This is the published version of a paper published in *Neurobiology of Aging*.

Citation for the original published paper (version of record):

Longitudinal association between hippocampus atrophy and episodic-memory decline.
*Neurobiology of Aging*, 51: 167-176

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-128725
Longitudinal association between hippocampus atrophy and episodic-memory decline

Tetiana Gorbach a,*, Sara Pudas b, c, Anders Lundquist a, b, Greger Orädd d, MariaJosefsson e, Alireza Salami h, c, f, Xavier de Luna a, Lars Nyberg h, c, d

a Department of Statistics, Umeå School of Business and Economics, Umeå University, Umeå, Sweden
b Umeå Center for Functional Brain Imaging, Umeå University, Umeå, Sweden
c Department of Integrative Medical Biology, Umeå University, Umeå, Sweden
d Department of Radiation Sciences, Umeå University, Umeå, Sweden
e Centre for Demographic and Ageing Research at Umeå University (CEDAR), Umeå University, Umeå, Sweden
f Umeå Center for Functional Brain Imaging, Umeå University, Umeå, Sweden
g Aging Research Center, Karolinska Institutet and Stockholm University, Stockholm, Sweden
h Department of Integrative Medical Biology, Umeå University, Umeå, Sweden
i Department of Radiation Sciences, Umeå University, Umeå, Sweden

ARTICLE INFO

Article history:
Received 30 June 2016
Received in revised form 1 December 2016
Accepted 4 December 2016
Available online 11 December 2016

Keywords:
Aging
Cognitive decline
Episodic memory
Hippocampus
Longitudinal changes
Nonignorable attrition

ABSTRACT

There is marked variability in both onset and rate of episodic-memory decline in aging. Structural magnetic resonance imaging studies have revealed that the extent of age-related brain changes varies markedly across individuals. Past studies of whether regional atrophy accounts for episodic-memory decline in aging have yielded inconclusive findings. Here we related 15-year changes in episodic memory to 4-year changes in cortical and subcortical gray matter volume and in white-matter connectivity and lesions. In addition, changes in word fluency, fluid IQ (Block Design), and processing speed were estimated and related to structural brain changes. Significant negative change over time was observed for all cognitive and brain measures. A robust brain-cognition change-change association was observed for episodic-memory decline and atrophy in the hippocampus. This association was significant for older (65–80 years) but not middle-aged (55–60 years) participants and not sensitive to the assumption of ignorable attrition. Thus, these longitudinal findings highlight medial-temporal lobe system integrity as particularly crucial for maintaining episodic-memory functioning in older age.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Age-related decline in episodic long-term memory is typically observed after the age of 60 years (Rönnlund et al., 2005; Schaie, 1994), but there is marked variability in both onset and rate of decline (Josefsson et al., 2012). Different kinds of structural brain changes have been suggested to account for episodic-memory change in aging (Becker et al., 2015; Buckner, 2004; Hedden and Gabrieli, 2004; Nyberg and Bäckman, 2011). The hippocampus and other medial-temporal lobe (MTL) regions are vital for episodic memory (Cabeza, 2006; Dickerson and Eichenbaum, 2010; Eichenbaum, 2004; Squire, 1992). Longitudinal magnetic resonance imaging (MRI) studies have revealed marked age-related shrinkage in the hippocampus (Fjell et al., 2009; Raz et al., 2005) and hippocampus atrophy has been linked to episodic memory decline in both cross-sectional and longitudinal studies (Buckner et al., 2005; Hedden et al., 2016; Kramer et al., 2007; Persson et al., 2012; Petersen et al., 2000; Ward et al., 2015). Other studies, however, observed no relation between hippocampus atrophy and episodic-memory decline (Charlton et al., 2010; Rodrigue and Raz, 2004; see also Van Petten, 2004 for a similar conclusion based on cross-sectional data). Age-related changes outside the MTL-region have also been considered, notably in prefrontal cortex (Becker et al., 2015; West, 1996) but also in other cortical regions (Greenwood, 2000). A systematic review of brain structural correlates of successful aging confirmed the impression of variable outcomes across studies (Kaup et al., 2011) but with a fairly robust association between memory and the integrity of the hippocampal formation.

Longitudinal MRI studies have also revealed sizable white-matter (WM) alterations in aging (Madden and Parks, 2016), including loss of tract-specific integrity (Bender and Raz, 2015; Sexton et al., 2014) and accumulation of lesions (Schmidt et al., 2012b; van Dijk et al., 2008). WM lesions have been linked to memory decline in aging (Lockhart et al., 2012; Vernooij et al., 2009) but more consistently to impairments in processing speed

* Corresponding author at: Department of Statistics, Umeå School of Business and Economics, Umeå University, SE-901 87 Umeå, Sweden. Tel.: +46907865213; fax: +46907860955.
E-mail address: tetiana.gorbach@umu.se (T. Gorbach).
and fluid IQ (Prins and Scheltens, 2015; Prins et al., 2005; Raz et al., 2007). Age-related decline in WM integrity has been suggested to underlie cognitive decline in aging (Charlton et al., 2010; O’Sullivan et al., 2001). There is some evidence for a relation to episodic memory of WM integrity in the genu of corpus callosum (Bucur et al., 2008; Davis et al., 2009; Kennedy and Raz, 2009; Maddên et al., 2009; Persson et al., 2006), but some cross-sectional studies found relations to WM connectivity for processing speed but not for episodic memory (Laukka et al., 2013; Salami et al., 2012). A similar pattern was seen in a longitudinal study of very old adults for the corticospinal tract (Lövdén et al., 2014). Finally, a longitudinal study found that global changes in WM microstructure related to fluid IQ but not to memory (Ritchie et al., 2015).

Taken together, longitudinal evidence for change-change association between cognition and brain structure across the adult lifespan is scarce and it remains unclear how different kinds of brain changes in aging relate to memory change. Here we predicted that gray-matter (GM) changes in MTL/hippocampus would account for age-related changes in episodic memory, possibly along with WM changes in the genu. Using longitudinal data from the Betula project (Nilsson et al., 1997), we examined how episodic-memory change over 15 years related to 4-year changes in brain GM (cortical and subcortical) and WM (tracts and lesions). Decline in episodic memory covaries with decline in other cognitive functions (Ghisletta et al., 2012; Habib et al., 2007; Tucker-Drob et al., 2011). Therefore, to examine generality versus specificity in brain-episodic memory relations, we included tests of processing speed, word fluency, and fluid IQ (Block Design). There are some indications of stronger correlations of brain structure with cognitive performance in older than younger individuals (Kaup et al., 2011). Therefore, based on findings of average decline in episodic long-term memory after 60 years of age (Rönnlund et al., 2005; Schaeie, 1994), we report results for episodic memory for the entire age range (55–80 years) as well as after dividing the sample into younger (55–60) and older (65+) individuals.

2. Materials and methods

2.1. Participants

Participants in this study are part of the longitudinal Betula project (Nilsson et al., 1997). Within the project, 6 waves of cognitive assessment have been conducted (Fig. 1). The first wave was completed in 1988–1990 (T1), where T represents ‘time point’, followed by waves in 1993–1995 (T2), 1998–2000 (T3), 2003–2005 (T4), 2008–2010 (T5), and the sixth in 2013–2014 (T6). In this study, the sample was based on 264 healthy (no severe neurological disorders or dementia) older adults from sample 1 and sample 3 of Betula project who entered the study at T1 and T2, respectively, underwent MRI at T5 and fulfilled inclusion criteria (no MRI contraindications or artifacts on the MRI data). The age of the participants at T5-scanning ranged from 55 to 80 years (mean = 66.3, standard deviation (SD) = 7.9 years; 54.5% female), mean education level was 12.6 years (SD = 7.9 years). Of the 264 initially scanned participants, 155 remained healthy and underwent a follow-up MRI examination at T6 (hereafter referred to as healthy returnees). Of the rest, 83 participants declined to participate in T6 MRI, 9 people moved or were not contactable, 7 were scanned but not healthy at T6, and other reasons for attrition included: “MRI contraindication” (implant; 1), unknown reason (1), and death (8). Note that additional attrition (between 4 and 13) arose depending on availability of the different brain markers (Table 1).

The Betula Study was approved by the Regional Ethical Vetting Board, and written consent was obtained from all participants.

The Block Design test [while Block Design is not a pure measure of fluid intelligence, it is estimated to correlate 0.66 with full scale IQ from WAIS-IV (Groth-Marnat and Wright, 2016) and likely taps fluid processes more than crystallized ones. We therefore consider it a reasonable estimate of fluid IQ, like many prior studies (e.g., Bugg et al., 2006; Leaper et al., 2001)] from the revised form of Wechsler Adult Intelligence Scale (Wechsler, 1981) was used to assess visuospatial ability and fluid IQ. This test requires the participants to use colored blocks to recreate spatial patterns shown to them on cards. The raw score from Block Design (maximum 51) was used in the current analyses.

Word fluency was measured by 3 conditions in which participants orally generated as many words as possible, during 1 minute, satisfying the following criteria: (1) starting with the letter A; (2) 5-letter words with the initial letter M; and (3) names of professions beginning with the letter B (Nilsson et al., 1997).

The measure of processing speed was based on 3 paper-pencil tests. The first was a letter-digit substitution test requiring participants to pair letters with digits according to a letter-digit transformation key, which was given on the top of the paper form. The score was the number of correct digits that the participant managed to fill in during 1 minute (maximum 125). The second measure was letter comparison, in which participants were instructed to compare pairs of nonword strings of 3–9 letters, in order to judge
whether they were the same or different. The score was the number of correctly judged pairs during 30 seconds (out of a maximum of 21). A similar test, figure comparison, was the final measure of speed. Here participants compared pairs of abstract line figures during 30 seconds (maximum 30). All 3 versions of the tests listed all test items simultaneously on one A4-sized paper.

### 2.3. MRI data acquisition and analyses

MRI data (see Fig. 2) was collected on the same 3 T general electric scanner, equipped with a 32 channel head coil, at both baseline and follow-up. T1-weighted images were acquired with a 3D fast spoiled gradient echo sequence (180 slices with a 1 mm thickness; TR: 8.2 ms, TE: 3.2 ms, flip angle: 12°, field of view: 25 × 25 cm). To obtain measures of GM volume, the T1-weighted images from baseline and follow-up were first processed separately using the standard processing stream in FreeSurfer v.5.3 (http://surfer.nmr.mgh.harvard.edu/). Technical details of this procedure have been documented online and in previous publications (e.g., Fischl et al., 2002). Briefly, the processing includes motion correction, normalization of multiple T1-images, applying a hybrid watershed/surface deformation procedure to remove nonbrain tissue, Talairach transformation, segmentation of subcortical WM and GM structures, intensity normalization, tessellation of the GM/WM boundary, topology correction, as well as surface deformation to optimize placement of GM/WM and gray/cerebrospinal fluid border. Subsequently the images from baseline and follow-up were processed through the FreeSurfer longitudinal processing stream.

#### Table 1

<table>
<thead>
<tr>
<th>Brain markers</th>
<th>Tri</th>
<th>Cau</th>
<th>Ros</th>
<th>Sup</th>
<th>Pre</th>
<th>Inf</th>
<th>Lat</th>
<th>HC</th>
<th>PHG</th>
<th>Put</th>
<th>Genu</th>
<th>Body</th>
<th>Sple</th>
<th>WMH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N₅</strong></td>
<td>255</td>
<td>254</td>
<td>255</td>
<td>255</td>
<td>253</td>
<td>255</td>
<td>255</td>
<td>254</td>
<td>255</td>
<td>254</td>
<td>142</td>
<td>142</td>
<td>142</td>
<td>251</td>
</tr>
<tr>
<td><strong>N₆</strong></td>
<td>151</td>
<td>150</td>
<td>151</td>
<td>151</td>
<td>150</td>
<td>151</td>
<td>151</td>
<td>150</td>
<td>151</td>
<td>150</td>
<td>142</td>
<td>142</td>
<td>142</td>
<td>145</td>
</tr>
<tr>
<td><strong>Mean₅</strong></td>
<td>7258</td>
<td>12,656</td>
<td>5018</td>
<td>26,373</td>
<td>19,837</td>
<td>20,831</td>
<td>22,653</td>
<td>7829</td>
<td>4241</td>
<td>10,177</td>
<td>0.47</td>
<td>0.48</td>
<td>0.60</td>
<td>4.43</td>
</tr>
<tr>
<td><strong>Mean₆</strong></td>
<td>7091</td>
<td>12,451</td>
<td>4965</td>
<td>25,696</td>
<td>19,198</td>
<td>20,211</td>
<td>21,829</td>
<td>7672</td>
<td>4154</td>
<td>9781</td>
<td>0.43</td>
<td>0.47</td>
<td>0.59</td>
<td>5.69</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>-10.8</td>
<td>-7.1</td>
<td>-4.0</td>
<td>-9.2</td>
<td>-14.5</td>
<td>-11.8</td>
<td>-15.2</td>
<td>-8.0</td>
<td>-8.2</td>
<td>-14.6</td>
<td>-26.1</td>
<td>-12.7</td>
<td>-9.3</td>
<td>10</td>
</tr>
</tbody>
</table>

Key: Body, body of corpus callosum; Cau, caudal middle frontal; Genu, genu of corpus callosum; HC, hippocampus; Inf, inferior temporal lobe; Lat, lateral occipital cortex; Mean₅, T5 mean of corresponding brain marker measure for healthy returnees; Mean₆, T6 mean of corresponding brain marker measure for healthy returnees; N₅, number of healthy at T5 adults that have corresponding brain marker measure; N₆, number of healthy returnees that have corresponding brain marker measure; PHG, parahippocampal gyrus; Pre, precuneus; Put, putamen; Ros, rostral anterior cingulate cortex; Sple, splenium of corpus callosum; Sup, superior parietal lobe; T, one sample t-test statistic for testing: average relative difference (T6/T5) to be significantly less than 1 for GM volume and WM fractional anisotropy, and greater than 1 for WMH volumes, degrees of freedom are N₆-1 and bold values show significance at 0.1% significance level. The unit of measurement for GM volume is mm³, WMH volume = cm³, FA measures are unitless; Tri, pars triangularis of the inferior frontal gyrus; WMH, white matter hyperintensities.

![Fig. 2. Brain markers. Top left: example of white matter hyperintensities. Top right: genu (blue), body (yellow), and splenium (pink) of the corpus callosum. Bottom: lateral (left) and medial (right) views of investigated gray matter regions. Lateral view: pars triangularis of the inferior frontal gyrus (turquoise); caudal middle frontal gyrus (dark red-brown); superior parietal cortex (pale pink); inferior temporal gyrus (purple); and lateral occipital cortex (orange). Medial view: rostral anterior cingulate cortex (violet); precuneus (lavender); hippocampus (pale yellow); parahippocampal gyrus (purple); and putamen (teal).](image-url)
which creates an unbiased within-subject template image of the baseline and follow-up data, to increase reliability of the segmentation and parcelation of brain regions over time (Reuter et al., 2012). The reported data are derived from the longitudinal FreeSurfer pipeline. For the cortical regions of interests (ROIs) used in our analyses, the parcelation was based on the “Desikan-Killiany” atlas in FreeSurfer (Desikan et al., 2006), while the subcortical segmentation was based on Fischl et al. (2002).

All extracted volumetric measurements were then subjected to a quality control, which involved inspecting the following scatter plots for outliers: left versus right hemisphere measurements, baseline versus follow-up measures, as well as each regions association with chronological age for baseline and follow-up data separately. The segmentation result of each outlier was visually inspected by overlaying it on the original T1-image. Observations for which the segmentation was clearly unsuccessful (fragmented, extended, or incomplete) were excluded. The data from 1 participant was completely discarded due to multiple segmentation errors across the brain. For 5 participants, data from 1 region was discarded due to segmentation errors in that specific region. The exclusions were done prior to any statistical analyses, and the rater was blind to the cognitive performance level of the participants. Finally, GM volume for every ROI was calculated as a sum of corresponding measures in left and right hemispheres.

WM integrity was examined with diffusion tensor imaging (DTI). Images were acquired by a spin-echo-planar T2-weighted sequence, using 3 repetitions and 32 independent directions. The total slice number was 64, TR: 8000 ms, a TE: 84.4 ms, flip angle: 90°, field of view: 25 × 25 cm, and b = 1000 s/mm² and six b = 0 images.

Diffusion-weighted data analysis was performed using the University of Oxford’s Center for Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library package (http://www.fmrib.ox.ac.uk/fsl) and the tract-based spatial statistics as part of the FMRIB software package. The 3 subject-specific diffusion acquisitions were concatenated in time followed by eddy current correction to correct for head motion and eddy current distortions. Accordingly, the b-matrix was reoriented based on the transformation matrix (Leemans and Jones, 2009). DTIfit (Jenkinson and Smith, 2001) was used to fit a diffusion tensor to each voxel included in the brain mask, which was generated (using brain extraction tool) from thresholding the first volume with no gradient (b = 0). As such, voxel-wise maps of fractional anisotropy (FA) were yielded. We closely followed longitudinal tract-based spatial statistic processing stream as suggested by Engvig and colleagues (2012). The initial alignment between FA maps of 2 time points (baseline and follow-up) was carried out using linear registration algorithm (FMRIB’s Linear Image Registration Tool [FLIRT]; Jenkinson and Smith, 2001). Both FA maps were resampled into the space halfway between them, which reduces registration bias toward one of the time points. Then, subject-specific mid-space maps was generated by averaging the 2 halfway registered FA maps. All subject-specific mid-space templates were nonlinearly normalized to a standard space (FMRIB58_FA), and study-specific mean FA map was generated. Finally, the above mentioned transformation was applied to properly aligned FA maps from the 2 time points after applying a small smoothing (sigma = 2). Average FA along the spatial course of 3 segments of the corpus callosum (genu, body, and splenium), one of the major WM pathways in the brain, were computed with reference to JHU ICBM-DTI-81 WM labels, developed at Johns Hopkins University and distributed with the Functional Software Library package (Wakana et al., 2004).

T2-weighted Fluid-Attenuated Inversion Recovery (FLAIR) images were acquired with a 2D T2 FLAIR sequence (48 slices with 3 mm thickness; TR: 8000 ms, TE: 120 ms, field of view: 24 × 24 cm). Lesions were segmented by the lesion growth algorithm (Schmidt et al., 2012a) as implemented in the Lesion Segmentation Tool (LST) version 2.0.14 (http://www.statistical-modelling.de/lst.html) for SPM12. The algorithm first segments the T1 images into the 3 main tissue classes (cerebrospinal fluid, GM, and WM). This information is then combined with the coregistered FLAIR intensities in order to calculate lesion belief maps. By thresholding these maps with a prechosen initial threshold (kappa = 0.3, defined by visual inspection) an initial binary lesion map is obtained which is subsequently grown along voxels that appear hyperintense in the FLAIR image. The result is a lesion probability map that is then thresholded to obtain the binary map of lesions. The threshold was chosen as 50%, however the same result would have been obtained for any threshold between 0.4–0.99.

2.4. Scanner stability

During the time period between T5 and T6 several changes occurred in the scanner hardware, and it is important to check that these did not affect the image quality. Therefore, a quality assurance program based on Friedman and Glover (2006) study was run on a weekly basis on the scanner since November 2010 and is briefly described in the following paragraph.

For the structural data, the same T1-weighted fast spoiled gradient echo protocol as in the study was used to obtain volume data for the GE phantom. Data were thresholded well above the noise level and the selected voxels were used to calculate the volume of the phantom. The relative volume change between the beginning of the quality assurance measurements and the time for T6 data collection was 0.45%. This change is small compared to the expected average volume change in cortical regions in healthy elders, which is of the order of 2% over the time period of 4 years between T5 and T6 (Fjell et al., 2009). Furthermore, the positive change in the measured volume of the phantom suggests that the significant volumetric decline observed in the study might be slightly underestimated.

As a measure of diffusion-related scanner performance, the DTI protocol used in the study was run on the GE magnetic resonance spectroscopy phantom and the mean and standard deviation of the measured diffusion coefficient and the FA was recorded for an ROI in the center of the phantom. For the isotropic diffusion in the phantom the mean diffusion is equal to the self-diffusion coefficient of water and the anisotropy is zero. Therefore the diffusion coefficient can be used as a measure of overall gradient performance and the FA as an indicator of deviations in individual gradient channels. For the period between June 2013 and June 2014 (which corresponded roughly with the data collection at T6) the mean diffusion was 2093 μm²/s (SD = 26) and the FA was constant at 0.03 (SD = 0.01). The diffusion was comparable to the diffusion in water at the same temperature and the observed variance in the diffusion can mainly be attributed to variations in temperature of the phantom. For the typical temperatures in the scanner (21 °C–23 °C), the diffusion coefficient for water is 2077–2187 μm²/s (Holz et al., 2000). The nonzero value of FA is expected in an isotropic medium, due to the magnitude calculation of FA. This value is expected to be very sensitive to changes in performance of individual gradient amplifiers and the constant value indicates that the gradient output was not deviating from the nominal values. Taken together, the stable results from the DTI scan on the magnetic resonance spectroscopy phantom is characteristic of a well performing scanner.

2.5. Selection of regions of interest

GM ROIs for the current analyses were selected based on previously established links between regional volume or GM integrity and cognitive function from the literature and particularly so for the aging brain. To be able to probe the specificity and generality of
brain–cognition relationships, ROIs were chosen so that all major lobes of the brain were represented. The chosen regions included the ventrolateral (pars triangularis) and dorsolateral (caudal middle frontal) aspects of the lateral frontal cortex, as well as the rostral anterior cingulate cortex (Jung and Haier, 2007; Pardo et al., 2007; West, 1996). Temporal lobe ROIs included the hippocampus (e.g., Kramer et al., 2007; Persson et al., 2012; Petersen et al., 2000) and parahippocampal gyrus (Köhler et al., 1998; Pantel et al., 2003) medially, as well as the inferior temporal cortex (Murphy et al., 2010; Tisserand et al., 2004). The superior parietal cortex (Jung and Haier, 2007) and precuneus (Cavanna and Trimble, 2006) were chosen from the parietal lobe. As for subcortical structures, the putamen (de Jong et al., 2008; Hedden et al., 2016) was selected, whereas the caudate nucleus was omitted due to high between-person variance, coupled with a high incidence of outliers that was found to be driven by poor segmentation results for older individuals with enlarged ventricles. Finally, the lateral occipital cortex was examined (see McDonald et al., 2009).

2.6. Defining cognitive and brain change

Concerning retests effects, Rabitt et al. (2004) argued that gains due to practice effects for cognitive testing tend to be greatest between the first and second testing occasions and more moderate afterward. Thus, we considered data starting from the second measurement occasion (or later) for each individual. Since only part of the investigated sample was examined for the first time at T2, and in order to investigate equal time span for all individuals, we considered scores from T3 to T6 for episodic memory, word fluency and fluid IQ (Block Design). For processing speed, measurements started at T4, that is, at most 3 measurements were available and all were used in the analysis.

In order to adjust for different scales, individual tasks were standardized using mean and standard deviation of the corresponding task scores from the third Betula wave (T3). The sum of corresponding standardized scores was used as a composite score for each cognitive domain and time point.

Test-retest reliability of the episodic memory, word fluency, Block Design and processing speed composite scores were respectively, 0.76, 0.75, 0.81, and 0.84, as assessed by Pearson correlations between scores of the fourth and fifth Betula waves (measurement occasions with maximal number of available data for each domain) based on the participants included in the current analysis. We refer the reader to Supplementary Material for the presentation of internal consistency and reliability of individual measures included in the composites.

Change was estimated individually for each subject and cognitive domain as the ordinary least squares slope estimate in a linear model used for cognitive change estimation is described in the Supplementary Material. Brieﬂy, the sensitivity analysis assessed whether dependence between attrition and brain marker change, given cognition change, age, and hypertension, would alter the p-values obtained in the complete cases analysis. If assuming such dependency does increase the latter p-value above the significance level chosen, then the complete case analysis is sensitive to the assumption of ignorable attrition.

3. Results

3.1. Cognitive changes

Fig. 3 presents mean change functions across age together with individual change. As can be seen for all cognitive functions, there were marked individual differences in both level and slope. For processing speed, the decline occurred from the youngest recordings at the age of 50 years. For fluid IQ/Block Design, a similar negative trend was seen. The memory-age function was, as expected, curve-linear (the hypothesis of linearity is rejected using likelihood ratio test, see Supplementary Material) with accelerated onset of decline at around the age of 65 years. Equivalently, onset of decline occurred late, with stability characterizing the period from 45 to 70 years of age. Statistically, a significant decline was observed for processing speed (t_{262} = −18.15, p < 0.001), episodic memory (t_{262} = −7.70, p < 0.001), fluid IQ/Block Design (t_{263} = −12.18, p < 0.001), and word fluency (t_{263} = −3.89, p < 0.001).

3.2. Brain changes

There was significant mean atrophy for all 10 cortical and subcortical regions (Table 1). The global nature of changes is on par with previous findings (e.g., Walhovd et al., 2005). For the hippocampus, the average onset of atrophy was after the age of 60 years (Fig. 4). Significant change was also observed for WM lesions (Fig. 4) and for FA in the genu, body, and splenium of the corpus callosum (Table 1, Fig. 4).
3.3. Linking brain-cognition changes

All brain-cognition change-change correlations are presented in Fig. 5, Supplementary Fig. 1 and Supplementary Table 2. For episodic memory, significant at 5% level associations were found for 3 GM markers: hippocampus (Fig. 6), parahippocampal gyrus, and the lateral occipital cortex. In the age-stratified analyses, all of these correlations remained significant at the 5% level in the older (65+ years) group, but they were nonsignificant in the younger subgroup. The main result of significant episodic memory—hippocampus correlation for older but not younger adults holds true even when division 55–65 and 70- to 80-year old is used. Only the hippocampus-episodic memory association remained significant after Bonferroni correction. No significant associations to episodic memory were found for the WM brain markers of corpus callosum.

For the other cognitive domains, associations to structural brain changes were generally weak and nonsignificant. For word fluency, an association was observed to atrophy in the parahippocampal gyrus, but this relation did not survive Bonferroni correction. No significant associations were seen for perceptual speed or Block Design.

3.4. Sensitivity analyses

The healthy returnees were younger than the dropouts (mean age at T5 = 64.97 vs. 68.41 years; $t_{205.7} = -3.39, p < 0.001$) and more educated (mean education = 13.1 vs. 11.91 years; $t_{232.8} = 2.30, p = 0.02$).

Sensitivity analyses were conducted for the significance at 5% level findings (episodic memory change—hippocampus, parahippocampal gyrus, and lateral occipital cortex change, fluency—parahippocampal gyrus change) and revealed that the partial correlations estimated remained significant at the 5% level when taking into account plausible (positive, see Supplementary Material for more details) dependency between attrition and brain-marker change.

4. Discussion

The main purpose of this study was to relate age-related episodic-memory change to structural brain changes. Episodic memory was measured longitudinally over 15 years, and was found to show average onset of decline around the age of 65 years. This change function differed from the other 3 examined cognitive functions. For verbal fluency, change was modest and only observed around the age of 70 years. By contrast, an earlier onset of change was seen for the measures of speed and fluid IQ/Block Design. The estimates of average onset of decline should be considered as approximations, as the data were not corrected for practice effects and the sample was most likely higher performing than the general population (Pudas, 2013). Still, with this caveat in mind, the
observed patterns for the different cognitive functions are in good agreement with past longitudinal observations (see Schaie, 1994).

A noteworthy feature of the present cognitive and brain data is the existence of marked individual differences. Thus, for all examined cognitive functions (Fig. 3) as well as brain measures (Fig. 4) the extent of age-related differences and changes varied markedly across individuals (Lindenberger, 2014). Here we capitalized on this variability and analyzed, across individuals, the relations between rates of change in brain and cognition.

As predicted, atrophy in the hippocampus was related to episodic-memory decline. This was the strongest observed association in the present set of analyses, and it survived correction for multiple comparisons. A relation between hippocampus atrophy and episodic memory decline is in keeping with the findings from several previous studies (Buckner et al., 2005; Hedden et al., 2016; Kaup et al., 2011; Kramer et al., 2007; Persson et al., 2012; Petersen et al., 2000; Ward et al., 2015). Critically, the present findings were based on 15-year changes in episodic memory and 4-year changes in hippocampus volume, which extends past, typically cross-sectional, observations.

The remaining brain-cognition change-change relations were weak and nonsignificant. The absence of relations to cognition for the DTI measures of WM tract integrity is particularly noteworthy in view of several previous reports, especially for speed (Hedden et al., 2016; Laukka et al., 2013; Lövén et al., 2014; Madden et al., 2004; Penke et al., 2010; Persson et al., 2006; Salami et al., 2012; Sullivan et al., 2001; Turken et al., 2008). However, the lack of association to measures of higher-order cognition (episodic, fluency, Block Design) is in line with some past cross-sectional (Salami et al., 2012) and longitudinal (Lövén et al., 2014) findings, suggesting that WM microstructure is not a major brain correlate of cognitive decline in normal aging. In this study, the change-change association between the WM and cognition was only investigated for the corpus callosum. Thus, the lack of change-change association does not rule out the possibility of an association between cognition decline and changes in other WM tracts. It is also important to note that FA changes may not only reflect alterations in WM microstructure (e.g., integrity of axonal connections) but may also express changes in WM macrostructures (Jones, 2010).

When analyzed separately for the 2 age-groups, the observed brain-memory relations consistently were in the direction of a significant relation for the older but not the younger individuals. This direction of effect is in line with limited evidence from past studies of relationships across adulthood (Kaup et al., 2011). A likely explanation for this difference is that both brain markers and cognitive measures show more pronounced change after the age of 65 years (cf., Figs. 3 and 4), which enables detection of change-change associations.

Fig. 4. Individual trajectories for brain markers change (for 142–150 participants based on availability of brain marker measures). The bold gray line indicates mean change, estimated using a Generalized Additive Mixed Model (Wood, 2006).
From a methodological perspective, there is a necessity to conduct a sensitivity analysis for nonignorable attrition since we indeed expect attrition to be positively correlated with brain marker deterioration given observed covariates. As a result, we find that the statistically significant results based on complete case analysis are not sensitive to the plausible existence of dependence between attrition and brain marker changes. This means that significant results are not driven by nonignorable attrition effects. It should also be noted here that we can not rule out that we miss brain-cognition change relations due to nonignorable attrition. Furthermore, we could have controlled for baseline cognition and/ or baseline brain marker in estimation of associations, although as noted by Glymour et al. (2005) this may introduce spurious correlations due to confounding. However, note that the trend for stronger association in later adulthood holds even when additionally controlling for baseline cognition.

The strengths of the current study are longitudinal assessments of multiple cognitive domains (with several measures per each domain), diverse brain imaging markers, and relatively large and well-characterized age-heterogeneous sample. To date, very few longitudinal studies investigated change-change associations using such comprehensive data, and to our knowledge, none have taken into account nonignorable attrition.

A possible limitation of this study is that the relatively large size of the sample necessitated the use of automated procedures for extraction of regional brain volumes, which may be less accurate than manual segmentation (e.g., Wenger et al., 2014). Due to this, we performed a thorough quality control and excluded outliers in which the volumetric estimation had clearly failed. Furthermore, we were only interested in within-person volumetric changes, which were estimated with the longitudinal image processing stream with improved reliability (Reuter et al., 2012). Indeed, our data did show good test-retest reliability (>0.96). However, some noise introduced by the automated volumetric extraction was still likely present in our data. This may have made it harder to detect significant relationships with cognitive change and possibly caused a slight underestimation of the effects we did obtain.

The findings from the present longitudinal dataset highlight the MTL system as the primary pathway of episodic-memory decline also in normal, nondemented aging. This is consistent with...
suggestions of a continuum between normal cognitive aging, mild cognitive impairment, and Alzheimer’s disease (Devanand et al., 2007; Petersen et al., 2000; Stoub et al., 2010). Strong conclusions in this regard may, however, require more detailed analyses of MTL and hippocampal subregions (Frisoni et al., 2008; La Joie et al., 2013; Perrottin et al., 2015; Ward et al., 2015). More generally, our findings can be related to the notion of brain maintenance (Nyberg et al., 2012). By this view, individual differences in the manifestation of brain pathology explain why some but not other individuals show cognitive decline. The present findings demonstrate that maintaining the hippocampal/MTL system is a key to preserved episodic long-term memory in aging.

Disclosure statement

The authors have no actual or potential conflicts of interest.

Acknowledgements

This work was supported by Swedish Research Council (grant number 340-2012-5931), Knut and Alice Wallenberg Foundation (Wallenberg scholar grant to L.N.), and Ragnar Söderberg’s Foundation (grant number KVA/2011/88/65 to L.N.). The FreeSurfer analyses were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at HPC2N in Umeå. The authors thank Minna Genbäck for providing R code used for the sensitivity analysis and Micael Andersson for running the FreeSurfer analyses.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neurobiolaging.2016.12.002.

References


