Hippocampus and amygdala volumes in a random community-based sample of 60–64 year olds and their relationship to cognition

Jerome J. Maller⁎, Kaarin J. Anstey, Chantal Réglade-Meslin, Helen Christensen, Wei Wen, Perminder Sachdev

Centre for Mental Health Research, The Australian National University, Canberra, Australia
School of Psychiatry, University of New South Wales, and Neuropsychiatric Institute, Prince of Wales Hospital, Sydney, Australia

Received 5 May 2005; received in revised form 8 December 2005; accepted 21 June 2007

Abstract

Reduced volumes of the hippocampus (HC) and amygdala (AG) are potential biomarkers for Alzheimer’s disease (AD) and other neuropsychiatric disorders. Published studies on HC and AG volumes suffer from methodological limitations, and a valid and reliable normative database does not exist. This study aimed to establish a database of HC and AG volumes from a large community sample of participants 60–64 years old and relate them to cognition. A total of 452 randomly selected participants (from 622 approached) were retained in the study (238 males, 214 females), and all received brain MRI scans, as well as cognitive and physical assessments. HC and AG volumes were estimated from manual tracings on T1-weighted images, and intracranial volume (ICV) was obtained from an automated program. In both sexes, right hippocampi were larger than left, while left amygdalae were larger than right. The only correlation to remain significant after normalization was left HC volume and percent retention of a word list in females. This study provides a HC and AG volumetrics database and describes its relationship with cognitive performance in a representative sample using a standard methodology that will be a reference for future studies. It will therefore have clinical applicability in early AD and other disorders.

Keywords: Hippocampus; Amygdala; MRI; Volumetrics; Sex; Memory

1. Introduction

1.1. Normative volumetrics

Medial temporal lobe (MTL) structures, in particular the amygdala (AG) and hippocampus (HC), show atrophy in the early stages of Alzheimer’s disease (AD) and are potential markers for pre-clinical AD (Bobinski et al., 1995; Golebiowski et al., 1999; Laakso, 2002). Volumetric abnormalities in these structures have also been noted in other neuropsychiatric disorders such as schizophrenia, depression, and post-traumatic stress disorder (Garcia, 2002; Nelson et al., 1998). A number of case-control studies have examined the volumes of these structures (e.g. Convit et al., 2000; Mega et al., 2002; Seab et al., 1988), and related them to clinical parameters, but these studies often have small control groups, making for low generalizability to the broader community. Magnetic resonance imaging (MRI) parameters,
delineation of structures, and methods of normalization for brain volume also vary among studies. All of these factors may affect volumetric results (Brierly et al., 2002; Jack, 1994; Laakso et al., 1997; Luft et al., 1996). Many of these important past studies were performed as pilot investigations and identified important clinical effects that illustrated the ultimate need for a normative database.

Brierly et al. (2002) conducted a systematic review and meta-analysis of 51 studies on volumetric MRI of the AG. An overall volumetric estimate was not meaningful, and they concluded that the main methodological factor to influence AG measurement was anatomical definition. Hippocampal volumetrics also differ greatly between studies, mostly due to the studies’ inclusion or exclusion of different portions of the HC, i.e. the head, body and tail. A recent review by Anstey and Maller (2003) expanded on these differences in the context of studies investigating mild cognitive impairment. They and others have concluded that normative data on volumes of medial temporal lobe structures are required to enable comparability between studies and for clinical applications of volumetric research. Toward that end, this article reports normative data for a large representative sample of 60–64 individuals by using comprehensive, consistent and reliable anatomical methodology, and clear statistical analytical techniques.

1.2. Normative volumetrics and cognition

Gradual decline in cognitive ability is characteristic of longitudinal studies of the elderly (DeCarli, 2003), with 25% of individuals over 65 years of age having sufficient cognitive problems, short of dementia, to affect the quality of their lives (Unverzagt et al. 2001). In particular, memory is commonly the first neuropsychological skill to decline (Bartres-Faz et al., 2001; Convit et al., 2003) and may be a key indicator of the development of neurodevelopmental disorders such as Alzheimer’s disease, the most common form of dementia. Investigators have suggested that clinically recognized memory dysfunction may be a feature of normal ageing (Goldman and Morris, 2001), but that substantive loss in memory and other cognitive abilities is not part of ‘normal’ ageing (Rubin et al., 1998; Wilson et al., 1999).

Reduced HC volume has been found to be significantly associated with memory loss (e.g. Convit et al., 2003; Hackert et al., 2002; Lye et al., 2004). Similarly, the entorhinal cortex has also been found to be reduced in size in those with pre-clinical dementia (Strange et al., 1999), consistent with the distribution of plaques and tangles in the earliest stages of AD (Braak and Braak, 1997; Krasuski et al., 1998). However, the “bigger is better” hypothesis of the relationship between HC volume and memory performance has not been supported in many populations investigated. Differences in the type and amount of material to be recalled, the length of the delay, the modality in which the information was presented, or a combination of these factors may have contributed to the mixed findings of reported brain–behavior associations (e.g. Chantome et al., 1999; Sullivan et al., 1995; van Petten et al., 2004). Furthermore, there have been differences in sample sizes, sex composition, and ranges of age and education; MRI acquisitions varied as well. Recently, a meta-analysis of 33 studies of HC volumes and memory performance (van Petten, 2004) and of 15 studies of HC volumes and memory performance by the current authors (unpublished) found that the finding of a positive relationship between HC volume and verbal memory (immediate or delayed recall) was inconsistent, and that the relationship was more likely to be found in older age groups; the relationship in young groups was more often null or negative. This suggests that the relationship between HC volume and memory performance may vary with age, such that there is a weak or significant negative effect in youth, null in middle-age, and positive in older aged individuals. However, it may be that no significant relationship early on becomes significant because of decline in some individuals, i.e. the base volume is unrelated to memory, but decline in volume is related to decline in memory.

Structural and functional studies suggest a major role for the AG in the emotional modulation of memory (Abe, 2001; Cahill et al., 1996; Phelps and Anderson, 1997), while the HC has been shown to be more related to the formation, storage and recall of any type of memory (Brewer and Moghekar, 2002; Preston and Gabrieli, 2002). Clearly what is needed is a study to not only estimate HC and AG volumes in a reliable and detailed fashion, but to also measure memory performances in a large sample of normal individuals. This article responds to this need. We hypothesize that (1) population estimates of HC and AG volumetry are influenced by sex and intracranial volume (ICV, used as a proxy for premorbid brain size), (2) HC volumes have a significantly negative correlation with memory performance, specifically the left hippocampus with verbal delayed recall (and by measuring only verbal episodic memory, a dissociation will be shown, i.e. episodic memory for words will be related to HC volume but not AG volume), and (3) ICV will confound
the relationships of HC and AG volumetry with cognition.

2. Method

2.1. Subjects

Participants were sampled randomly from the electoral rolls for Canberra, ACT, and the neighboring city of Queanbeyan, NSW, Australia (enrolment to vote is compulsory for Australian citizens), as part of the PATH Through Life Project, which is a longitudinal study of social, psychological and biological risk factors for high-prevalence mental health problems. The project involves cohorts initially aged 20–24, 40–44, or 60–64 years who are to be followed up every 4 years for 20 years. Further details on the full PATH Through Life Project have been published previously (Anstey et al., 2004; Sachdev et al., 2004). The response rate for the 60–64 age group was 58.3%. When characteristics of this sample were compared with census data on the population, the sample was found to be better educated, but to be similar in marital status and employment status.

During the community survey, respondents were asked whether they would be willing to undergo an MRI scan, and 2076 out of 2551 (81.5%) said they would. Subsequently, a randomly selected sub-sample of 622 of the willing 2076 people were invited to undergo a scan and 478 participated (76.8%). Those who said during the initial interview that they were unwilling to have a scan were significantly ($P<0.05$) more likely to be female, to be of non-English speaking background, to be less educated, with poorer physical health and to have lower cognitive test scores. Those who refused to undergo a scan when actually invited were significantly ($b$) more likely to be of non-English speaking background, to be less educated, with poorer physical health and to have lower cognitive scores. There was no significant difference in depression or anxiety (as measured by the Goldstein scales; Goldberg et al., 1988) at either stage.

2.2. Procedure

2.2.1. Survey procedure and cognitive testing

Each participant was interviewed individually by a member from a team of professional survey interviewers experienced in epidemiological and social surveys. The screening was conducted by trained interviewers i.e. information from the general PATH interview was used, with medical conditions being self-reported. Individuals in the 60–64 year old cohort were given the Mini-Mental State Examination (MMSE; Folstein et al., 1975), as well as other cognitive tasks to perform including the immediate and delayed recalls of the first trial from the California Verbal Learning Test (CVLT; Delis et al., 1987), which were used to assess short- and long-term episodic memory. Verbal working memory was assessed by the Digit Span Backward (DSB) subtest from the Wechsler Memory Scale (Wechsler, 1945), psychomotor speed by the Symbol–Digit Modalities Test (SDMT; Smith, 1982), and reading vocabulary as a proxy for premorbid intelligence by the Spot-the-Word Test Version A (STW; Baddeley et al., 1992). Reaction time (RT) was tested using a small box held with both hands, with left and right buttons at the top to be depressed by the index fingers. There were four blocks of 20 trials measuring simple RT, followed by two blocks of 20 trials measuring choice RT. For simple RT everyone used their right hand regardless of dominance. Simple RT was measured over 80 trials, and choice RT was measured over 40 trials. More details on the RT protocol are available in Anstey et al. (2004).

Approval for the study was obtained from the human research ethics committees of the Australian National University and the University of New South Wales, and all participants provided written informed consent. Assessments were conducted between April 2001 and July 2002, and the time between assessments and MRI scanning was no more than 12 months for any participant.

2.2.2. Image acquisition

Imaging was conducted on a 1.5 Tesla Gyroscan MRI scanner (ACS-NT, Philips Medical Systems, Best, Netherlands) for T1-weighted 3-D structural and T2-weighted FLAIR (fluid attenuated inversion recovery) sequence MRI. A 2D scout mid-sagittal cut for anterior–posterior commissure plane alignment was first acquired. Then 3-D structural MRI was acquired in coronal orientation using a T1-weighted fast field echo sequence (TR/TE/NEX=28.05/2.64/2; flip angle=30; matrix size=256×256; FOV=260×260 mm; slice thickness=2.0 mm, inter-slice distance=1.0 mm), yielding over-contiguous coronal slices 1.0 mm thick with an in-plane spatial resolution of 1.016×1.016 mm/pixel. The FLAIR sequence was acquired in coronal orientation (TR/TE/TI/NEX=11000/140/2600/2; matrix size=256×256; FOV=230×230 mm; slice thickness=4.0 mm with no gap between slices) with in-plane spatial resolution of 0.898×0.898 mm/pixel. The total time of each subject’s scanning session was approximately 20 min. As 26 of the scans were untraceable due to excessive artifact (15 males and 11 females), the total number of participants with traceable MR images totaled 452 (238 males, 214 females).
2.2.3. Volumetric measurement

The volumes of brain anatomical regions were determined by two researchers (JM, CM), each manually outlining the periphery of the regions of interest (ROI) on a subset of the coronal T1-weighted slices (approximately 50:50) using Analyze 5.0 (Brain Imaging Resource, Mayo Clinic, Rochester, MN). The boundaries of the ROIs were outlined with a cordless infrared light-driven cursor, and the number of voxels within the regions was calculated to produce a total volume for the ROI. The outlining of the structures always proceeded from anterior to posterior, and the AG was traced first. The HC included the dentate gyrus, the hippocampus proper, and the subicular complex (Duvernoy, 1999). We often used axial orientation to validate neuroanatomical boundaries.

Amygdala boundaries were based on studies by Watson et al. (1997) with a modification as suggested by Brierly et al. (2002). That is, we checked for the most anterior margin (nucleus) of the AG (Fig. 1A) by continuing to move anteriorly and verifying that this ovoid structure disappeared from view. Watson’s method of determining the anterior margin of the AG is the section where the lateral sulcus closes to form the endorhinal sulcus (Watson’s method); however, this method may be less reliable in atrophy-prone elderly. Hippocampal boundaries were also based on those outlined by Watson et al. (1997), but we were careful to exclude hippocampal sulcal cavities (HSC) at the lateral aspect of the hippocampal fissure as HC volume and not to mistake the uncus recess of the temporal horn for an HSC. An HSC was defined as round or curvilinear in shape on coronal orientation and crescent shaped on axial, and followed the isotense signal characteristics of CSF (Fig. 1D). The most posterior and final slice of the HC

Fig. 1. Examples from MRI slices illustrating boundaries of the amygdala and hippocampus from anterior to posterior. (A) Anterior tip of the amygdala. (B) Anterior separation of the amygdala from the hippocampus. Note the protrusion of the IHLV and the clear strip of white matter (alveus) as the boundary between the two structures. (C) IHLV becomes clearer and HC becomes larger. (D) Hippocampal boundaries posterior to last slice of amygdala showing the choroidal fissure. Note the exclusion of HSC volume from the HC volume area outlined. (E) Most posterior slice taken for the right HC. (F) One slice posterior to (E) showing clear separation of the crus of the fornix from the wall of the lateral ventricle on the right (arrow head).
was defined as the slice before the most anterior section where the crus of the fornix was seen to clearly separate from the wall of the lateral ventricle (Fig. 1E and F). Boundaries were CSF-GM and GM-WM contrasts. Fig. 2 is an illustration of a right-side hippocampus and amygdala rendered from the full set of manual tracings from the above MRI example.

Data were normalized by dividing each HC or AG volume by the individual’s total intracranial volume (ICV). This was estimated from their T1-weighted images. The ICV procedure was achieved through an automated process using SPM99 (Wellcome, UK). In order to determine correspondence between the automated method used in this article and results from manual tracings, five T1-weighted scans were randomly selected and the ICV determined by manually outlining the cranial cavity of each slice (approximately 252 slices per scan); the error between these volumes and those produced by the automated procedure was below 5% (see Wen and Sachdev, 2004, for a full description of the process).

Reliability of the regional volumetric measures was assessed by an intra- and inter-class correlation formula that presumes random selection of raters, ICC (Shrout and Fleiss, 1979). The two tracers yielded ICCs in a pilot study (n=5) that were well within acceptable limits (>0.90). ICC between raters was determined (n=10) after every 50 brain measurements. Each tracer further measured 10 brains in order to examine inter-rater consistency. ICCs were recalculated once 450 scans had been traced for the ROIs.

2.3. Statistical analysis

The data were analyzed by using SPSS-PC version 11.5 software (Chicago, Ill.). The data were screened for variability and potential outliers. Analysis of variance (ANOVA) and paired t-tests were used to compare the means between volumes of both sides of the brains, and for the effect of sex and ICV.

3. Results

3.1. Demographic variables

Men and women had a mean age of 62.62 (1.42) years and 62.60 (1.45) years, respectively. Female participants had significantly fewer years of education than did males (males: 14.52 (2.46), females: 13.61 (2.69); F (1, 451)= 13.93, P<0.01), but they did not differ on MMSE scores (males: mean=29.25 (1.17), females: mean=29.39

Fig. 2. Four different views of the 3D rendered right-side hippocampus and amygdala. (A) Sagittal: approximately aligned with long axis of HC; (B) Coronal: approximately aligned with AC–PC; (C) Coronal: as in (B) but rotated left approximately 45° to more clearly expose detail of medial body of the hippocampus; (D) Coronal: tilted anterior–posterior. Green=hippocampus; Red=amygdala. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
11.08; $F(1,451) = 1.01, P > 0.05$). Eleven participants scored below 27 on the MMSE (males = 7, females = 4), of whom one male and two females scored below 24 on the MMSE (commonly regarded as the threshold for clinical dementia). As this study’s sample is intended to represent a random population, these participants were retained. Likewise, the 329 participants with medical conditions (54 with heart trouble, 40 with diabetes, 19 who had had a stroke, 184 with hypertension, 32 with asthma) and the 22 who were found on MRI to have a neuroanatomical abnormality (4 adenomas, 3 meningiomas, and 1 of each of the following: tumour, haemangionioma, lipoma, normal pressure hydrocephalus, infarct-like lesion, cyst, aneurysm, sequelae of temporal stroke) were also retained.

3.2. Volumetrics

ICCs between the two raters for the right HC and left HC were 0.970 and 0.937, respectively, and 0.943 and 0.974 for the right and left AG, respectively (overall average for the four structures was 0.956). ICCs between raters ranged from 0.948 to 0.989 and 0.981 to 0.993 for the right and left HC, and from 0.975 to 0.989 and 0.995 to 0.996 for the right and left AG, respectively (averaged 0.982 and 0.996).

3.2.1. ICV

The mean ICV of males (1.59 l, S.D. = 0.15) was larger than those of females (1.40 l, S.D. = 0.13; $F(1,451) = 225.44, P < 0.01$).

3.2.2. Hippocampus

Hippocampal volumes were normally distributed.

3.2.2.1. Non-normalized hippocampal volumes.

The average non-normalized right and left hippocampal volumes were larger in males than in females ($F(1,451) = 50.59, P < 0.01$ and $F(1,451) = 48.75, P < 0.01$, respectively). For both sexes, the average right volume was significantly larger than for the left HC (males; $t(237) = 3.01, P < 0.01$; females; $t(213) = 2.63, P < 0.01$).

3.2.2.2. Normalized hippocampal volumes.

Dividing the raw right and left HC volumes by ICV produced significantly larger right and left hippocampal volumes for females than for males (right; $F(1,451) = 6.27, P < 0.05$, left; $F(1,451) = 7.10, P < 0.01$). Males and females both had significantly larger right HC than left HC volumes (males; $t(237) = 3.56, P < 0.01$; females; $t(213) = 2.75, P < 0.01$).

3.2.3. Amygdala

Amygdalar volumes were normally distributed.

3.2.3.1. Non-normalized amygdalar volumes.

The average right and left non-normalized AG volumes were significantly larger in males than in females (right $F(1,451) = 30.62, P < 0.01$; left $F(1,451) = 54.66, P < 0.01$). While the average left AG volume was larger than the average volume of the right AG for both sexes, this was only significant for males (males $t(237) = 3.56, P < 0.01$).

3.2.3.2. Normalized amygdalar volumes.

Right and left AG volumes did not differ between males and females when they were normalized by ICV. Mean left AG volumes normalized by ICV were larger than those

### Table 1

<table>
<thead>
<tr>
<th>(R) HC</th>
<th>(L) HC</th>
<th>(R) AG</th>
<th>(L) AG</th>
<th>(R) HC divided by ICV</th>
<th>(L) HC divided by ICV</th>
<th>(R) AG divided by ICV</th>
<th>(L) AG divided by ICV</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>238</td>
<td>238</td>
<td>238</td>
<td>238</td>
<td>238</td>
<td>238</td>
<td>238</td>
</tr>
<tr>
<td>Mean</td>
<td>3029.07</td>
<td>2975.27</td>
<td>1268.96</td>
<td>1318.27</td>
<td>1908.28</td>
<td>1876.54</td>
<td>799.56</td>
</tr>
<tr>
<td>Percentiles</td>
<td>245.59</td>
<td>418.45</td>
<td>269.73</td>
<td>259.38</td>
<td>259.66</td>
<td>255.69</td>
<td>162.65</td>
</tr>
</tbody>
</table>

For males (1.59 l, S.D. = 0.15) was larger than those of females (1.40 l, S.D. = 0.13; $F(1,451) = 225.44, P < 0.01$). ICCs between raters ranged from 0.948 to 0.989 and 0.981 to 0.993 for the right and left HC, and from 0.975 to 0.989 and 0.995 to 0.996 for the right and left AG, respectively (averaged 0.982 and 0.996).

3.2.1. ICV

The mean ICV of males (1.59 l, S.D. = 0.15) was larger than those of females (1.40 l, S.D. = 0.13; $F(1,451) = 225.44, P < 0.01$). ICCs between raters ranged from 0.948 to 0.989 and 0.981 to 0.993 for the right and left HC, and from 0.975 to 0.989 and 0.995 to 0.996 for the right and left AG, respectively (averaged 0.982 and 0.996).

3.2.2. Hippocampus

Hippocampal volumes were normally distributed.

3.2.2.1. Non-normalized hippocampal volumes.

The average non-normalized right and left hippocampal volumes were larger in males than in females ($F(1,451) = 50.59, P < 0.01$ and $F(1,451) = 48.75, P < 0.01$, respectively). For both sexes, the average right volume was significantly larger than for the left HC (males; $t(237) = 3.01, P < 0.01$; females; $t(213) = 2.63, P < 0.01$).

3.2.2.2. Normalized hippocampal volumes.

Dividing the raw right and left HC volumes by ICV produced significantly larger right and left hippocampal volumes for females than for males (right; $F(1,451) = 6.27, P < 0.05$, left; $F(1,451) = 7.10, P < 0.01$). Males and females both had significantly larger right HC than left HC volumes (males; $t(237) = 3.56, P < 0.01$; females; $t(213) = 2.75, P < 0.01$).

3.2.3. Amygdala

Amygdalar volumes were normally distributed.

3.2.3.1. Non-normalized amygdalar volumes.

The average right and left non-normalized AG volumes were significantly larger in males than in females (right $F(1,451) = 30.62, P < 0.01$; left $F(1,451) = 54.66, P < 0.01$). While the average left AG volume was larger than the average volume of the right AG for both sexes, this was only significant for males (males $t(237) = 3.56, P < 0.01$).

3.2.3.2. Normalized amygdalar volumes.

Right and left AG volumes did not differ between males and females when they were normalized by ICV. Mean left AG volumes normalized by ICV were larger than those

<table>
<thead>
<tr>
<th>(R) HC</th>
<th>(L) HC</th>
<th>(R) AG</th>
<th>(L) AG</th>
<th>(R) HC divided by ICV</th>
<th>(L) HC divided by ICV</th>
<th>(R) AG divided by ICV</th>
<th>(L) AG divided by ICV</th>
</tr>
</thead>
</table>
on the right for males \((t\,(237)=3.62, \ P<0.01)\). Normative HC and AG volumetrics are presented in Tables 1 and 2.

### 3.3. Behavior, comorbidities, MMSE performance, and volumetrics

Post hoc analyses revealed no effect of any depression or anxiety variable upon volumetrics. Excluding the 22 individuals with abnormalities on brain MRI or with medical conditions made no difference to the results. In addition, there were no significant differences between those with any of these comorbidities and those without.

While there was insufficient statistical power to compare the volumetrics of those who scored below 27 on the MMSE against those who scored 27 or over (i.e. to include those within 2 S.D. of the mean), their exclusion made no difference to the significance levels of the volumetric analyses. Furthermore, there were no significant correlations between MMSE performance and volumetrics, nor between years of education and any of the raw volumetrics for either sex (all \(P>0.05\)).

### 3.4. Laterality

From the tests of symmetry in Table 3, the differences between the right and left HC volumes were not significant between the sexes. There was a significantly larger difference between the right and left amygdalae in males for raw volumes only \((F\,(1, 450)=4.22, \ P<0.05)\). The proportion of right-to-left HC or AG did not reach significance between the sexes.

### 3.5. Volumetrics and cognition: males versus females

Table 4 displays the average performances for the subjects separately for males and females. There were significant differences between males and females in their performances on the CVLT-Immediate Recall \((F\,(1, 450)=26.02, \ P<0.01)\), CVLT-Delayed Recall \((F\,(1, 450)=23.53, \ P<0.01)\), and DS-B \((F\,(1, 450)=7.35, \ P<0.01)\), whereby the females performed significantly higher on the CVLT scores but lower on DS-B. No differences were found for CVLT-DIR, STW, SDMT, or MMSE. The mean simple RT (80 trials) performance was lower (faster) for males \((F\,(1, 437)=11.60, \ P<0.01)\), and there was a trend

| Raw (mm³) and normalized hippocampal and amygdalar volumetrics for females 60–64 years of age |
|-----------------------------------------------------|---------------------|---------------------|---------------------|
| \(N\) | \(214\) | \(214\) | \(214\) | \(214\) | \(197.52\) | \(1934.55\) | \(1924.21\) | \(19.06\) | \(10.18\) | \(10.24\) |
| Mean | \(2744.92\) | \(2706.51\) | \(1140.92\) | \(1152.06\) | \(818.73\) | \(827.06\) |
| Standard error of mean | \(26.74\) | \(27.16\) | \(14.74\) | \(14.57\) | \(10.18\) | \(10.24\) |
| S.D. | \(397.32\) | \(397.32\) | \(215.70\) | \(213.21\) | \(148.96\) | \(149.83\) |
| Percentiles | \(1961.72\) | \(1913.90\) | \(672.65\) | \(644.79\) | \(477.52\) | \(491.72\) |
| \(10\) | \(2235.79\) | \(2222.89\) | \(870.59\) | \(909.79\) | \(633.87\) | \(652.44\) |
| \(20\) | \(2352.86\) | \(2319.85\) | \(957.24\) | \(979.00\) | \(697.31\) | \(710.69\) |
| \(30\) | \(2552.86\) | \(2458.59\) | \(1026.11\) | \(1023.25\) | \(744.09\) | \(739.63\) |
| \(40\) | \(2640.65\) | \(2562.26\) | \(1090.30\) | \(1088.24\) | \(776.71\) | \(777.87\) |
| \(50\) | \(2730.91\) | \(2711.31\) | \(1144.46\) | \(1125.37\) | \(739.63\) | \(739.63\) |
| \(60\) | \(2838.70\) | \(2812.91\) | \(1191.39\) | \(1181.00\) | \(825.35\) | \(848.58\) |
| \(70\) | \(2944.95\) | \(2956.13\) | \(1237.81\) | \(1248.51\) | \(891.85\) | \(896.42\) |
| \(80\) | \(3076.98\) | \(3054.28\) | \(1323.42\) | \(1328.58\) | \(946.10\) | \(943.82\) |
| \(90\) | \(3269.08\) | \(3215.72\) | \(1395.63\) | \(1463.19\) | \(1012.29\) | \(1034.27\) |
| \(99\) | \(3877.84\) | \(4054.22\) | \(1811.53\) | \(1673.72\) | \(1207.61\) | \(1191.06\) |

Table 3

<table>
<thead>
<tr>
<th>Hippocampal and amygdalar laterality by sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
</tr>
<tr>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Hippocampus (right–left): raw (\dagger)</td>
</tr>
<tr>
<td>Hippocampus (right–left): /ICV</td>
</tr>
<tr>
<td>Hippocampus (right–left): /TBV</td>
</tr>
<tr>
<td>Hippocampus (right: left)</td>
</tr>
<tr>
<td>Amygdala (right–left): raw (\dagger)</td>
</tr>
<tr>
<td>Amygdala (right–left): /ICV</td>
</tr>
<tr>
<td>Amygdala (right–left): /TBV</td>
</tr>
<tr>
<td>Amygdala (right: left)</td>
</tr>
</tbody>
</table>

\(P\ (M/F)\) is the significance of the difference between the volumes of males and females; \(\dagger\) \(cm³\). The value rendered in bold is considered significant at \(P<0.05\).
for the males to also have performed faster on the choice RT (40 trials) task ($F(1, 435)=3.40, P=0.066)$.

### 3.5.1. Hippocampus and cognition

#### 3.5.1.1. Males.
Mean raw right HC volume did not correlate significantly with any of the CVLT scores or any other cognitive variable except STW ($r=0.139, P<0.05$) and SDMT ($r=0.138, P<0.05$; Table 5). Left raw HC correlated significantly with SDMT ($r=0.130, P<0.05$). However, when controlling for years of education, or when division was by ICV or TBV, there were no significant differences.

Left and right normalized HC volumes did not correlate with the cognitive variables, with or without education as a covariate. There were no significant correlations between asymmetry volumetrics and cognitive performances for either sex.

#### 3.5.1.2. Females.
For females, the mean raw left HC volumes correlated significantly with the ratio of Delayed recall to Immediate recall (DIR) to beyond the 0.05 level ($r=0.146, P=0.032$). This association was still significant after normalizing by ICV and TBV (ICV normalized: $r=0.145, P=0.034$; TBV normalized $r=0.161, P=0.018$), and controlling for years of education (raw: $r=0.138, P=0.044$; /ICV: $r=0.142, P=0.039$; /TBV: $r=0.158, P=0.021$). Right HC volumes did not correlate significantly with any of the cognitive variables, except there was a trend for the SDMT to correlate with raw left HC volume ($r=0.131, P=0.056$) but not after normalization.

The significant correlations between left HC volume and DIR are beyond chance and hence not Type I errors. That is, DIR was one of nine (11.11%) variables tested for correlations with volumetrics, which is greater than chance (i.e. 5%).

### 3.5.2. Amygdala and cognition

There were no significant correlations between mean AG volumes (raw, ICV or TBV normalized) and any of the cognitive variables, for either sex. Controlling for years of education made no difference to the results.

### 4. Discussion

This study reports data on the volumes of the hippocampus and amygdala and on cognitive performances in a large representative community-based sample of 60–64 year old men and women. Quantitative volumetric measures introduce a level of precision in the estimation of volume that is simply not available by visually inspecting a set of MR images (Jack, 1994). Hence, these norms have practical use for clinicians. For example, subtle yet significant hippocampal atrophy over time may be more detectable by volumetric analysis of the MRI scans but not from visual inspection of them. Our hippocampal volumes are in line with those reported from other large studies of similarly aged individuals (den Heijer et al., 2002; Hackert et al., 2002; MacLullich et al., 2002; Raz et al., 1997), although slightly smaller, which may be due to the use of thinner slices to trace on than in these other studies and different planes of acquisition (e.g. perpendicular to the AC–PC axis,

---

**Table 4**

Average performances on tests of cognition

<table>
<thead>
<tr>
<th>Variable</th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td></td>
</tr>
<tr>
<td>CVLT-immediate recall</td>
<td>6.98</td>
<td>1.93</td>
<td>7.96</td>
<td>2.15</td>
<td>**</td>
</tr>
<tr>
<td>CVLT-delayed recall</td>
<td>6.11</td>
<td>2.11</td>
<td>7.13</td>
<td>2.35</td>
<td>**</td>
</tr>
<tr>
<td>Delayed/immediate recall</td>
<td>0.88</td>
<td>.21</td>
<td>.89</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>29.22</td>
<td>1.26</td>
<td>29.39</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>Spot-the-Word</td>
<td>52.68</td>
<td>6.15</td>
<td>51.97</td>
<td>5.68</td>
<td></td>
</tr>
<tr>
<td>SDMT</td>
<td>50.94</td>
<td>8.12</td>
<td>51.10</td>
<td>8.90</td>
<td></td>
</tr>
<tr>
<td>Digit span − backwards</td>
<td>5.22</td>
<td>2.18</td>
<td>4.68</td>
<td>2.04</td>
<td>**</td>
</tr>
<tr>
<td>RT − simple (80 trials)</td>
<td>0.24</td>
<td>0.04</td>
<td>0.26</td>
<td>0.06</td>
<td>**</td>
</tr>
<tr>
<td>RT − choice (40 trials)</td>
<td>0.31</td>
<td>0.04</td>
<td>0.32</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

*P is the significance of the difference between the performances of males and females; ** = $P<0.01$. For RT (simple 80 trials), male $n=228$ and female $n=212$; for RT (choice 40 trials) male $n=227$ and female $n=211$.

**Table 5**

Correlations of neuropsychological performances with right and left hippocampal volumes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Subjects</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r$</td>
<td>$P$</td>
<td>$r$</td>
</tr>
<tr>
<td>Right hippocampal volume</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spot-the-Word-# correct</td>
<td>0.139</td>
<td>0.032</td>
<td>0.112</td>
</tr>
<tr>
<td>(raw volume)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normalized ICV</td>
<td>0.044</td>
<td>0.500</td>
<td>0.078</td>
</tr>
<tr>
<td>Normalized TBV</td>
<td>0.046</td>
<td>0.483</td>
<td>0.097</td>
</tr>
<tr>
<td>SDMT (raw volume)</td>
<td>0.138</td>
<td>0.033</td>
<td>0.080</td>
</tr>
<tr>
<td>Normalized ICV</td>
<td>0.053</td>
<td>0.413</td>
<td>0.046</td>
</tr>
<tr>
<td>Normalized TBV</td>
<td>0.031</td>
<td>0.630</td>
<td>0.052</td>
</tr>
<tr>
<td>Left hippocampal volume</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delayed recall/1st trial (raw volume)</td>
<td>-0.011</td>
<td>0.871</td>
<td>0.146</td>
</tr>
<tr>
<td>Normalized ICV</td>
<td>0.053</td>
<td>0.414</td>
<td>0.145</td>
</tr>
<tr>
<td>Normalized TBV</td>
<td>0.055</td>
<td>0.396</td>
<td>0.161</td>
</tr>
<tr>
<td>SDMT (raw volume)</td>
<td>0.130</td>
<td>0.045</td>
<td>0.131</td>
</tr>
<tr>
<td>Normalized ICV</td>
<td>0.039</td>
<td>0.550</td>
<td>0.097</td>
</tr>
<tr>
<td>Normalized TBV</td>
<td>0.017</td>
<td>0.789</td>
<td>0.104</td>
</tr>
</tbody>
</table>

$r$ = correlation, $P$ = significance. Values rendered in bold represent statistics that were considered significant at $<0.05$. 

---
perpendicular to the long axis of the hippocampus, or no correction at all). Furthermore, we were careful to exclude hippocampal sulcal cavities and other fluid-filled spaces from hippocampal volume calculations. Our amygdalar volumes were also within reported ranges (Brierly et al., 2002; Convit et al., 1999; Gur et al., 2002). Studies reporting larger volumes generally have unfavorable attributes such as poor inter- and/or intra-rater reliabilities, small samples, or have not described their volumetric protocols (Filipek et al., 1994; Goldstein et al., 2001; Lange et al., 1997; some used younger subjects. Men had larger raw HC volumes, while the average male HC was smaller than the average female HC when considered as a proportion to ICV. The right hippocampi were larger in both sexes, while the opposite was the case for the amygdalae. These results are consistent with the literature (e.g. Giedd et al., 1996; Gur et al., 2002; Bilir et al., 1998; Duchesne et al., 2002). Furthermore, raw HC and AG volumes are most commonly reported to be larger in males than in females (Gur et al., 2002; Lange et al., 1997; Pruessner et al., 2000; Raz et al., 1997). While the average male ICV may be larger than that of females (Jenkins et al., 2000; Reynolds et al., 1999), the difference between the sexes in their raw HC and AG volumes as a fraction of their ICV may not be proportional; this may lead to no significant difference in normalized volumes between sexes (Gur et al., 2002; Pruessner et al., 2000). Consistent with the literature (e.g. Pedraza et al., 2004), an examination of the asymmetry showed that the ratio of right-to-left HC or AG did not reach significance between the sexes.

4.1. Volumetrics and cognition

A supposition presented by van Petten et al. (2004) was “that, in cognitively superior elderly samples (such as that employed by that particular study), variation in memory ability across individuals may be largely determined by pre-existing differences rather than differential degrees of age-related decline", p. 17). The results of the meta-analysis by van Petten (2004) yielded no support for the ‘bigger is better’ hypothesis, but rather found some support for both the developmental and ageing hypotheses, in that the HC volume-memory correlation tended to shift from negative to neutral to positive as the age of the sample increased. While the current study does not have a large enough age range to comprehensively test this, the finding in the current study lends support to a neutral to positive HC volume-memory relationship when considering the differences in this relationship between the sexes in this sample.

There was a significant correlation between HC volume and the ratio between verbal delayed recall and immediate recall (DIR) for females, and not with RT, and AG volume did not relate to any cognitive variable assessed. While performance on the SDMT significantly correlated with right and left raw HC volume for males, and STW performances significantly correlated with right raw HC volumes (for males), these relationships became non-significant when the influence of ICV (or TBV) was considered through simple division, hence showing that ICV is a confound upon the relationships of HC and AG volumes with cognition. However, DIR remained significantly correlated with left HC volume (for females) after applying various normalization methods, hence not supporting the relationship in this case.

Similarly, Lye et al. (2004) recently reported that the left HC volume was an important determinant of retention on the CVLT in community-dwelling individuals aged 81 to 92 years. Although they did not report on the relationship between volumetrics and immediate recall, their results were consistent with our finding of the relationship between retention (DIR) and left HC volume.

4.2. The effect of normalization

A re-analysis of the data controlling for years of education instead of ICV or TBV also rendered all of the significant correlations between HC volume and cognition amongst the males non-significant, although it did not alter the significant relationship between left HC volume and DIR. This significant correlation between raw left HC volume and DIR in females was virtually unaltered by ICV normalization and higher when normalized by TBV. That reflects the fact that DIR did not correlate significantly with ICV (r=0.030, P=0.665), or with TBV (r=0.003, P=0.965). If significant correlations are reduced to non-significant when normalized, then the results logically relate to the normalization factor. That is, the two original factors are only related through their strong associations with a third variable (Atchley et al., 1976). In the current situation, ICV is a primary normalization factor. One complicating factor is that ICV was associated with education in our sample (see Maller et al., submitted for publication).

4.3. Memory and the hippocampus

On average, females in our sample recalled 14% (7.96/6.98) more information than men on the first trial, and 17% (7.13/6.11) more on the delayed trial. Women
have previously been reported to outperform men on a range of memory measures (e.g., Aartsen et al., 2004; Larrabee and Crook, 1993), including the CVLT (e.g., Kramer et al., 1988; Lye et al., 2004). However, we did not find such a difference in the mean retention in the current sample as was reported in the sample of Lye et al. (2004). If retention is dependent upon the level of attention previously directed to encoding the information (which is indicated by the immediate recall performance), then it may be that a threshold of attention has to be reached before the left HC mediates the retention. That is, the relationship between left HC volume and retention in our sample may be related to attention and therefore possibly effort.

It is important at this point to acknowledge that a non-significant correlation with HC volume does not necessarily imply that the HC is not involved as volume differs from function. Memory variability may result from a number of factors including stimulus differences, varying levels of attention and effort put forth across stimuli, and other uncontrolled factors (Reber et al., 2002a,b). The proper functioning of the HC depends on its interactions with subcortical regions concerned with emotion, motivation, arousal and attention (Shastri, 2002). For example, Thoenissen et al. (2002) reported activity in the prefrontal and temporal regions to be signs of awareness of regulation of actions. In other words, under particular conditions (when effort is high), the left HC mediates retention in our sample of females. There is evidence in the literature to support this assumption. For example, Montaldi et al. (2002) found that MTL activation increased in participants as a proportion of response words successfully recalled and that bilateral frontal lobe activation increased in proportion to retrieval effort that was greater when learning had been good. Hence, more effort leads to increased MTL activation during memory recall. Similarly, Reber et al. (2002a,b) showed that left MTL activation was crucial for the encoding that actually led to successful memory on that subsequent test, while prefrontal activation was strongly associated with intentional verbal encoding.

It may be that those in our sample with poor memories had average size hippocampi relative to others in the sample when we scanned them, but they may have had much larger relative maximal volumes at some previous point in time than others in the sample, and are now in the process of (relatively) accelerated atrophy. Alternatively, it may be that those in our sample with better memories had average HC volumes when we scanned them, but they were also close to their maximal HC volumes at some prior point in time (i.e. less HC atrophy). If this scenario is valid, the relationship between HC and memory is different for the normal and pathological ageing brains, which supports previous findings of medial temporal lobe volumetric correlations with memory performance in those with AD but not in normal controls (Basso et al., 2006). Hence, it may be that specific changes in HC volume may only become an important influence upon verbal memory in the presence of pathology (Squire, 2004). It is also possible that some of our participants were in the pre-clinical stages of dementia by virtue of their age.

The finding in the current study of no significant relationships between AG volumes and cognitive performances therefore supports those of past studies (den Heijer et al., 2002; Mackay et al., 1998; Yurgelun-Todd et al., 2003). Furthermore, no relationships were found between HC volumes and RT performances.

4.4. Limitations

While participants in this study did not undergo medical or neurological assessments, self-reported medical and neurological conditions had no effect upon reported volumes. Although this study reports on results from healthy subjects, we did not exclude those with silent brain lesions (Xu et al., 2000) such as subcortical silent brain infarction, WMH on T2-weighted imaging, and periventricular hyperintensity (Kobayashi et al., 1997). However, this is not necessarily a limitation as we sought a representative sample rather than a ‘super-normal’ sample.

We only administered Trial 1 and Immediate Recall of the CVLT and not the full version. The full version would have allowed us to more comprehensively evaluate verbal memory performance in this sample from perspectives such as learning styles of semantic versus rote memory, for example.

4.5. Summary

The present study has many positive attributes, such as employing a large sample randomly selected from the community, representing both sexes equally, and capturing a tight cohort in order to enhance the representativeness of these subvolumes for this age group. Furthermore, very high reliability of the regional volumetric measures was achieved through regular assessment and the application of high-resolution MRI utilizing acquisition protocols regarded to be imperative for valid and reproducible results. Analyses showed that medical and neurological conditions had no effect upon reported volumes.
The pattern in the interpretation of the current study’s results supports the assumption that there are at least two types of ageing: ‘normal’ and ‘pathological’, and that while larger hippocampal volume may not equate to any cognitive advantage, this relationship may differ by sex. Thus, it is imperative that the structures, functions, connectivity and vulnerabilities of the MTL be elucidated. If structural volumetrics of MTL regions are to be understood on a larger scale in community samples, then the normative database presented here, which is the first wave of a longitudinal study, will be a valuable resource.

Acknowledgements

The authors are grateful to the staff of the Centre for Mental Health Research ANU, in particular Patricia Jacomb, Karen Maxwell, June Cullen, Tony Jorm, and Rajeev Kumar, the National Capital Diagnostic Imaging group, particularly Jeremy Price, and the Neuroimaging Group, NPI, Prince of Wales Hospital, in particular Michael Valenzuela. Grant support was provided by NHMRC.

References


