Effects of Age and Alzheimer's Disease on Hippocampal Subfields: Comparison Between Manual and FreeSurfer Volumetry

Robin de Flores,^{1,2,3,4} Renaud La Joie,^{1,2,3,4} Brigitte Landeau,^{1,2,3,4}
Audrey Perrotin,^{1,2,3,4} Florence Mézenge,^{1,2,3,4} Vincent de La Sayette,^{1,2,3,5}
Francis Eustache,^{1,2,3,4} Béatrice Desgranges,^{1,2,3,4} and Gaël Chételat^{1,2,3,4,*}

¹INSERM, U1077, Caen, France

²Université de Caen Basse-Normandie, UMR-S1077, Caen, France

³Ecole Pratique des Hautes Etudes, UMR-S1077, Caen, France

⁴CHU de Caen, U1077, Caen, France

⁵CHU de Caen, Service de Neurologie, Caen, France

Abstract: Growing interest has developed in hippocampal subfield volumetry over the past few years and an increasing number of studies use the automatic segmentation algorithm implemented in FreeSurfer. However, this approach has not been validated on standard resolution T1-weighted magnetic resonance (MR) as used in most studies. We aimed at comparing hippocampal subfield segmentation using FreeSurfer on standard T1-weighted images versus manual delineation on dedicated high-resolution hippocampal scans. Hippocampal subfields were segmented in 133 individuals including 98 cognitively normal controls aged 19–84 years, 17 mild cognitive impairment and 18 Alzheimer's disease (AD) patients using both methods. Intraclass correlation coefficients (ICC) and Bland–Altman plots were computed to assess the consistency between both methods, and the effects of age and diagnosis were assessed from both measures. Low to moderate ICC (0.31–0.74) were found for the subiculum and other subfields as well as for the whole hippocampus, and the correlations were very low for cornu ammonis (CA)1 (<0.1). FreeSurfer CA1 volume estimates were found to be much lower than those obtained from manual segmentation, and this bias was proportional to the volume of this structure so that no effect of age or AD could be detected on FreeSurfer CA1 volumes. This study points to the differences in the anatomic definition of the subfields between FreeSurfer and manual delineation, especially for CA1, and provides clue for improvement of this automatic technique for potential clinical application on standard T1-weighted MR. *Hum Brain Mapp* 36:463–474, 2015. © 2014 Wiley Periodicals, Inc.

Key words: hippocampal subfields; aging; Alzheimer's disease; manual segmentation; FreeSurfer segmentation; magnetic resonance imaging

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

Contract grant sponsors: Association France Alzheimer, Fondation Plan Alzheimer (Alzheimer Plan 2008–2012), Programme Hospitalier de Recherche Clinique (PHRC National 2011), Agence Nationale de la Recherche (ANR LONGVIE 2007), Région Basse Normandie, and Institut National de la Santé et de la Recherche Médicale (INSERM).

*Correspondence to: Gaël Chételat; Inserm—EPHE—Université de Caen/Basse-Normandie, Unité U1077, GIP Cyceron, Bd H Becquerel, 14074 Caen Cedex, France. E-mail: chetelat@cyceron.fr

Received for publication 22 April 2014; Revised 26 August 2014; Accepted 9 September 2014.

DOI: 10.1002/hbm.22640

Published online 18 September 2014 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

The hippocampus has been the focus of extensive research over the last decades, especially with the development and improvement of neuroimaging techniques. The particular interest of neuroscientists for this structure arises from its implication in cognitive processes, especially episodic memory [Lepage et al., 1998; Squire et al., 1992; Tulving and Markowitsch, 1998; see Spaniol et al., 2009 for review] and spatial navigation [Burgess et al., 2002; Ekstrom et al., 2003; Maguire et al., 1998; see Bird and Burgess, 2008 for review], as well as its structural alteration in several neurological and psychiatric disorders, such as Alzheimer's disease (AD), temporal lobe epilepsy, schizophrenia, post-traumatic stress disorder, and major depression [see Geuze et al., 2005 for review].

The hippocampus is composed of several histologically defined and interconnected subfields, including the Dentate Gyrus (DG), the four fields of the cornu ammonis (CA1–4) and the subiculum [Duvernoy, 2005]. Recent advances in neuroimaging have stressed the relevance to consider these different substructures instead of assessing the hippocampus as a whole. Over normal aging, for example, studies agree on the fact that age exerts a differential effect on hippocampal subfield volumes. The subiculum (SUB) seems to be the earliest affected in aging, although results differ across studies, probably reflecting the use of different methodologies [Chételat et al., 2008; Frisoni et al., 2008; La Joie et al., 2010; Mueller et al., 2007; Wang et al., 2003; Wisse et al., 2014; Ziegler et al., 2011].

Interestingly, hippocampal subfields also show differential vulnerability to neurological disorders and particularly to AD, with CA1 showing greatest and earliest sensitivity in both neuropathological [Rössler et al., 2002; West et al., 1994] and neuroimaging [Apostolova et al., 2010; Chételat et al., 2008; Csernansky et al., 2005; Frisoni et al., 2008; Kerchner et al., 2013; Wang et al., 2006] studies. Thus, hippocampal subfield, and more specifically CA1 volumetry, may be more accurate than global hippocampal volumetry to detect AD especially in early stages, for example, in mild cognitive impairment (MCI) [Apostolova et al., 2006; Chételat et al., 2008; La Joie et al., 2013; Mueller et al., 2010; Pluta et al., 2012].

Both to facilitate research development in this area and for clinical application, validated automatic algorithms of hippocampal subfield segmentation are urgently needed. Several automatic algorithms have been developed [Van Leemput et al., 2009; Yushkevich et al., 2010], the most widely used being that implemented in FreeSurfer [Van Leemput et al., 2009]. FreeSurfer is free and easy-to-use software, allowing automatic segmentation of hippocampal subfields from T1-weighted magnetic resonance (MR). This method has been validated on ultrahigh resolution T1 images acquired in healthy subjects [Van Leemput et al., 2009] while it is mostly applied to standard resolution T1-weighted MR [Durazzo et al., 2013; Hanseeuw et al., 2011;

Lim et al., 2012a, b, 2013; Pereira et al., 2013, 2014; Teicher et al., 2012; see also Table I]. The segmentation of such small structures on standard resolution T1-weighted MR appears particularly challenging as the limited spatial resolution of the images may compromise the accurate delineation of subfield boundaries, and especially given the rather large number (7) of hippocampal substructures segmented by FreeSurfer. Moreover, the hippocampal subfield segmentation obtained with FreeSurfer has not been validated so far on clinical populations of MCI or AD patients or on lower field strengths. Besides, several recent studies using FreeSurfer in MCI and AD failed to report the well-documented CA1 atrophy [Hanseeuw et al., 2011; Lim et al., 2012a, 2013], which leads some doubts on the validity of this automatic segmentation approach applied on standard T1-weighted images, at least in early AD.

The aim of this study was therefore to compare hippocampal subfield volumes obtained using FreeSurfer on standard T1-weighted MR to those obtained by manual delineation on dedicated high-resolution hippocampal scan in a sample of 133 individuals including cognitively normal adults, MCI and AD patients. A secondary objective was to further our understanding of the effect of normal aging, MCI and AD on hippocampal subfield volumes.

MATERIAL AND METHODS

Participants

All participants were included in the Imagerie Multimodale de la maladie d'Alzheimer à un stade Précoce (IMAP) study (Caen, France) and part of them were included in previous publications from our lab [Arenaza-Urquijo et al., 2013; La Joie et al., 2010, 2012, 2013; Mevel et al., 2013]. They were all right-handed, had at least 7 years of education and had no history of alcoholism, drug abuse, head trauma or psychiatric disorder. The IMAP Study was approved by a regional ethics committee (Comité de Protection des Personnes Nord-Ouest III) and is registered with ClinicalTrials.gov (number NCT01638949). All participants gave written informed consent to the study prior to the investigation.

Patients were recruited from local memory clinics and selected according to corresponding internationally agreed criteria. Clinical diagnosis was assigned by consensus under the supervision of a senior neurologist (VdIS) and neuropsychologists. Briefly, 18 patients with AD fulfilling standard National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) clinical criteria for probable AD [McKhann et al., 1984] were included. Seventeen MCI patients were also recruited from local memory clinics and selected according to Petersen's criteria for MCI [Petersen and Morris, 2005]. Finally, 98 healthy individuals (aged 19–84) were recruited from the community by flyers and advertisements in local

TABLE I. MRI studies assessing hippocampal subfield volumes

Study	Field (T)	Sequence	Resolution before interpolation (mm ³)	Segmentation	Population number (mean age ± SD) (n)	Ratio CA1/HCP (%)
Mueller et al. (2010)	4	T2, FSE	0.4 × 0.4 × 2.0	Manual	53 (69.3 ± 7.3)	42.77
Wisse et al. (2012)	7	T2, 3D FSE	0.7 × 0.7 × 0.7	Manual	14 (63.2 ± 7.8)	46.09
Alder et al. (2013)	9.4	T2, SEMS	0.2 × 0.2 × 0.2	Manual	1 (89)	50.33
Winterburn et al. (2013)	3	T1, FSGRE T2, FSE	0.57 × 0.57 × 0.6	Manual	5 (37)	31.08
Wisse et al. (2014)	7	T2, 3D FSE	0.7 × 0.7 × 0.7	Manual	29 (69.5 ± 7.3)	43.21
Present study	3	PD	0.375 × 0.375 × 2.0	Manual	98 (45.7 ± 19.2)	32.03
Hanseeuw et al. (2011)	3	T1, GRE	0.81 × 0.95 × 1.0	FreeSurfer	15 (69.4 ± 4.8)	9.80
Lim et al. (2012)	3	T1, 3D MPRAGE	1.0 × 1.0 × 1.0	FreeSurfer	30 (72.4 ± 4.5)	8.82
Lim et al. (2012)	3	T1, 3D MPRAGE	1.0 × 1.0 × 1.0	FreeSurfer	49 (72.4 ± 4.5)	11.59
Lim et al. (2013)	3	T1, 3D MPRAGE	0.8 × 0.8 × 0.8	FreeSurfer	33 (75.6 ± 4.2)	8.62
Pereira et al. (2014)	3	T1	1.0 × 1.0 × 1.0	FreeSurfer	50 (63.7 ± 7)	10.11
Present study	3	T1, FFE	1.0 × 1.0 × 1.0	FreeSurfer	98 (45.7 ± 19.2)	13.48

HCP, Total hippocampus.

newspapers. They performed in the normal range on all neuropsychological tests from a cognitive battery assessing multiple domains of cognition (verbal and visual episodic memory, semantic memory, language skills, executive functions, visuospatial functions, and praxis). A subgroup of 30 healthy elderly from this 98-individual sample was used as age-, gender- and education-matched controls for comparison to MCI and AD patients (for demographic information of the samples see Table II).

MR Data Acquisition

Each subject underwent an MR scan at the CYCERON center (Caen, France) using a 3T Philips (Eindhoven, The Netherlands) scanner. First, T1-weighted structural images were acquired (Repetition Time (TR) = 20 ms; Echo Time (TE) = 4.6 ms; flip angle = 10°; 180 slices; slice thickness = 1 mm; no gap; Field of View (FoV) = 256 × 256 mm²; matrix = 256 × 256; in-plane resolution = 1 × 1 mm²; acquisition time = 9.4 min). Then, a high resolution proton density weighted sequence was acquired perpendicularly to the long axis of the hippocampus (TR = 3,500 ms; TE = 19 ms; flip angle = 90°; 13 slices; slice thickness = 2 mm; interslices gap = 2 mm; in-plane reso-

lution = 0.375 × 0.375 mm², acquisition time = 7.4 min). Note that the 2 mm gap was chosen as the best compromise to limit the scanning time and the related motion artifacts; this is particularly important in elderly clinical populations.

Neuroimaging Data Processing

Manual delineation of hippocampal subfields

Hippocampal subfields were manually segmented on the high resolution proton density weighted scan according to a protocol detailed in a previous publication [La Joie et al., 2010]. Briefly, three hippocampal regions were delineated: (i) the SUB_{manual}; (ii) CA1; (CA1_{manual}), and (iii) CA2–CA3–CA4 and DG pooled together in a unique region termed “OTHER_{manual}” in what follows. Indeed, the very limited size of CA2, CA3, and CA4 with CA4 surrounded by DG makes difficult the accurate and reliable delineation of each individual subfield [La Joie et al., 2010]. Manual delineations were all performed by the same rater blind to the identity of the participants [RLJ], the same as in La Joie et al., 2010, 2013]. High intrarater reliability was reported with this method [intraclass correlation coefficients (ICC) = 0.94, 0.89, and 0.96 for CA1, SUB

TABLE II. Demographics of the samples

	Healthy individuals	Healthy elderly ^a	MCI	AD
Number (women/men)	98 (57/41)	30 (13/17)	17 (9/8)	18 (12/6)
Age (years; mean ± SD)	45.7 ± 19.2	70 ± 6.4	71.7 ± 6	67.4 ± 9.9
Education (years; mean ± SD)	13.5 ± 3.5	12.2 ± 4	10.5 ± 3.3	10.7 ± 3.8
MMSE (mean ± SD)	29.4 ± 0.8	29.4 ± 0.7	27.2 ± 1.3*	21.4 ± 3.9*

SD, standard deviation.

MCI and AD patients were compared for age, education, and MMSE with healthy elderly using 2-sample t-tests and compared for gender with a chi-square test. *P < 0.05.

^asubsample from the group of healthy individuals.

and other, respectively [La Joie et al., 2010]. To further assess the inter-rater reliability in this study, subfield manual delineations were also performed by an independent trained rater (RDF) on a subsample of 30 individuals, and high ICC values were also found between the two raters (0.82, 0.8, and 0.9). The volume of the whole hippocampus corresponded to the sum of the volumes of the three resulting hippocampal regions.

Automated delineation of hippocampal subfields

Hippocampal subfields were automatically segmented on the T1-weighted scans using FreeSurfer 5.1.0. This method was applied to T1-weighted images with 1 mm isotropic resolution to be in the same condition as most studies using FreeSurfer [Durazzo et al., 2013; Hanseeuw et al., 2011; Lim et al., 2012a, b, 2013; Pereira et al., 2013, 2014; Teicher et al., 2012; see also Table I]. Seven subregions were obtained: CA1, CA2–3, CA4-DG, pre-SUB, SUB, fimbria, and hippocampal fissure. All subregions of all participants were visually checked to detect visible errors in segmentation (no errors were detected). To compare manual and FreeSurfer delineations, FreeSurfer SUB and pre-SUB were combined in a single region called “SUB_{FreeSurfer}” to correspond to the “SUB_{manual}” ROI, and FreeSurfer CA2–3 and CA4-DG were combined in a single region called “OTHER_{FreeSurfer}” to correspond to the “OTHER_{manual}” ROI (Fig. 1). The whole hippocampus volume was obtained by adding all hippocampal subfields.

Total intracranial volume

Individual total intracranial volume (TIV) values were obtained from the T1-weighted images using the VBM5 toolbox implemented in the Statistical Parametric Mapping software (SPM5; Wellcome Trust Center for Neuroimaging, Institute of Neurology, London, England) by summing the volumes of the grey matter, white matter and cerebrospinal fluid.

Statistical Analyses

To evaluate the consistency between the two methods, ICCs were calculated for each ROI and within each group with a two-way random ANOVA model with absolute agreement. Then, the agreement between the two methods was assessed using Bland–Altman plots (see Fig. 2 caption for details). Both ICC and Bland–Altman plots were performed for each ROI on raw volumes in mm³ within the healthy individuals, MCI and AD groups separately.

Then, to evaluate the effects of age, MCI and AD on hippocampal subfields, raw volumetric measures normalized to the TIV were used to compensate for interindividual variability in head size. The effects of age, MCI, and AD were assessed separately for manual and FreeSurfer delineations. The effect of age was assessed using both linear and quadratic regressions between TIV-normalized vol-

umes and age for each subfield. The model with the best-fitting robustness was selected for display. The effect of MCI and AD was assessed with repeated measure analyses of covariance (ANCOVAs) with TIV-normalized volumes as dependent variables and group (healthy elderly, MCI, and AD), age, gender, and education as independent variables. When the subfield $\times$ group interaction was significant, post hoc analyses with Fisher’s LSD were conducted.

All data were analyzed using Statistica 10. Note that, for the sake of simplicity, only the results on the pooled right and left volumes will be presented as the results were unchanged when considering the right and left volumes separately.

RESULTS

Direct Comparison of FreeSurfer Versus Manual Segmentations

As displayed in Table III, ICCs between FreeSurfer and manual measurements were low to moderate. The highest correlations were observed for the SUB region and the whole hippocampus. For the OTHER region, the ICC was higher for the MCI than for the healthy individuals and AD patients. Finally ICCs were close to zero in all groups for CA1.

The Bland–Altman plots showed that FreeSurfer volumes were larger than manual volumes for SUB, OTHER and whole hippocampus within all groups (e.g., 224 mm³, 736 mm³, and 745 mm³ larger, respectively, within the healthy individuals). On the contrary, FreeSurfer CA1 estimates were much smaller than those obtained from manual segmentation in all groups (e.g., 1239 mm³ smaller within the healthy individuals; Fig. 2). All these differences were significant (within the healthy individuals: $P < 0.001$ for SUB, OTHER, whole hippocampus and CA1). Noteworthy, the differences between manual and FreeSurfer volumes highly correlated with the average volume for CA1, ($r^2 = 0.79$; $P < 0.001$) while the correlation was much weaker for SUB ($r^2 = 0.11$; $P < 0.001$) and nonsignificant for OTHER and for the whole hippocampus ($r^2 = 0.02$; $P = 0.1$ and $r^2 = 4.10^{-4}$; $P = 0.8$ respectively). In other words, the bias was proportional to the volume for CA1 (and to a lesser degree for SUB) which is thus expected to impact on detection of age- and AD-related volume changes. For instance for CA1, larger volumes are more under-estimated than smaller volumes, which compromise detection of atrophy.

Effects of Age and Disease

The results of the effects of age on hippocampal subfield volumes delineated manually are detailed in Table IV and shown in Figure 3. For CA1_{manual}, the model with the best fit was a quadratic regression, showing an effect of age from about 50 years (inflection point of the curve). The strongest effect of age was observed on the SUB_{manual},

Figure 1.

Manual (left) and FreeSurfer (right) delineation on coronal slices. For the sake of comparison to manual delineation, CA2–3 and CA4-DG were merged into a single so-called “OTHER_{FreeSurfer}” region to correspond to the manually delineated “OTHER_{manual}” ROI, and the pre-SUB and SUB were merged into a single “SUB_{FreeSurfer}” region to correspond to the manually delineated “SUB_{manual}” ROI. Note that the major difference between the two

segmentation protocols seems to be the location of the CA1/SUB border. The border is very lateral in FreeSurfer’s segmentation (accounting for the rather small size of the CA1 subfield) whereas according to our method (based on Harding et al. [1998] and Duvernoy 2005), the border progressively moves medially as we go from anterior to posterior slices. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

with only the linear model being significant, while no model was significant for the OTHER_{manual} ROI. Results remained the same when correcting for both gender and education (data not shown).

For the FreeSurfer delineation, no models were significant for CA1_{FreeSurfer}. For SUB_{FreeSurfer} and OTHER_{FreeSurfer}, the quadratic model presented the best fit (see Table IV).

The effects of MCI and AD on manually delineated hippocampal subfields are presented in Figure 4. The

repeated measure ANCOVA revealed a significant subfield $\times$ group interaction ($P = < 0.001$). The post hoc analysis showed a significant decrease in CA1_{manual} volume for MCI patients ($P < 0.001$; mean volume loss = -20%) and AD patients ($P < 0.001$; mean volume loss = -25%), with no significant difference between both patient groups. For the SUB_{manual}, a significant decrease was also found for MCI patients ($P = < 0.01$; mean volume loss = -15%) and AD patients ($P = < 0.001$; mean volume

Figure 2.

Bland–Altman plots of raw volumes (in mm^3) of bilateral hippocampal subfields and whole hippocampus in healthy individuals. The x-axis represents the average of the two measurements and the y-axis represents the difference between the two measurements (FreeSurfer—Manual). The black line indicates the origin 0 of the y-axis, the blue line represents the mean of the differences between the two methods (i.e., the bias) and the red lines

represent the 95% confidence intervals. A blue line below 0 (i.e., below the black line) indicates that FreeSurfer estimates are smaller than manual estimates while the reverse (a blue line above the black line) indicates that FreeSurfer estimates are larger than manual estimates. Similar results were obtained in MCI and AD patients. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

loss = −25%), with a trend for a difference between both patient groups ($P = 0.055$). The $\text{OTHER}_{\text{manual}}$ region was only significantly decreased in AD patients ($P < 0.001$; mean volume loss = −15%) while there was no difference in MCI patients ($P = 0.8$) compared to healthy elderly (for mean volume differences with corresponding 95% CI, see Table V).

The effects of MCI and AD on the hippocampal subfields delineated with FreeSurfer are presented in Figure 5. The repeated measure ANCOVA revealed a significant subfield $\times$ group interaction ($P < 0.001$). The post hoc analysis showed no significant difference between either MCI or AD versus healthy elderly in $\text{CA1}_{\text{FreeSurfer}}$ volume. For both $\text{SUB}_{\text{FreeSurfer}}$ and $\text{OTHER}_{\text{FreeSurfer}}$, a significant decrease was found in both MCI and AD patients compared to controls, with a significant difference between both patient groups (MCI > AD) (for mean volume differences with corresponding 95% CI, see Table V).

Supplementary Analysis

Based on visual inspection of the segmentation, FreeSurfer's delineation of the SUB seems to include a portion of CA1, compared with atlases [Duvernoy, 2005; Harding

TABLE III. ICC between manual and FreeSurfer delineation

	Healthy individuals ($n = 98$)	MCI ($n = 17$)	AD ($n = 18$)
CA1	0.02	0.07	0.05
SUB	0.51	0.59	0.58
OTHER (CA2/3/4/DG)	0.39	0.62	0.31
Whole Hippocampus	0.67	0.74	0.64

TABLE IV. Linear and quadratic regressions between TIV-normalized hippocampal subfield volumes and age

	Linear regression		Quadratic regression		
	r^2	age	r^2	age	age ²
Manual					
CA1 _{manual}	0.17	$P < 0.001$	0.25	$P = < 0.05$	$P = < 0.01$
SUB _{manual}	0.35	$P < 0.001$	0.35	$P = 0.9$	$P = 0.2$
OTHER _{manual}	0.03	$P = 0.07$	0.05	$P = 0.4$	$P = 0.2$
FreeSurfer					
CA1 _{FreeSurfer}	3.10^{-4}	$P = 0.8$	0.03	$P = 0.1$	$P = 0.1$
SUB _{FreeSurfer}	6.10^{-2}	$P = 0.01$	0.20	$P < 0.001$	$P < 0.001$
OTHER _{FreeSurfer}	1.10^{-2}	$P = 0.3$	0.13	$P < 0.001$	$P < 0.001$

OTHER: CA2/3/4/DG

et al., 1998] (Fig. 1). Interestingly, these observations were also highlighted in Lim et al. (2012). This may explain the fact that the FreeSurfer CA1 volume was not correlated with that obtained with manual delineation and failed to show an effect of MCI or AD. To confirm this view, all analyses have been repeated by considering CA1 and the SUB delineated with FreeSurfer as a single region, to be compared to the manually delineated CA1 region.

Much higher ICCs were obtained for CA1 when pooling CA1 and SUB volumes from FreeSurfer (0.57 for healthy individuals, 0.72 for MCI and 0.8 for AD) than when using only CA1 FreeSurfer measure (see above). Also, the biases calculated from the Bland–Altman plots were much lower when considering the FreeSurfer CA1–SUB pooled volumes than when using only CA1 FreeSurfer measurement (98.12 mm^3 vs. 1239 mm^3 in the healthy elderly, P value of the difference < 0.001). Moreover, a significant effect of age was detected on the FreeSurfer CA1–SUB pooled volumes, the model with the best fit being a quadratic regression ($r^2=0.12$; age: $P = < 0.01$; age²: $P = < 0.01$). Also, a significant effect of MCI and AD was found on the FreeSurfer CA1–SUB pooled volumes: the repeated measure ANCOVA revealed a significant subfield $\times$ group interac-

tion ($P < 0.001$), and the post hoc analysis showed a significant decrease in CA1–SUB pooled volumes for MCI patients ($P < 0.05$) and AD patients ($P < 0.001$), with a significant difference between both patient groups ($P < 0.05$).

DISCUSSION

In this study, we provided a comparison between hippocampal subfield volumes obtained by manual delineation versus automatic segmentation with FreeSurfer in the same samples of 133 individuals including cognitively normal subjects (aged 19–84), MCI and AD patients, which represents the largest sample to date (see Table I). Our results revealed overall moderate ICCs between both methods for the SUB and CA2–3–4–DG subfields pooled together (OTHER) as well as for the whole hippocampus. Moreover, although subtle differences were found between both methods, the effects of MCI and AD on the SUB, OTHER, and whole hippocampus were also overall comparable. However, significantly larger estimates were obtained for FreeSurfer than for manual delineation. These results are consistent with previous studies that only assessed the whole hippocampus but also reported larger FreeSurfer estimates compared to manual delineation [Cherbuin et al., 2009; Tae et al., 2008; Wenger et al., 2014] and they suggest that, despite the reasonable correlations between FreeSurfer and manual measurements for these subfields, the two methods poorly overlap (see also below).

As for CA1, manual versus FreeSurfer estimates did not correlate. Moreover, Bland–Altman plots show much smaller volume estimates for FreeSurfer for all groups and this difference increases with the volume of the structure so that this bias significantly impacts detection of atrophy with age or in MCI and AD. This view is confirmed by the fact that the relative size of CA1 (ie., the ratio between the volumes of CA1 and whole hippocampus) reported in previous studies using manual versus FreeSurfer segmentations was clearly smaller for the later (see Table I). Finally, as also reported in most [Hanseeuw et al., 2011; Lim et al., 2012a, 2013] though not all [Li et al., 2013] previous

Figure 3.

Scatterplots of the bilateral TIV-normalized volume of manually delineated hippocampal subfields against age. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 4.

Mean TIV-normalized volume of manually delineated bilateral hippocampal subfields in healthy elderly (white), MCI (grey), and AD (black). Error bars represent SEM = SD/ $\sqrt{n}$. t, trend; ** $P < 0.01$; *** $P < 0.001$.

studies using FreeSurfer, no effect of MCI or AD was found on FreeSurfer CA1 estimates, while these effects were highly significant on manual estimates [La Joie et al., 2013; Mueller et al., 2010; Wisse et al., 2014]. The lack of agreement between both methods may reflect a lack of sensitivity of FreeSurfer delineation based on classical T1-weighted MR, but may also likely reflects differences in the anatomic definition of the subfields between the two methods. Note that automatic versus manual delineations were obtained on different images so that it was not possible to compare directly their spatial overlap (using Dice coefficients for example) without introducing interpolation-related bias. However, when compared visually, differences between both delineations were easily detected. First, as noted above, it appeared that a signifi-

cant portion of CA1 is included in the SUB subfield when using FreeSurfer. This was confirmed by our supplementary analyses showing stronger correlations and much lower biases between CA1 manual and FreeSurfer estimates when pooling CA1 and SUB FreeSurfer segments together. Also, the effects of age, MCI and AD were closer to those observed using the manual delineation when pooling the FreeSurfer CA1-SUB volumes together. Another difference between both delineation protocols concerns the segmentation of the hippocampal head. While we started delineating the OTHER subfield only several mm after the most anterior pole of the hippocampus, in accordance with the atlases we used [Duvernoy, 2005; Harding et al., 1998], the same scheme is applied all along the hippocampus axis in the FreeSurfer protocol, with all

TABLE V. Between group differences in subfield volumes

	Healthy elderly versus MCI Mean difference (95% CI)	Healthy elderly versus AD Mean difference (95% CI)	MCI versus AD Mean difference (95% CI)
Manual			
CA1 _{manual}	0.21 (0.11; 0.3)***	0.25 (0.16; 0.35)***	0.05 (−0.06; 0.16)
SUB _{manual}	0.13 (0.04; 0.23)**	0.23 (0.14; 0.33)***	0.1 (−0.03; 0.21) ^t
OTHER _{manual}	−0.01 (−0.11; 0.08)	0.17 (0.08; 0.26)***	0.18 (0.08; 0.29)***
FreeSurfer			
CA1 _{FreeSurfer}	0.03 (−0.05; 0.1)	0.05 (−0.02; 0.12)	0.02 (−0.06; 0.11)
SUB _{FreeSurfer}	0.12 (0.05; 0.2)**	0.25 (0.18; 0.33)***	0.14 (0.05; 0.22)**
OTHER _{FreeSurfer}	0.15 (0.08; 0.23)***	0.33 (0.26; 0.41)***	0.18 (0.1; 0.27)***

Difference in subfield TIV-normalized volume between each pair of groups (mean and 95% confidence interval CI) and corresponding posthoc LSD test statistics (repeated measure ANCOVAs adjusted for age, sex and education); t, trend; ** $P < 0.01$ *** $P < 0.001$; OTHER: CA2/3/4/DG.

Figure 5.

Mean TIV-normalized volumes of the bilateral hippocampal subfields delineated with FreeSurfer in healthy elderly (white), MCI (grey), and AD (black). Error bars represent $SEM = SD/\sqrt{n}$. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

subfields being segmented from the most anterior pole of the hippocampus. As a result, a large part of CA1 is counted as OTHER in the hippocampal head using FreeSurfer. This is likely to have a significant impact on the assessment of the effect of AD as they are thought to predominate in the head of the hippocampus [Martin et al., 2010]. Altogether, our findings indicate that the volume of CA1 provided by FreeSurfer is not comparable to that obtained by manual delineation, and doesn't allow to detect changes in normal aging or pathological conditions. Moreover, they also suggest that, for the SUB and OTHER subfields, even though the correlations between the two methods were relatively high, the segmentations only partly overlap (based on visual inspection) and may therefore not be reliable, as part of CA1 is included in the SUB and OTHER subfields.

Over and above the comparison between both methods, as a secondary aim, this study provides insights on the effects of normal aging on hippocampal subfield volumes delineated manually on a relatively large sample of healthy individuals, as well as the effects of MCI and AD. The results highlighted a differential effect of age on hippocampal subfields, with a linear decrease of the volume of the SUB_{manual} from 20 years of age, a nonlinear decrease of CA1_{manual} volume starting around 50 years, and no significant changes in the other subfields. These results are consistent with a post mortem study [Simic et al., 1997] as well as a previous neuroimaging report [Ziegler et al., 2011], both showing that the SUB is the earliest affected subfield in aging and that CA1 volume decreases later. Note that discrepant results were also reported as some

studies only found a significant effect of age on the CA1 subfield [Mueller et al., 2007; Mueller and Weiner, 2009; Shing et al., 2011; Wisse et al., 2014], and other studies did not find a significant effect of age on any subfield [Price et al., 2001; Rössler et al., 2002]. These discrepancies may reflect methodological differences in the segmentation methods—for example, the head of the hippocampus is excluded in Mueller et al. (2009) and Shing et al. (2011) but not in the others, or in the age range of the samples—for example, some studies considered the entire adult lifespan [La Joie et al., 2010; Mueller et al., 2007; Mueller and Weiner, 2009; Ziegler et al., 2011] while others only included elderly people [Price et al., 2001; Rössler et al., 2002; Wisse et al., 2014]. In a previous work on a partly overlapping sample [La Joie et al., 2010], results were similar for the SUB and other regions but we failed to evidence a significant effect of age on CA1. This was probably due to the fact that the sample was smaller ($n = 50$ vs. 98 in this study) and the maximum age lower (68 years) as compared to this study (84 years). This study also highlighted a differential effect of MCI and AD on hippocampal subfield volumes delineated manually, with a similar major decrease of CA1_{manual} in both MCI and AD patients, a more gradual effect on the SUB_{manual}, while the OTHER_{manual} region was impaired only from the AD stage. These results are overall consistent with previous studies showing that CA1 is the most sensitive subfield in MCI and AD [Apostolova et al., 2010; Chételat et al., 2008; Csernansky et al., 2005; Frisoni et al., 2008; Mueller et al., 2010; Wang et al., 2006]. They further support the view that CA1 atrophy may be a sensitive marker of AD in early stages

[Apostolova et al., 2006; Chételat et al., 2008; La Joie et al., 2013; Mueller et al., 2010; Pluta et al., 2012].

This study has several limitations. The two segmentation methods have been applied to different kinds of images: the manual segmentation was performed on high resolution ($0.375 \times 0.375 \times 2$ mm) PD-weighted images while the FreeSurfer segmentation was applied on T1-weighted 1 mm isotropic images. The objective of this study was to test the validity of T1-MR-based FreeSurfer segmentation as performed in most studies using FreeSurfer (see Table I); therefore, we used, for comparison, the images that appear as the best compromise for contrast, resolution, and reproducibility of the delineation (i.e., high resolution PD-weighted images, excluding the most posterior slice—see La Joie et al., 2010). Of course, the gold standard reference would have been histologically-based delineation on postmortem tissue with concomitant ex vivo MR acquisitions. In their study, Schoene-Bake et al., (2014) have investigated the relationship between the FreeSurfer segmentation applied on conventional T1-weighted MR and postsurgical hippocampal histology in patients with mesial temporal lobe epilepsy. This study reported that the neural density of CA1 significantly correlated with the volume of CA1 determined on T1-weighted magnetic resonance imaging. Yet, this study did not assess the entire volume of CA1 so it couldn't inform on the validity of T1-MR-based FreeSurfer segmentation for hippocampal subfield volumetry. Further histologically-based studies are thus needed to assess this question.

CONCLUSION

FreeSurfer is a free and easy-to-use software allowing to delineate automatically the hippocampal subfields. While the method has been validated on ultrahigh resolution T1 images, it is mostly applied to standard resolution T1-weighted MR. This study however raises concerns as regard to the validity of using FreeSurfer to measure hippocampal subfields on standard T1 MRI. It highlights differences in the anatomic definition of the subfields, especially for CA1, when compared to manual delineation on high resolution images. While the correlations between FreeSurfer and manual measurements were reasonable for the SUB and CA2-3-4-DG subfields pooled together (OTHER), the two methods systematically differ in their assessment of the volume of these subfields with significantly larger estimates for FreeSurfer. Moreover, FreeSurfer does not seem to provide a reliable estimate of CA1 volume (no correlation and much lower volume estimates) and failed to detect the widely-described AD-related changes in this subfield. This is particularly a concern for clinical application in early AD diagnosis as earliest AD-related changes seem to occur in CA1.

ACKNOWLEDGMENTS

The funding sources were not involved in study design, data collection, statistical analysis, results interpretation,

writing of the report, or in the decision to submit the article for publication. The authors have no disclosure. Additional contributions: The authors are grateful to S. Benbrika, F. Dégeilh, M. Fouquet, M. Gaubert, J. Gonneaud, M. Leblond, K. Mevel, J. Mutlu, A. Pélerin, A. Quillard, C. Schupp, C. Tomadesso, N. Villain, and the Cyceron MR-PET staff members for their help with patients and imaging examination.

REFERENCES

- Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, Thompson PM (2006): Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Arch Neurol* 63:693–699.
- Apostolova LG, Mosconi L, Thompson PM, Green AE, Hwang KS, Ramirez A, Mistur R, Tsui WH, De Leon MJ (2010): Subregional hippocampal atrophy predicts Alzheimer's dementia in the cognitively normal. *Neurobiol Aging* 31:1077–1088.
- Arenaza-Urquijo EM, Landeau B, La Joie R, Mevel K, Mézence F, Perrotin A, Desgranges B, Bartrés-Faz D, Eustache F, Chételat G (2013): Relationships between years of education and gray matter volume, metabolism and functional connectivity in healthy elders. *NeuroImage* 83:450–457.
- Bird CM, Burgess N (2008): The hippocampus and memory: insights from spatial processing. *Nat Rev Neurosci* 9:182–194.
- Burgess N, Maguire E a, O'Keefe J (2002): The human hippocampus and spatial and episodic memory. *Neuron* 35:625–641.
- Cherbuin N, Anstey KJ, Réglade-Meslin C, Sachdev PS (2009): In vivo hippocampal measurement and memory: A comparison of manual tracing and automated segmentation in a large community-based sample. *PLoS one* 4:e5265.
- Chételat G, Fouquet M, Kalpouzos G, Denghien I, De la Sayette V, Viader F, Mézence F, Landeau B, Baron JC, Eustache F, Desgranges B (2008): Three-dimensional surface mapping of hippocampal atrophy progression from MCI to AD and over normal aging as assessed using voxel-based morphometry. *Neuropsychologia* 46:1721–1731.
- Csernansky JG, Wang L, Swank J, Miller JP, Gado M, McKeel D, Miller MI, Morris JC (2005): Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *NeuroImage* 25:783–792.
- Durazzo TC, Meyerhoff DJ, Nixon SJ (2013): Interactive effects of chronic cigarette smoking and age on hippocampal volumes. *Drug and Alcohol Depend* 133:704–711.
- Duvernoy HM (2005): *The Human Hippocampus, Functional Anatomy, Vascularization and Serial Sections with MRI*, 3rd ed. Berlin, Germany: Springer-Verlag.
- Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003): Cellular networks underlying human spatial navigation. *Nature* 425:184–188.
- Frisoni GB, Ganzola R, Canu E, Rüb U, Pizzini FB, Alessandrini F, Zoccatelli G, Beltramello A, Caltagirone C, Thompson PM (2008): Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain* 131: 3266–76.
- Geuze E, Vermetten E, Bremner JD (2005): MR-based in vivo hippocampal volumetrics: 2. Findings in neuropsychiatric disorders. *Mol Psychiatry* 10:160–184.

- Hanseeuw BJ, Van Leemput K, Kavec M, Grandin C, Seron X, Ivanoiu a (2011): Mild cognitive impairment: Differential atrophy in the hippocampal subfields. *AJNR Am J Neuroradiol* 32: 1658–1661.
- Harding AJ, Halliday GM, Kril JJ (1998): Variation in hippocampal neuron number with age and brain volume. *Cereb Cortex* 8: 710–718.
- Kerchner GA, Bernstein JD, Fenesy MC, Deutsch GK, Saranathan M, Zeineh MM, Rutt BK (2013): Shared vulnerability of two synaptically-connected medial temporal lobe areas to age and cognitive decline: A seven tesla magnetic resonance imaging study. *J Neurosci* 33:16666–16672.
- La Joie R, Fouquet M, Mézengue F, Landeau B, Villain N, Mevel K, Pélerin A, Eustache F, Desgranges B, Chételat G (2010): Differential effect of age on hippocampal subfields assessed using a new high-resolution 3T MR sequence. *NeuroImage* 53: 506–514.
- La Joie R, Perrotin A, Barré L, Hommet C, Mézengue F, Ibazizene M, Camus V, Abbas A, Landeau B, Guilloteau D, De La Sayette V, Eustache F, Desgranges B, Chételat G (2012): Region-specific hierarchy between atrophy, hypometabolism, and β -amyloid (A β) load in Alzheimer's disease dementia. *J Neurosci* 32:16265–73.
- La Joie R, Perrotin A, De La Sayette V, Egret S, Doeuvre L, Belliard S, Eustache F, Desgranges B, Chételat G (2013): Hippocampal subfield volumetry in mild cognitive impairment, Alzheimer's disease and semantic dementia. *NeuroImage Clin* 3: 155–162.
- Lepage M, Habib R, Tulving E (1998): Hippocampal PET activations of memory encoding and retrieval: The HIPER model. *Hippocampus* 8:313–322.
- Li Y-D, Dong H-B, Xie G-M, Zhang L (2013): Discriminative analysis of mild Alzheimer's disease and normal aging using volume of hippocampal subfields and hippocampal mean diffusivity: An in vivo magnetic resonance imaging study. *Am J Alzheimer's Dis Other Dement* 28:627–633.
- Lim HK, Hong SC, Jung WS, Ahn KJ, Won WY, Hahn C, Kim IS, Lee CU (2012a): Automated hippocampal subfield segmentation in amnesic mild cognitive impairments. *Dement and Geriatr Cogn Disord* 33:327–333.
- Lim HK, Hong SC, Jung WS, Ahn KJ, Won WY, Hahn C, Kim I, Lee CU (2012b): Automated hippocampal subfields segmentation in late life depression. *J Affect Disord* 143:253–256.
- Lim HK, Hong SC, Jung WS, Ahn KJ, Won WY, Hahn C, Kim IS, Lee CU (2013): Automated segmentation of hippocampal subfields in drug-naïve patients with Alzheimer disease. *AJNR Am J Neuroradiol* 34:747–751.
- Maguire EA, Burgess N, Donnett JG, Frackowiak RS, Frith CD, O'Keefe J (1998): Knowing where and getting there: A human navigation network. *Science* 280:921–924.
- Martin SB, Smith CD, Collins HR, Schmitt FA, Gold BT (2010): Evidence that volume of anterior medial temporal lobe is reduced in seniors destined for mild cognitive impairment. *Neurobiol Aging* 31:1099–1106.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984): 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* 34:939–939.
- Mevel K, Landeau B, Fouquet M, La Joie R, Villain N, Mézengue F, Perrotin A, Eustache F, Desgranges B, Chételat G (2013): Age effect on the default mode network, inner thoughts, and cognitive abilities. *Neurobiol Aging* 34:1292–1301.
- Mueller SG, Weiner MW (2009): Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields 564:558–564.
- Mueller SG, Stables L, Du AT, Schuff N, Truran D, Cashdollar N, Weiner MW (2007): Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4T. *Neurobiol Dis* 28:719–726.
- Mueller SG, Schuff N, Yaffe K, Madison C, Miller B, Weiner MW (2010): Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease. *Hum Brain Mapp* 31: 1339–1347.
- Pereira JB, Junqué C, Bartrés-Faz D, Ramírez-Ruiz B, Martí M-J, Tolosa E (2013): Regional vulnerability of hippocampal subfields and memory deficits in Parkinson's disease. *Hippocampus* 23:720–728.
- Pereira JB, Valls-Pedret C, Ros E, Palacios E, Falcón C, Bargalló N, Bartrés-Faz D, Wahlund L-O, Westman E, Junque C (2014): Regional vulnerability of hippocampal subfields to aging measured by structural and diffusion MRI. *Hippocampus* 24: 403–414.
- Petersen RC, Morris JC (2005): Mild cognitive impairment as a clinical entity and treatment target. *Arch Neurol* 62:1160–1163; discussion 1167.
- Pluta J, Yushkevich P, Das S, Wolk D (2012): In vivo analysis of hippocampal subfield atrophy in mild cognitive impairment via semi-automatic segmentation of T2-weighted MRI. *J Alzheimer's Dis* 31:85–99.
- Price JL, Ko a I, Wade MJ, Tsou SK, McKeel DW, Morris JC (2001): Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. *Arch Neurol* 58:1395–1402.
- Rössler M, Zarski R, Bohl J, Ohm TG (2002): Stage-dependent and sector-specific neuronal loss in hippocampus during Alzheimer's disease. *Acta Neuropathol* 103:363–369.
- Schoene-Bake J-C, Keller SS, Niehusmann P, Volmering E, Elger C, Deppe M, Weber B (2014): In vivo mapping of hippocampal subfields in mesial temporal lobe epilepsy: Relation to histopathology. *Hum Brain Mapp* 35:4718–4728.
- Shing YL, Rodrigue KM, Kennedy KM, Fandakova Y, Bodammer N, Werkle-Bergner M, Lindenberger U, Raz N (2011): Hippocampal subfield volumes: age, vascular risk, and correlation with associative memory. *Front Aging Neurosci* 3:2.
- Simic G, Kostovic I, Winblad B, Bogdanovic N (1997): Volume and number of neurons of the human hippocampal formation in normal aging and Alzheimer's disease. *J Comp Neurol* 379: 482–494.
- Spaniol J, Davidson PSR, Kim ASN, Han H, Moscovitch M, Grady CL (2009): Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia* 47:1765–1779.
- Squire LR, Ojemann JG, Miezin FM, Petersen SE, Videen TO, Raichle ME (1992): Activation of the hippocampus in normal humans: A functional anatomical study of memory. *Proc Natl Acad Sci USA* 89:1837–1841.
- Tae WS, Kim SS, Lee KU, Nam E-C, Kim KW (2008): Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology* 50:569–581.
- Teicher MH, Anderson CM, Polcari A (2012): Childhood maltreatment is associated with reduced volume in the hippocampal

- subfields CA3, dentate gyrus, and subiculum. *Proc Natl Acad Sci USA* 109:E563–E572.
- Tulving E, Markowitsch HJ (1998): Episodic and declarative memory: Role of the hippocampus. *Hippocampus* 8:198–204.
- Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B (2009): Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* 19:549–557.
- Wang L, Swank JS, Glick IE, Gado MH, Miller MI, Morris JC, Csernansky JG (2003): Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging *NeuroImage* 20:667–682.
- Wang L, Miller JP, Gado MH, McKeel DW, Rothermich M, Miller MI, Morris JC, Csernansky JG (2006): Abnormalities of hippocampal surface structure in very mild dementia of the Alzheimer type. *NeuroImage* 30:52–60.
- Wenger E, Mårtensson J, Noack H, Bodammer NC, Kühn S, Schaefer S, Heinze H-J, Düzel E, Bäckman L, Lindenberger U, Lövdén M (2014): Comparing manual and automatic segmentation of hippocampal volumes: Reliability and validity issues in younger and older brains. *Hum Brain Mapp* 35:4236–4238.
- West MJ, Coleman PD, Flood DG, Troncoso JC (1994): Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer’s disease. *Lancet* 344:769–772.
- Wisse LEM, Biessels GJ, Heringa SM, Kuijf HJ, Koek D (H. . L, Luijten PR, Geerlings MI (2014): Hippocampal subfield volumes at 7T in early Alzheimer disease and normal aging. *Neurobiol Aging*.
- Yushkevich PA, Wang H, Pluta J, Das SR, Craige C, Avants BB, Weiner MW, Mueller S (2010): Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. *NeuroImage* 53:1208–1224.
- Ziegler G, Dahnke R, Jäncke L, Yotter RA, May A, Gaser C (2011): Brain structural trajectories over the adult lifespan. *Hum Brain Mapp* 33:2377–2389.