute Agreement (95% CI)		
	HC	aMCI	AD	HC	aMCI	AD
<b>ASHS-T1 vs. Manual</b>						
Entorhinal Cortex						
Left	0.43 (0.11, 0.67)	0.79 (0.63, 0.89)	0.69 (0.44, 0.84)	0.06 (-0.03, 0.23)	0.27 (-0.05, 0.65)	0.25 (-0.07, 0.63)
Right	0.56 (0.28, 0.75)	0.67 (0.45, 0.82)	0.67 (0.40, 0.83)	0.09 (-0.04, 0.34)	0.19 (-0.05, 0.53)	0.19 (-0.06, 0.54)
Transentorhinal Cortex						
Left	0.57 (0.29, 0.76)	0.54 (0.27, 0.73)	0.60 (0.31, 0.79)	0.37 (-0.08, 0.68)	0.35 (-0.08, 0.66)	0.35 (-0.10, 0.69)
Right	0.37 (0.05, 0.63)	0.60 (0.35, 0.77)	0.33 (-0.04, 0.62)	0.25 (-0.07, 0.53)	0.35 (-0.10, 0.67)	0.14 (-0.09, 0.42)
<b>ASHS-T1 vs. FreeSurfer</b>						
Entorhinal Cortex						
Left	0.43 (0.11, 0.67)	0.72 (0.52, 0.85)	0.59 (0.29, 0.78)	0.07 (-0.04, 0.26)	0.28 (-0.07, 0.65)	0.26 (-0.10, 0.61)
Right	0.31 (-0.03, 0.59)	0.54 (0.27, 0.73)	0.51 (0.18, 0.73)	0.03 (-0.02, 0.14)	0.11 (-0.05, 0.39)	0.14 (-0.06, 0.44)
Transentorhinal Cortex						
Left	-0.11 (-0.43, 0.23)	-0.06 (-0.37, 0.27)	0.13 (-0.24, 0.47)	-0.05 (-0.21, 0.16)	-0.03 (-0.17, 0.17)	0.07 (-0.11, 0.30)
Right	0.11 (-0.24, 0.42)	0.17 (-0.16, 0.46)	0.06 (-0.31, 0.41)	0.11 (-0.24, 0.43)	0.17 (-0.17, 0.47)	0.05 (-0.27, 0.38)

Note: AD = Alzheimer's disease; aMCI = amnesic mild cognitive impairment; ASHS-T1 = Automatic Segmentation of Hippocampal Subfields-T1; CI = confidence interval; HC = healthy control; ICC = intraclass correlation coefficient.

to -101 mm<sup>3</sup>) and right entorhinal cortex ( $\pm 1.96 SD = -614$  to  $-115$  mm<sup>3</sup>). Regression analyses revealed a significant proportional bias for all three diagnostic groups (all *ps* < 0.01), whereby the magnitude of the difference in the volumes increased as the mean of the volumes increased.

3.2.2. Transentorhinal cortex

The dependent-samples *t*-tests indicated that, compared to manual segmentation, ASHS-T1 significantly underestimated transentorhinal cortex volumes for all three diagnostic groups, though to a lesser extent than entorhinal cortex volumes, with the differences medium to large in magnitude (*d* range =  $-0.61$  to  $-1.46$ ). The degree of underestimation was similar across the diagnostic groups. Overall, ICC consistency (range = 0.33 to 0.60) was higher, but not significantly so, than ICC agreement (range = 0.14 to 0.37). Both ICC consistency and agreement were similar across the diagnostic groups. Likewise, the partial correlation coefficients were similar across the diagnostic groups. The

Bland-Altman plots showed very wide 95% limits of agreement for both the left transentorhinal cortex ( $\pm 1.96 SD = -294$  to  $75.9$  mm<sup>3</sup>) and right transentorhinal cortex ( $\pm 1.96 SD = -347$  to  $82$  mm<sup>3</sup>) but less so than those for the entorhinal cortices. Regression analyses indicated that there was no proportional bias for any of the diagnostic groups (all *ps* > 0.05).

3.3. Comparing ASHS-T1 and FreeSurfer

Mean differences, standard deviation of the differences, *p* values, and Cohen's *d* estimates between ASHS-T1 and manual segmentation volumes are shown in Table 2. ICC consistency and absolute agreement statistics between ASHS-T1 and FreeSurfer volumes are shown in Table 3. Partial Pearson correlation coefficients between ASHS-T1 and FreeSurfer volumes, controlling for ICV, are shown in Table 4. Scatter plots of ASHS-T1 and FreeSurfer volumes are presented in Fig. 3. Bland-Altman plots of ASHS-T1 and FreeSurfer volumes are presented in

**Table 4**

Partial correlation coefficients between ASHS-T1 vs. manual segmentation and ASHS-T1 vs. FreeSurfer entorhinal and transentorhinal cortical volumes, controlling for intracranial volume.

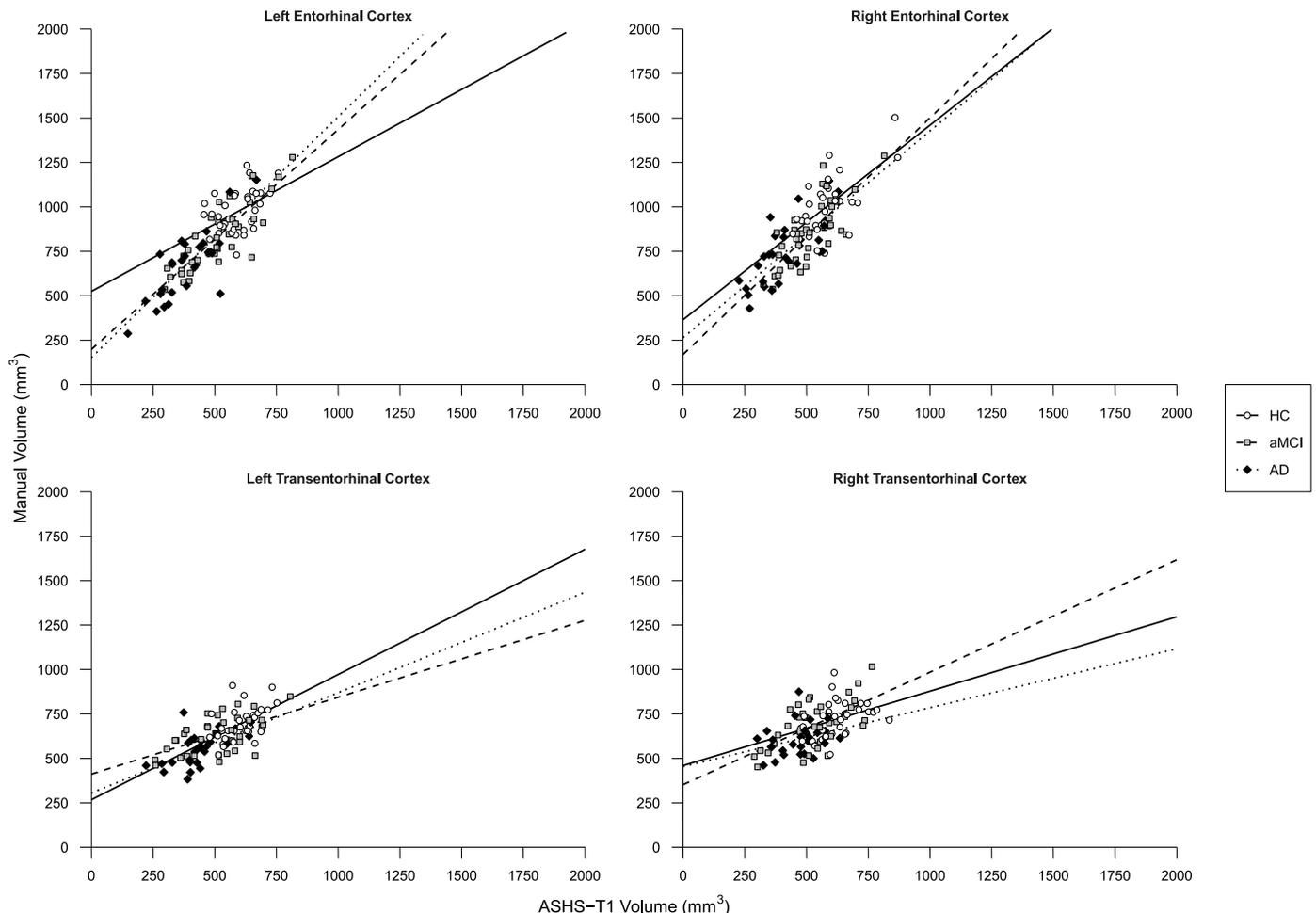
Method/ Region	Partial $r$ , $p$ Value HC	aMCI	AD
<b>ASHS-T1 vs. Manual</b>			
Entorhinal Cortex			
Left	$r_p = 0.55, p = 0.001$	$r_p = 0.83, p < 0.001$	$r_p = 0.79, p < 0.001$
Right	$r_p = 0.63, p < 0.001$	$r_p = 0.75, p < 0.001$	$r_p = 0.75, p < 0.001$
Transentorhinal Cortex			
Left	$r_p = 0.57, p < 0.001$	$r_p = 0.51, p = 0.001$	$r_p = 0.59, p = 0.001$
Right	$r_p = 0.35, p = 0.043$	$r_p = 0.54, p < 0.001$	$r_p = 0.33, p = 0.088$
<b>ASHS-T1 vs. FreeSurfer</b>			
Entorhinal Cortex			
Left	$r_p = 0.41, p = 0.017$	$r_p = 0.80, p < 0.001$	$r_p = 0.76, p < 0.001$
Right	$r_p = 0.33, p = 0.063$	$r_p = 0.62, p < 0.001$	$r_p = 0.71, p < 0.001$
Transentorhinal Cortex			
Left	$r_p = -0.28, p = 0.112$	$r_p = -0.05, p = 0.784$	$r_p = 0.22, p = 0.270$
Right	$r_p = 0.08, p = 0.642$	$r_p = 0.14, p = 0.404$	$r_p = 0.08, p = 0.701$

Note: AD = Alzheimer's disease; aMCI = amnestic mild cognitive impairment; ASHS-T1 = Automatic Segmentation of Hippocampal Subfields-T1; HC = healthy control.

**Fig. 4.**

**3.3.1. Entorhinal cortex**

The dependent-samples *t*-tests indicated that ASHS-T1 significantly underestimated entorhinal cortex volumes compared to FreeSurfer for all three diagnostic groups, with the differences large in magnitude (*d* range =  $-1.76$  to  $-3.78$ ). There was a trend toward less underestimation of both left and right entorhinal cortex volumes with increasing disease severity (HC *d* =  $-3.08$  and  $-3.78$  vs. aMCI *d* =  $-2.37$  and  $-2.86$  vs. AD *d* =  $-1.76$  and  $-2.36$ ). Overall, ICC consistency (range =  $0.31$  to  $0.72$ ) was higher, but not significantly so, than ICC agreement (range =  $0.03$  to  $0.28$ ). There was a trend toward higher ICC consistency and agreement for both left and right entorhinal cortex volumes in the aMCI group (ICC consistency =  $0.72$  and  $0.54$ , ICC agreement =  $0.28$  and  $0.11$ ) and AD group (ICC consistency =  $0.59$  and  $0.51$ , ICC agreement =  $0.26$  and  $0.14$ ) than in the HC group (ICC consistency =  $0.43$  and  $0.31$ , ICC agreement =  $0.07$  and  $0.03$ ). Likewise, there was a trend toward higher partial correlation coefficients for both left and right entorhinal cortex volumes in the aMCI group ( $r_p = 0.80$  and  $0.62$ ) and AD group ( $r_p = 0.76$  and  $0.71$ ) compared to the HC group ( $r_p = 0.41$  and  $0.33$ ). The Bland-Altman plots showed extremely wide 95% limits of agreement for both the left entorhinal cortex ( $\pm 1.96$  SD =  $-613$  to  $-49.2$  mm<sup>3</sup>) and right entorhinal cortex ( $\pm 1.96$  SD =  $-930$  to  $-158$  mm<sup>3</sup>). Regression analyses revealed a significant proportional bias for all three diagnostic groups (all *ps* <  $0.001$ ), whereby the magnitude of the difference in the volumes increased as the mean of the volumes increased.



**Fig. 1.** Scatter plots of ASHS-T1 and manual segmentation entorhinal and transentorhinal cortical volumes.

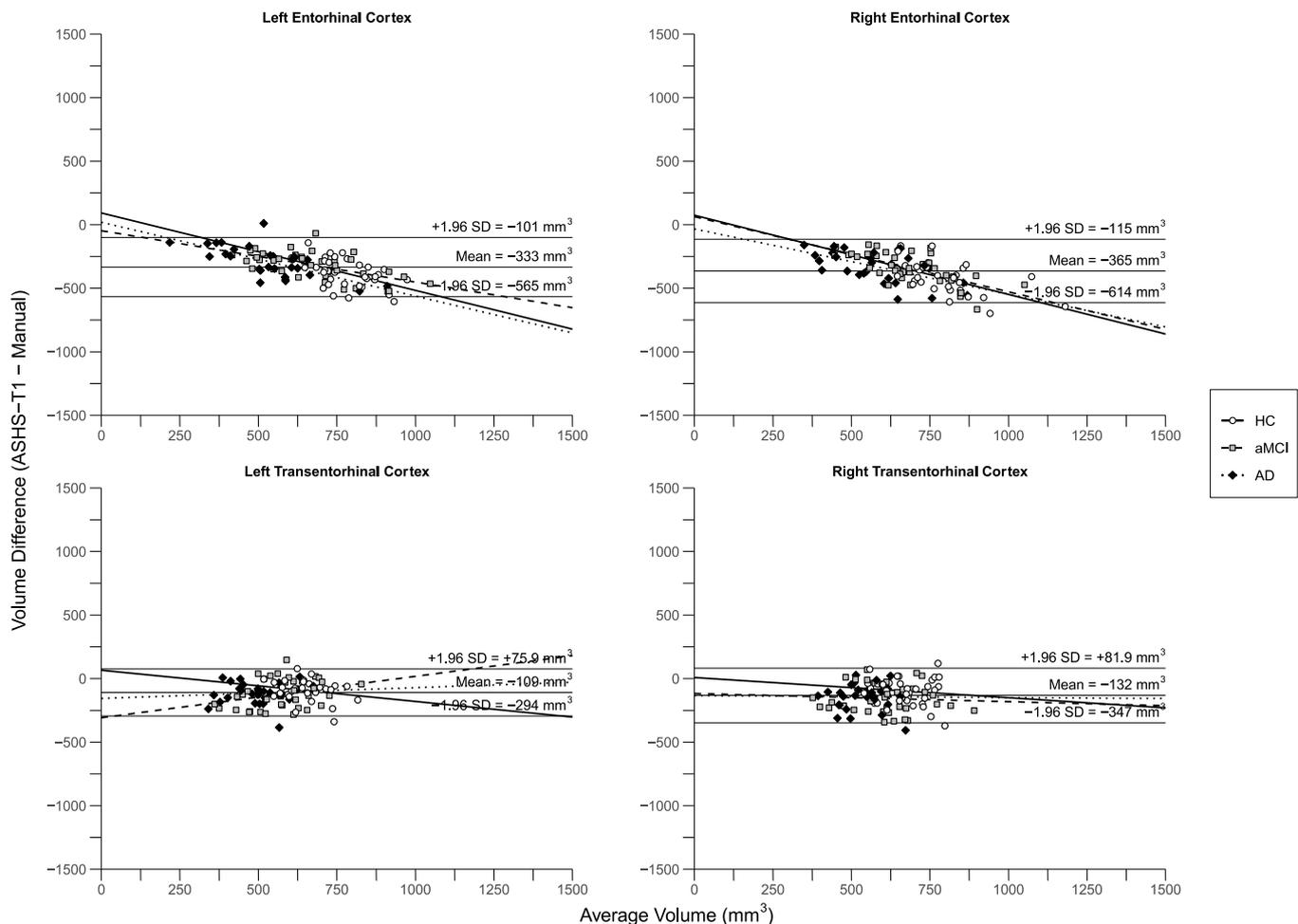


Fig. 2. Bland–Altman plots of ASHS-T1 and manual segmentation entorhinal and transentorhinal cortical volumes.

### 3.3.2. Transentorhinal cortex

The dependent-samples *t*-tests indicated that ASHS-T1 significantly underestimated left transentorhinal cortex volume compared to FreeSurfer for all three diagnostic groups, with the differences large in magnitude ( $d$  range =  $-0.98$  to  $-1.11$ ). There were no significant differences in right transentorhinal cortex volume between ASHS-T1 and FreeSurfer. However, for both the left and right transentorhinal cortices, the standard deviations of the differences in volumes were very large in all diagnostic groups ( $SD$  range =  $177.85$  to  $346.38$   $\text{mm}^3$ ). ICC consistency (range =  $-0.11$  to  $0.17$ ) was similar to ICC agreement (range =  $-0.05$  to  $0.17$ ). Both ICC consistency and agreement were similar across the diagnostic groups. Likewise, the partial correlation coefficients were similar across the diagnostic groups. The Bland–Altman plots showed extremely wide 95% limits of agreement for both the left transentorhinal cortex ( $\pm 1.96$   $SD = -930$  to  $295$   $\text{mm}^3$ ) and right transentorhinal cortex ( $\pm 1.96$   $SD = -433$  to  $357$   $\text{mm}^3$ ). Regression analyses revealed a significant proportional bias for all three diagnostic groups, whereby the magnitude of the difference in the volumes increased as the mean of the volumes increased (all  $p$ s  $< 0.001$ ), except for the right hemisphere in the aMCI group, which showed no proportional bias ( $p = 0.270$ ).

### 3.4. Evaluating volume differences across HC, aMCI, and AD groups

The means and standard deviations of the entorhinal and transentorhinal cortical volumes estimated by ASHS-T1, manual segmentation, and FreeSurfer for each diagnostic group as well as the  $p$  values and Cohen's  $d$  estimates between the diagnostic groups are presented in Table 5.

### 3.4.1. Entorhinal cortex

The one-way ANCOVAs revealed significant differences across the three diagnostic groups in the left and right entorhinal cortex volumes estimated by ASHS-T1 (left:  $F = 24.97$ ,  $p < 0.001$ ; right:  $F = 19.47$ ,  $p < 0.001$ ), manual segmentation (left:  $F = 21.22$ ,  $p < 0.001$ ; right:  $F = 16.55$ ,  $p < 0.001$ ), and FreeSurfer (left:  $F = 15.14$ ,  $p < 0.001$ ; right:  $F = 20.56$ ,  $p < 0.001$ ).

Post hoc pairwise comparisons of the ASHS-T1 entorhinal cortex volumes showed that the left and right entorhinal cortex volumes were significantly smaller in the AD and aMCI groups compared to the HC group and in the AD group compared to the aMCI group.

Post hoc pairwise comparisons of the manual segmentation and FreeSurfer entorhinal cortex volumes showed that the left and right entorhinal cortex volumes were significantly smaller in the AD and aMCI groups compared to the HC group but only the left entorhinal cortex volume was significantly smaller in the AD group compared to the aMCI group.

### 3.4.2. Transentorhinal cortex

The one-way ANCOVAs revealed significant differences across the three diagnostic groups in the left and right transentorhinal cortex volumes estimated by ASHS-T1 (left:  $F = 16.08$ ,  $p < 0.001$ ; right:  $F = 17.10$ ,  $p < 0.001$ ) and manual segmentation (left:  $F = 13.73$ ,  $p < 0.001$ ; right:  $F = 7.41$ ,  $p = 0.001$ ) but only in the right transentorhinal cortex volume estimated by FreeSurfer (left:  $F = 1.99$ ,  $p = 0.142$ ; right:  $F = 3.57$ ,  $p = 0.032$ ).

Post hoc pairwise comparisons of the ASHS-T1 transentorhinal cortex volumes showed that the left and right transentorhinal cortex

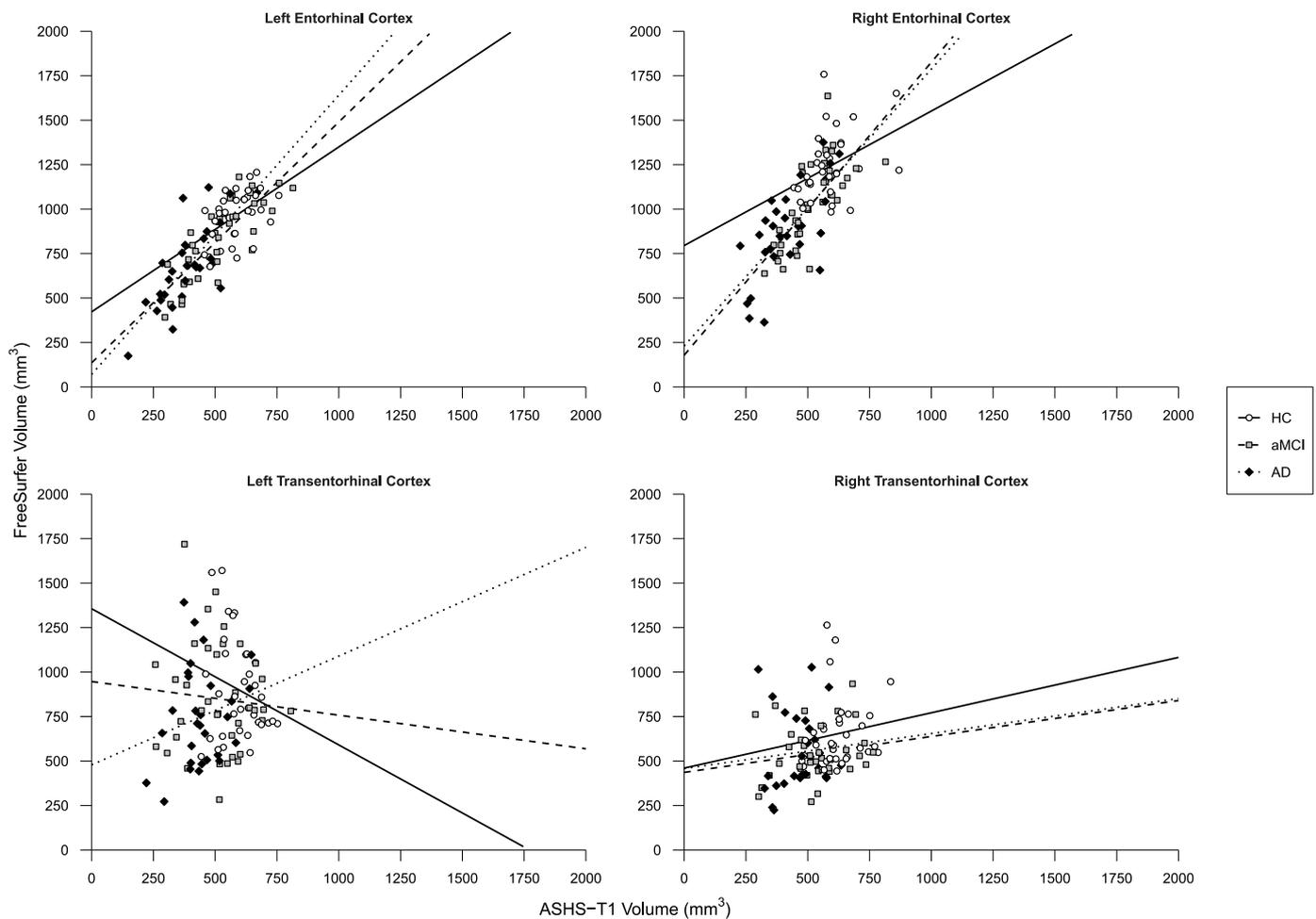


Fig. 3. Scatter plots of ASHS-T1 and FreeSurfer entorhinal and transentorhinal cortical volumes.

volumes were significantly smaller in the AD and aMCI groups compared to the HC group but were not significantly different between the AD and aMCI groups.

Post hoc pairwise comparisons of the manual segmentation transentorhinal cortex volumes showed that only the left transentorhinal cortex volume was significantly smaller in the aMCI group compared to the HC group but both the left and right transentorhinal cortex volumes were significantly smaller in the AD group compared to the aMCI and HC groups.

Post hoc pairwise comparisons of the FreeSurfer transentorhinal cortex volumes showed that only the right transentorhinal cortex volume was significantly smaller in the aMCI group compared to the HC group. Furthermore, there seemed to be some inaccuracy in the measurement of the right transentorhinal cortex volume as the mean right transentorhinal cortex volume for the AD group was slightly larger, though not significantly so, than for the aMCI group.

#### 4. Discussion

The current study has three key, novel findings. First, there was substantial discrepancy between the volumes measured by ASHS-T1 and the volumes measured by manual segmentation and FreeSurfer. Second, there was a trend toward less discrepancy and higher consistency and agreement between ASHS-T1 and manual segmentation entorhinal cortex volumes for individuals with aMCI and individuals with AD dementia than for HCs, suggesting that segmentation performance may be influenced by the presence of AD-related atrophy. Third, both the entorhinal and transentorhinal cortical volumes measured by ASHS-T1

were sensitive to atrophy in individuals with aMCI and individuals with AD dementia, whereas only the FreeSurfer entorhinal cortex volumes were able to reliably detect disease-related atrophy in individuals with aMCI and individuals with AD dementia.

##### 4.1. Comparison between ASHS-T1 and manual segmentation

There was overall variable consistency and low agreement between the ASHS-T1 and manual segmentation entorhinal and transentorhinal cortical volumes. Notably, ASHS-T1 tended to underestimate entorhinal and transentorhinal cortical volumes relative to manual segmentation. The consistently smaller volumes produced by ASHS-T1 may have been a result of differences between the segmentation protocol used to construct the ASHS-T1 atlas set and the segmentation protocol used in the current study. In particular, the segmentation protocol used to construct the ASHS-T1 atlas set defines the anterior boundary of the entorhinal and transentorhinal cortices as 1.3 mm anterior to the hippocampal head (Xie et al., 2019), whereas the segmentation protocol used in the current study commenced segmentation of the entorhinal and transentorhinal cortices 5 mm anterior to the hippocampal head. The smaller volumes produced by ASHS-T1 likely contributed to the low ICC absolute agreement values between ASHS-T1 and manual segmentation (entorhinal cortex = 0.06–0.27, transentorhinal cortex = 0.14–0.37) observed in the current study. The ICC consistency values, which ignores absolute differences between the two sets of measurements, were higher but still suboptimal (entorhinal cortex = 0.43–0.79, transentorhinal cortex = 0.33–0.60). The initial validation study by Xie et al. (2019), which validated ASHS-T1 using the same manual

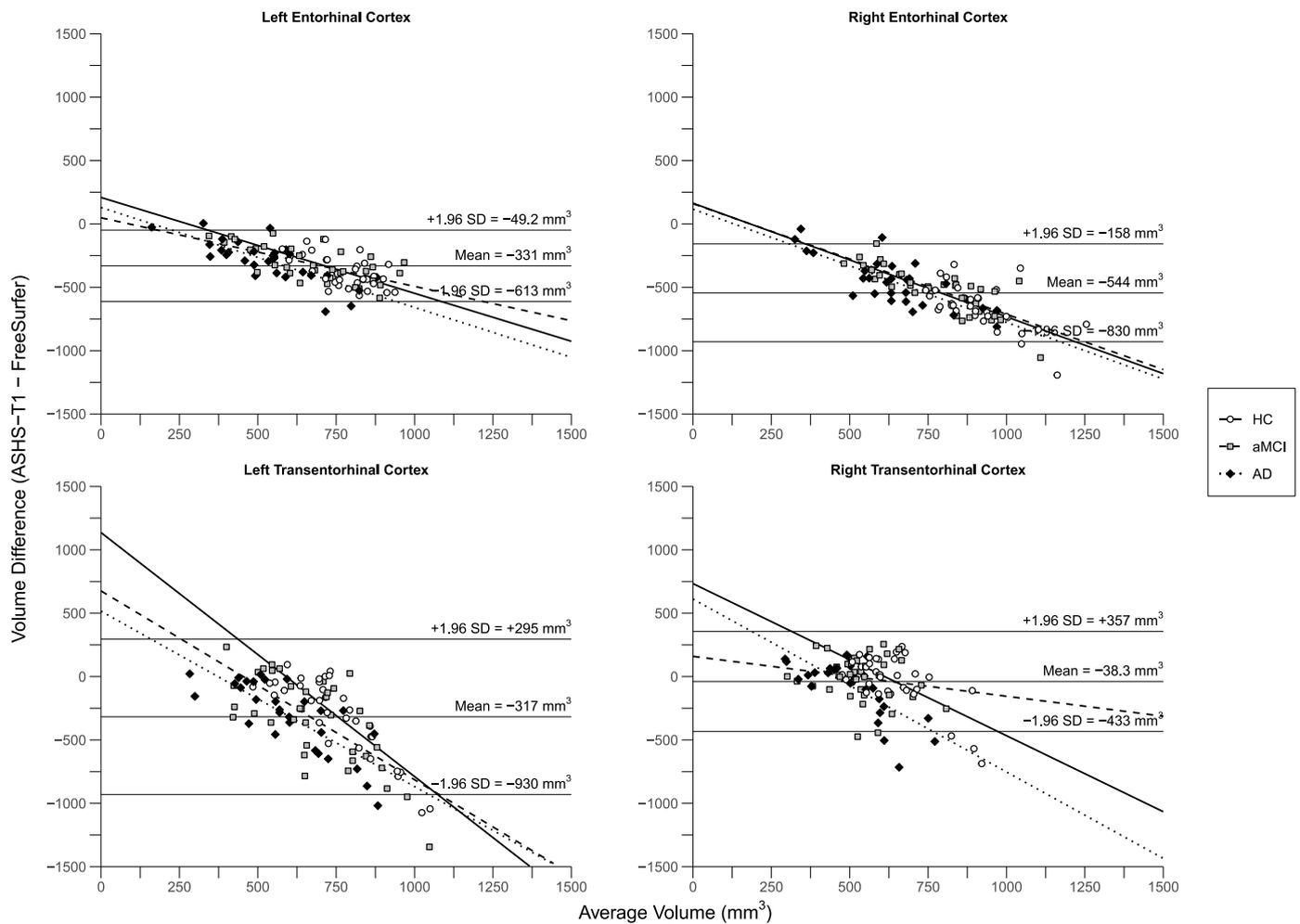


Fig. 4. Bland–Altman plots of ASHS-T1 and FreeSurfer entorhinal and transentorhinal cortical volumes.

segmentation data with which the ASHS-T1 multi-atlas was constructed, reported moderate-to-high ICC values between ASHS-T1 and manual segmentation (viz., entorhinal cortex = 0.69, transentorhinal cortex = 0.77). Whilst validating ASHS-T1 against its own atlas set would likely have resulted in over-optimistic comparison results, the ICC values reported by Xie et al. (2019) were nonetheless below the acceptable reliability threshold of 0.80 (Hopkins, 2000). Overall, the findings of both the current study and that of Xie et al. (2019) highlight less than optimal agreement between ASHS-T1 and manual segmentation entorhinal and transentorhinal cortical volumes.

#### 4.2. Comparison between ASHS-T1 and FreeSurfer

Comparisons between the ASHS-T1 and FreeSurfer entorhinal and transentorhinal cortical volumes showed overall low-to-moderate consistency and low agreement for the entorhinal cortex and low consistency and agreement for the transentorhinal cortex. Moreover, the Bland–Altman plots showed extremely wide 95% limits of agreement for both regions. Xie et al. (2019) compared ASHS-T1 and FreeSurfer volumes in individuals at different stages of AD and showed that FreeSurfer produced on average 37–76% larger entorhinal and transentorhinal cortical volumes compared to ASHS-T1. The current study replicates the previous findings of Xie et al. (2019) of large differences between ASHS-T1 and FreeSurfer volume measurements, demonstrating on average 2–112% larger entorhinal and transentorhinal cortical volumes by FreeSurfer compared to ASHS-T1. The discrepancy between the ASHS-T1 and FreeSurfer volumes is likely due to FreeSurfer mislabeling

the tentorium cerebelli as entorhinal or transentorhinal cortex, resulting in larger entorhinal and transentorhinal cortical volumes (Xie et al., 2019). Overall, these results highlight the substantial variability in volume measurements between different automated segmentation methods. A study comparing three widely used automated segmentation methods, SPM5, FSL, and FreeSurfer, also showed large differences in the volume measurements between the different segmentation methods (Klauschen et al., 2009). Taken together, these findings suggest that different automated segmentation methods may produce very discrepant volume measurements and supports a strong cautionary caveat against comparing volume measurements obtained with different automated methods.

#### 4.3. Influence of atrophy on segmentation performance

There was some evidence that the segmentation performance of ASHS-T1, particularly for the entorhinal cortex, was influenced by the presence of atrophy. The results showed a trend toward less discrepancy and greater consistency and agreement between ASHS-T1 and manual segmentation entorhinal cortex volumes for individuals with aMCI and individuals with AD dementia than for HCs (see Tables 2 and 3). Furthermore, significant proportional bias was observed in the Bland–Altman plots between ASHS-T1 and manual segmentation entorhinal cortex volumes, whereby ASHS-T1 tended to underestimate entorhinal cortex volumes that were larger and the degree of underestimation increased as entorhinal cortex volumes increased (see Fig. 1). Xie et al. (2019) also observed significant proportional bias in

Table 5

Comparisons of entorhinal and transentorhinal cortical volumes ( $\text{mm}^3$ ) estimated by ASHS-T1, manual segmentation, and FreeSurfer, adjusted for intracranial volume, across diagnostic groups.

Method/ Region	Mean Volume (SD)		AD (n = 29)	p Value, Cohen's d (95% CI)		
	HC (n = 34)	aMCI (n = 37)		HC vs. aMCI	HC vs. AD	aMCI vs. AD
<b>ASHS-T1</b>						
Entorhinal Cortex						
Left	587.60 (78.11)	517.31 (131.94)	387.10 (111.33)	<b><i>p</i> = 0.019, <i>d</i> = 0.66 (0.18, 1.14)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.79 (1.23, 2.36)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.12 (0.61, 1.65)</b>
Right	584.05 (94.61)	511.28 (106.27)	415.60 (112.51)	<b><i>p</i> = 0.007, <i>d</i> = 0.74 (0.26, 1.22)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.59 (1.04, 2.15)</b>	<b><i>p</i> = 0.003, <i>d</i> = 0.85 (0.34, 1.36)</b>
Transentorhinal Cortex						
Left	596.28 (78.83)	511.49 (129.90)	441.06 (97.52)	<b><i>p</i> = 0.002, <i>d</i> = 0.83 (0.34, 1.31)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.43 (0.88, 1.97)</b>	<i>p</i> = 0.055, <i>d</i> = 0.60 (0.10, 1.10)
Right	622.00 (89.08)	533.21 (124.36)	467.15 (86.26)	<b><i>p</i> &lt; 0.001, <i>d</i> = 0.89 (0.40, 1.38)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.46 (0.92, 2.01)</b>	<i>p</i> = 0.074, <i>d</i> = 0.57 (0.07, 1.07)
<b>Manual Segmentation</b>						
Entorhinal Cortex						
Left	969.51 (124.54)	838.44 (192.72)	677.16 (190.52)	<b><i>p</i> = 0.005, <i>d</i> = 0.78 (0.29, 1.26)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.66 (1.10, 2.22)</b>	<b><i>p</i> = 0.002, <i>d</i> = 0.88 (0.37, 1.40)</b>
Right	1004.40 (160.98)	850.48 (182.75)	748.29 (177.72)	<b><i>p</i> &lt; 0.001, <i>d</i> = 0.92 (0.43, 1.41)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.43 (0.88, 1.97)</b>	<i>p</i> = 0.131, <i>d</i> = 0.51 (0.01, 1.01)
Transentorhinal Cortex						
Left	687.65 (95.82)	632.99 (100.55)	554.62 (91.11)	<b><i>p</i> = 0.049, <i>d</i> = 0.58 (0.10, 1.06)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.34 (0.79, 1.88)</b>	<b><i>p</i> = 0.010, <i>d</i> = 0.76 (0.25, 1.26)</b>
Right	720.11 (99.36)	689.63 (130.93)	609.14 (87.13)	<i>p</i> = 0.697, <i>d</i> = 0.29 (-0.19, 0.76)	<b><i>p</i> &lt; 0.001, <i>d</i> = 0.96 (0.44, 1.49)</b>	<b><i>p</i> = 0.024, <i>d</i> = 0.68 (0.17, 1.18)</b>
<b>FreeSurfer</b>						
Entorhinal Cortex						
Left	967.03 (142.82)	835.73 (218.70)	678.07 (231.91)	<b><i>p</i> = 0.014, <i>d</i> = 0.69 (0.20, 1.17)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.40 (0.86, 1.95)</b>	<b><i>p</i> = 0.016, <i>d</i> = 0.71 (0.21, 1.22)</b>
Right	1236.76 (185.68)	1020.14 (239.90)	877.07 (254.73)	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.05 (0.55, 1.54)</b>	<b><i>p</i> &lt; 0.001, <i>d</i> = 1.58 (1.03, 2.14)</b>	<i>p</i> = 0.105, <i>d</i> = 0.54 (0.03, 1.04)
Transentorhinal Cortex						
Left	899.71 (284.30)	850.00 (311.00)	747.59 (279.79)	<i>p</i> = 1.00, <i>d</i> = 0.16 (-0.30, 0.64)	<i>p</i> = 0.153, <i>d</i> = 0.50 (-0.01, 1.01)	<i>p</i> = 0.552, <i>d</i> = 0.33 (-0.16, 0.83)
Right	653.71 (197.75)	543.41 (149.72)	549.17 (213.61)	<b><i>p</i> = 0.042, <i>d</i> = 0.59 (0.12, 1.07)</b>	<i>p</i> = 0.136, <i>d</i> = 0.52 (0.01, 1.03)	<i>p</i> = 1.00, <i>d</i> = -0.08 (-0.57, 0.42)

Note: Bold type indicates statistical significance. AD = Alzheimer's disease; aMCI = amnesic mild cognitive impairment; ASHS-T1 = Automatic Segmentation of Hippocampal Subfields-T1; CI = confidence interval; HC = healthy control.

the ASHS-T1 entorhinal cortex segmentations. Visual inspection of the ASHS-T1 segmentations suggested that ASHS-T1 tended to under-segment the entorhinal cortex particularly at its superior border and at the gray matter–white matter border. It is possible that the under-segmentation is exacerbated in larger entorhinal cortices compared to smaller entorhinal cortices, resulting in the observed volume-dependent bias. Other studies have also observed volume- or atrophy-dependent segmentation performance in other automated segmentation methods compared to manual segmentation for other brain regions, including the hippocampus, amygdala, superior temporal gyrus, and temporal lobe (Lehmann et al., 2010; Sánchez-Benavides et al., 2010). These findings suggest that automated segmentation methods may perform differently in healthy brains versus diseased brains, highlighting the importance of validating these segmentation methods across the populations in which they are intended to be used.

#### 4.4. Clinical utility of entorhinal and transentorhinal cortical volumes

The current study additionally assessed the clinical utility of ASHS-T1 by comparing the entorhinal and transentorhinal cortical volumes measured by ASHS-T1 between HCs, individuals with aMCI, and individuals with AD dementia. The entorhinal and transentorhinal cortical volumes were significantly smaller in the aMCI and AD groups compared to the HC group. Only the entorhinal cortex volumes, however, were significantly smaller in individuals with AD dementia compared to individuals with aMCI, whereas the transentorhinal cortex volumes were not significantly different between these two groups. The finding of volume loss in the entorhinal and transentorhinal cortices in individuals with aMCI and individuals with AD dementia is consistent with the findings of Xie et al. (2019), who showed significantly reduced ASHS-T1

entorhinal and transentorhinal cortical volumes in individuals with early prodromal AD, individuals with late prodromal AD, and individuals with AD dementia compared to HCs, and is also in line with previous characterizations of entorhinal and transentorhinal cortical atrophy in individuals with aMCI and individuals with AD dementia (Jessen et al., 2006; Juottonen et al., 1998; Kulason et al., 2019; Penanen et al., 2004; Tward et al., 2017). However, the finding of no difference in transentorhinal cortex volume between individuals with aMCI and individuals with AD dementia is surprising, given that AD-related neuropathological changes within the transentorhinal cortex continue to accumulate as the disease progresses (Braak and Braak, 1991). One possible explanation for such a result is that the current study was not sufficiently powered to detect a difference in transentorhinal cortex volume between the two groups, as suggested by the medium effect size (i.e.,  $d > 0.50$ ) for the comparisons despite statistical non-significance. Taken together, these findings suggest that ASHS-T1 is sufficiently sensitive to entorhinal and transentorhinal cortical volume changes to detect AD-related atrophic changes in both early and late disease stages. Further evaluation of ASHS-T1 in a larger sample is required to clarify its utility in monitoring disease progression in the transentorhinal cortex.

The FreeSurfer entorhinal and transentorhinal cortical volumes showed inconsistent results in detecting differences across the diagnostic groups. Whilst the FreeSurfer entorhinal cortex volumes showed a similar trend of differences across the diagnostic groups as the ASHS-T1 entorhinal cortex volumes, except for the lack of a difference between the aMCI and AD groups, some inconsistency was observed in the FreeSurfer transentorhinal cortex volumes. The left transentorhinal cortex volume was not significantly different between any of the diagnostic groups, whereas the right transentorhinal cortex volume was

significantly different only between the HC and aMCI groups. Moreover, when examining the mean volumes for each diagnostic group, a marginal increase in right transentorhinal cortex volume was observed in the individuals with AD dementia relative to the individuals with aMCI. Xie et al. (2019) found a similar fluctuation in FreeSurfer entorhinal cortex volumes, whereby the mean entorhinal cortex volume was smaller in individuals with preclinical AD compared to HCs but was larger in individuals with early prodromal AD relative to individuals with preclinical AD. These findings may be the result of the erroneous inclusion of portions of the tentorium cerebelli in the FreeSurfer entorhinal or transentorhinal cortex labels (Xie et al., 2019). The fluctuation in entorhinal and transentorhinal cortical volumes across the diagnostic groups suggests that such an issue in the FreeSurfer segmentations is inconsistently applied, which may consequently obscure true group differences in the volumes. Overall, these findings raise questions about the validity of FreeSurfer entorhinal and transentorhinal cortical volumes, especially when employed for detecting clinically meaningful differences across the spectrum of AD severity.

## 5. Limitations

Some potential limitations to the current study should be mentioned. First, there were several differences between the segmentation protocol used in the ASHS-T1 atlas set and the segmentation protocol used in the manual approach in the current study. These differences in the segmentation protocols likely contributed to the differences in the volumes of the entorhinal and transentorhinal cortices, thereby impacting the comparisons between the two segmentation methods. Accordingly, whilst it would be difficult to draw strong conclusions about the segmentation accuracy of ASHS-T1 compared to manual segmentation, the current study nonetheless showed that the volumes of the entorhinal and transentorhinal cortices measured by ASHS-T1 were sensitive to AD-related atrophy in both early and late disease. Second, the results of the current study are based on MRI scans that were obtained in the research setting and may thus not be generalizable to routinely acquired clinical MRI scans. For example, ADNI MRI scans undergo rigorous quality control to ensure high quality data (Jack et al., 2008). Scans acquired in the clinical setting, however, typically do not receive the same standard of quality control due to capacity constraints and are thus more likely to contain image artifacts, which may adversely affect the performance of automated segmentation methods (Reuter et al., 2015; Tisdall et al., 2016). Consequently, it would be important for the findings of the current study to be replicated on MRI scans acquired in routine clinical practice. Third, the current study did not examine cortical thickness measurements of the entorhinal and transentorhinal cortices. It has been previously argued that thickness, rather than volume, measurements of the entorhinal and transentorhinal cortices may be more sensitive to AD-related atrophy in these regions as thickness measurements are more robust to the anatomical variability exhibited by these regions (Feczko et al., 2009; Xie et al., 2019). However, several previous studies, including the initial ASHS-T1 validation study by Xie et al. (2019), did not find significantly superior performance of thickness measurements compared to volume measurements of the entorhinal and transentorhinal cortices in detecting early AD-related atrophy in these regions (Li et al., 2021; Schwarz et al., 2016). On the contrary, Xie et al. (2019) found a significant difference in entorhinal cortex volume, but not thickness, between HCs and individuals with early prodromal AD. Therefore, there is still considerable uncertainty about the value of thickness over volume measurements of the entorhinal and transentorhinal cortices in the early detection of AD-related atrophy.

## 6. Conclusion

In conclusion, despite significant differences between ASHS-T1 and manual segmentation entorhinal and transentorhinal cortical volumes, ASHS-T1 demonstrated potential clinical utility in detecting AD-related

atrophy by significantly discriminating individuals with aMCI and individuals with AD dementia from HCs. Moreover, ASHS-T1 performed more consistently than FreeSurfer in detecting volume differences across the diagnostic groups. Overall, the findings of the current study highlight ASHS-T1 as a promising tool for the automated segmentation of the entorhinal and transentorhinal cortices. Further research should examine the performance of ASHS-T1 on routinely acquired clinical MRI scans to clarify its potential to be translated to the clinical setting.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This work was supported by the Australian Commonwealth Government and the University of Melbourne.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.psychres.2023.111707](https://doi.org/10.1016/j.psychres.2023.111707).

## References

- Augustinack, J.C., Huber, K.E., Stevens, A.A., Roy, M., Frosch, M.P., van der Kouwe, A.J. W., Wald, L.L., Van Leemput, K., McKee, A.C., Fischl, B., 2013. Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. *NeuroImage* 64, 32–42.
- Bartzokis, G., Althuler, L.L., Greider, T., Curran, J., Keen, B., Dixon, W., 1998. Reliability of medial temporal lobe volume measurements using reformatted 3D images. *Psychiatry Res. Neuroimaging* 82, 11–24.
- Baumann, K., 2003. Cross-validation as the objective function for variable-selection techniques. *TrAC Trends Anal. Chem.* 22, 395–406.
- Berron, D., Vieweg, P., Hochkepler, A., Pluta, J.B., Ding, S.-L., Maass, A., Luther, A., Xie, L., Das, S.R., Wolk, D.A., Wolbers, T., Yushkevich, P.A., Düzel, E., Wisse, L.E.M., 2017. A protocol for manual segmentation of medial temporal lobe subregions in 7 Tesla MRI. *NeuroImage Clin.* 15, 466–482.
- Bobinski, M., De Leon, M.J., Wegiel, J., Desanti, S., Convit, A., Saint Louis, L.A., Rusinek, H., Wisniewski, H.M., 1999. The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. *Neuroscience* 95, 721–725.

- Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 82, 239–259.
- Cumming, G., Finch, S., 2005. Inference by eye: Confidence intervals and how to read pictures of data. *Am. Psychol.* 60, 170–180.
- Feczko, E., Augustinack, J.C., Fischl, B., Dickerson, B.C., 2009. An MRI-based method for measuring volume, thickness and surface area of entorhinal, perirhinal, and posterior parahippocampal cortex. *Neurobiol. Aging* 30, 420–431.
- Fischl, B., 2012. FreeSurfer. *NeuroImage* 62, 774–781.
- Fischl, B., Stevens, A.A., Rajendran, N., Yeo, B.T., Greve, D.N., Van Leemput, K., Polimeni, J.R., Kakunoori, S., Buckner, R.L., Pacheco, J., 2009. Predicting the location of entorhinal cortex from MRI. *NeuroImage* 47, 8–17.
- Fung, Y.L., Ng, K.E.T., Vogrin, S.J., Meade, C., Ngo, M., Collins, S.J., Bowden, S.C., 2019. Comparative utility of manual versus automated segmentation of hippocampus and entorhinal cortex volumes in a memory clinic sample. *J. Alzheimers Dis.* 68, 159–171.
- Hopkins, W.G., 2000. Measures of reliability in sports medicine and science. *Sports Med.* 30, 1–15.
- JrL. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J., Ward, C., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691.
- Jessen, F., Feyen, L., Freymann, K., Tepest, R., Maier, W., Heun, R., Schild, H.H., Scheef, L., 2006. Volume reduction of the entorhinal cortex in subjective memory impairment. *Neurobiol. Aging* 27, 1751–1756.
- Juottonen, K., Laakso, M.P., Insausti, R., Lehtovirta, M., Pitkänen, A., Partanen, K., Soinen, H., 1998. Volumes of the entorhinal and perirhinal cortices in Alzheimer's disease. *Neurobiol. Aging* 19, 15–22.
- Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., 2009. Evaluation of automated brain MR image segmentation and volumetry methods. *Hum. Brain Mapp.* 30, 1310–1327.
- Kulason, S., Tward, D.J., Brown, T., Sicut, C.S., Liu, C.-F., Ratnanather, J.T., Younes, L., Bakker, A., Gallagher, M., Albert, M., 2019. Cortical thickness atrophy in the transentorhinal cortex in mild cognitive impairment. *NeuroImage Clin.* 21.
- Lehmann, M., Douiri, A., Kim, L.G., Modat, M., Chan, D., Ourselin, S., Barnes, J., Fox, N. C., 2010. Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *NeuroImage* 49, 2264–2274.
- Li, Q., Wang, J., Liu, J., Wang, Y., Li, K., 2021. Magnetic resonance imaging measurement of entorhinal cortex in the diagnosis and differential diagnosis of mild cognitive impairment and Alzheimer's disease. *Brain Sci.* 11, 1129.
- Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hänninen, T., Laakso, M.P., Hallikainen, M., Vanhanen, M., Nissinen, A., Helkala, E.-L., 2004. Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol. Aging* 25, 303–310.
- Penumetcha, N., Kabadi, S., Jedynek, B., Walcutt, C., Gado, M.H., Wang, L., Ratnanather, J.T., 2011. Feasibility of geometric-intensity-based semi-automated delineation of the tentorium cerebelli from MRI scans. *J. Neuroimaging* 21, e148–e155.
- Jr Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., 2010. Alzheimer's disease neuroimaging initiative (ADNI): Clinical characterization. *Neurology* 74, 201–209.
- R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reuter, M., Tisdall, M.D., Qureshi, A., Buckner, R.L., van der Kouwe, A.J.W., Fischl, B., 2015. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage* 107, 107–115.
- Sánchez-Benavides, G., Gómez-Ansón, B., Sainz, A., Vives, Y., Delfino, M., Peña-Casanova, J., 2010. Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. *Psychiatry Res. Neuroimaging* 181, 219–225.
- Schwarz, C.G., Gunter, J.L., Wiste, H.J., Przybelski, S.A., Weigand, S.D., Ward, C.P., Senjem, M.L., Vemuri, P., Murray, M.E., Dickson, D.W., 2016. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage Clin.* 11, 802–812.
- Stoub, T.R., Rogalski, E.J., Leurgans, S., Bennett, D.A., deToledo-Morrell, L., 2010. Rate of entorhinal and hippocampal atrophy in incipient and mild AD: Relation to memory function. *Neurobiol. Aging* 31, 1089–1098.
- Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., Pihlajamäki, M., Laakso, M.P., Hallikainen, M., Hämmäläinen, A., 2008. MRI of hippocampus and entorhinal cortex in mild cognitive impairment: A follow-up study. *Neurobiol. Aging* 29, 31–38.
- Tisdall, M.D., Reuter, M., Qureshi, A., Buckner, R.L., Fischl, B., van der Kouwe, A.J.W., 2016. Prospective motion correction with volumetric navigators (vNavs) reduces the bias and variance in brain morphometry induced by subject motion. *NeuroImage* 127, 11–22.
- Tward, D.J., Sicut, C.S., Brown, T., Bakker, A., Gallagher, M., Albert, M., Miller, M., 2017. Entorhinal and transentorhinal atrophy in mild cognitive impairment using longitudinal diffeomorphometry. *Alzheimer's Dement. Assess. Dis. Monit.* 9, 41–50.
- Venneri, A., Gorgoglione, G., Toraci, C., Nocetti, L., Panzetti, P., Nichelli, P., 2011. Combining neuropsychological and structural neuroimaging indicators of conversion to Alzheimer's disease in amnesic mild cognitive impairment. *Curr. Alzheimer Res.* 8, 789–797.
- Xie, L., Das, S.R., Wisse, L.E.M., Ittyerah, R., Yushkevich, P.A., Wolk, D.A., 2018. Early tau burden correlates with higher rate of atrophy in transentorhinal cortex. *J. Alzheimers Dis.* 62, 85–92.
- Xie, L., Wisse, L.E.M., Das, S.R., Wang, H., Wolk, D.A., Manjón, J.V., Yushkevich, P.A., 2016. Accounting for the confound of meninges in segmenting entorhinal and perirhinal cortices in T1-weighted MRI. In: Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G., Wells, W. (Eds.), *Lecture notes in computer science: Vol 9901. Medical image computing and computer-assisted intervention*. Springer, pp. 564–571.
- Xie, L., Wisse, L.E.M., Pluta, J., de Flores, R., Piskin, V., Manjón, J.V., Wang, H., Das, S. R., Ding, S.L., Wolk, D.A., 2019. Automated segmentation of medial temporal lobe subregions on in vivo T1-weighted MRI in early stages of Alzheimer's disease. *Hum. Brain Mapp.* 40, 3431–3451.
- Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* 31, 1116–1128.
- Yushkevich, P.A., Pluta, J.B., Wang, H., Xie, L., Ding, S.L., Gertje, E.C., Mancuso, L., Kliot, D., Das, S.R., Wolk, D.A., 2015. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum. Brain Mapp.* 36, 258–287.