Journal of Alzheimer's Disease 33 (2013) 451–462
DOI 10.3233/JAD-2012-121315
IOS Press

Prediction of Amyloid-β Pathology in Amnestic Mild Cognitive Impairment with Neuropsychological Tests

Alex Bahar-Fuchs[a,b,*], Victor Villemagne[a,c], Kevin Ong[a], Gaël Chetélat[d], Fiona Lamb[a], Cornelia B. Reininger[e], Michael Woodward[f] and Christopher C. Rowe[a]
[a]*Department of Nuclear Medicine, Centre for PET, Austin Health, Melbourne, VIC, Australia*
[b]*Centre for Research on Ageing, Health, and Wellbeing, Australian National University, Canberra, ACT, Australia*
[c]*Mental Health Research Institute, Parkville, VIC, Australia*
[d]*Inserm- EPHE-Université de Caen/Basse-Normandie, Unité U923, GIP Cyceron, CHU Cote de Nacre, France*
[e]*Bayer Pharma AG, Berlin, Germany*
[f]*Aged Care Services, Austin Health, VIC, Australia*

Accepted 9 August 2012

Abstract. Assessment of disease biomarkers, particularly the *in vivo* assessment of amyloid-β (Aβ) burden with positron emission tomography (PET), is gradually becoming central to the diagnosis of mild cognitive impairment (MCI) due to Alzheimer's disease (AD). However, the incorporation of biomarker evidence to the diagnostic process is currently restricted mainly to research settings. The identification of memory measures that are associated with Aβ is of clinical relevance as this may enhance the confidence in making a diagnosis of MCI due to AD in clinical settings. Forty one persons with amnestic MCI underwent Aβ imaging with ^{18}F-Florbetaben PET, magnetic resonance imaging, and a comprehensive neuropsychological assessment. All measures of episodic memory were significantly correlated with Aβ burden, but regression analyses revealed a strong and selective association between story recall and Aβ over and beyond the effects of age, education, global cognition, hippocampal volume, or other memory tests. Analyses of sensitivity and specificity of memory measures to detect high versus low Aβ scans suggested that word-list recall performed better when high sensitivity was preferred, whereas story recall performed better when high specificity was preferred. In conclusion, a measure of story recall may increase the confidence in making a diagnosis of MCI due to AD in clinical settings.

Keywords: Aging, Alzheimer's disease, amyloid imaging, ^{18}F Florbetaben PET, memory, mild cognitive impairment

INTRODUCTION

Dementia, a mostly age-related clinical syndrome characterized by the gradual loss of cognitive, intellectual, and functional abilities, represents one of the most pressing health challenges of our time. Although Alzheimer's disease (AD), the most common cause of dementia in those over 65 years, was first described more than a century ago, it was mainly in the past two decades that the most significant breakthroughs in our understanding of clinicopathological and genetic aspects of AD occurred. It is well established now that AD pathology is present many years before the onset of detectable clinical symptoms. In recent years, various techniques emerged that allow the detection of disease biomarkers such as amyloid-β (Aβ) burden *in vivo*. The ability to detect AD pathology before the

*Correspondence to: Alex Bahar-Fuchs, Centre for Research on Ageing, Health, and Wellbeing, Australian National University, Canberra, ACT 0200, Australia. Tel.: +61 2 61259705; E-mail: alex.baharfuchs@anu.edu.au.

ISSN 1387-2877/13/$27.50

development of obvious clinical symptoms is likely to be of particular relevance when future disease-modifying therapies become available as these are most likely to be effective at earliest stages of the disease. The major advances in AD research are reflected in the recent publication of the first revision of the original diagnostic criteria for AD [1]. Building on and extending the work the International Working Group [2, 3], the research of three working parties, established by the Alzheimer's Association and the National Institute on Aging (NIA), has recently culminated in the publication of three separate guidelines for the diagnosis of AD at three phases: preclinical, mild cognitive impairment (MCI), and dementia [4–6]. These reports draw distinctions between the pathophysiological (AD-P) and clinical (AD-C) aspects of AD, and between the evidence for AD that can be used in clinical settings versus research settings.

An imbalance between the production and clearance of Aβ that leads to its extracellular accumulation in plaques is widely regarded as the initial event in a cascade of pathological processes that lead to synaptic dysfunction and neuronal death which is followed by the development of cognitive impairment and eventually dementia [5]. Data from postmortem studies [7] and from *in vivo* assessment of Aβ show that high levels of Aβ are present in 30%–40% of apparently healthy older adults [8, 9]. In individuals meeting criteria for MCI (particularly the amnestic type), between 50% and 60% show high Aβ burden [10], and while not all individuals with high Aβ burden progress to Alzheimer dementia, virtually all individuals with Alzheimer dementia have high Aβ burden [11]. These observations contributed to an increasing consensus that Aβ is a necessary but insufficient factor in the development of Alzheimer dementia [12].

The recently published criteria for MCI due to AD recommend that the incorporation of biomarker evidence of AD pathology to increase the certainty in the diagnosis be restricted to research settings, whereas the core criteria for use in clinical settings will continue to rely primarily on careful history gathering and the results of objective cognitive measures [4]. This recommendation reflects in part the limited availability of biomarker evidence, particularly in respect to cerebrospinal fluid sampling and molecular neuroimaging of Aβ burden *in vivo* [4].

The identification of reliable clinical correlates of Aβ burden in persons with amnestic MCI is therefore of considerable interest, as these may enhance confidence that AD is the underlying etiology in clinical settings. Research in recent years had begun to address the nature of the concurrent relationship between Aβ and cognition. Whereas this literature is usually focused on the impact of Aβ on cognition, the ability of cognitive measures to reliably predict elevated Aβ levels is a question of clinical interest.

Data from postmortem studies usually support the presence of a strong relationship between cognitive functions and neurofibrillary tangles, especially in the area of episodic memory [13]. Furthermore, memory function has an established association with hippocampal atrophy, which is believed to indirectly reflect tangle-related neuronal death. Vascular pathology, as reflected in white matter hyperintensities (WMH), has been shown to be associated with cognitive compromise, particularly affecting performance on working memory and executive function, as well as visuospatial abilities among people with MCI [14]. The extent to which Aβ is associated with cognition is, however, less clear. Although the association between cognition and Aβ in studies in which participants across the spectrum of cognitive and functional ability are combined tends to be significant, the nature of this relationship among non-demented older adults remains inconclusive.

In apparently healthy older adults, the evidence regarding the relationship between Aβ burden and episodic memory is mixed. Preliminary reports observed a moderate relationship [15] between episodic memory and Aβ burden as reflected in Pittsburgh Compound B retention. However, other studies failed to replicate this finding [16–18], and more recently it has been shown that the association between Aβ burden and memory in healthy older adults may be moderated by cognitive reserve [19].

In persons with MCI, the results more uniformly point to the presence of a relationship between Aβ and episodic memory function [15, 20]. However, the results of one study suggest that this relationship may be mostly mediated by hippocampal volume [21]. This finding is consistent with an earlier histopathological study in which a significant association between Aβ and cognition was eliminated once density of neurofibrillary tangles was controlled [22]. Recently, Chételat et al. showed that, unlike the case for global Aβ deposition, Aβ burden in the temporal cortex continued to be independently associated with episodic memory in persons with MCI and in clinically healthy older adults [23]. Performance in domains other than episodic memory has generally been found to show little relationship with Aβ burden [15, 17].

Importantly, in all studies mentioned, the association between Aβ and memory was usually examined using a composite memory score, which was calculated from scores on delayed recall trials of two or more individual memory tests [15, 18]. However, in clinical settings where individuals with suspected MCI are usually assessed, a variety of memory measures are used by clinicians, and it is not clear whether specific memory measures are better associated with Aβ. Identification of specific memory measures that show a strong association with Aβ burden may increase the confidence in making the diagnosis of MCI due to AD in clinical settings. Furthermore, in cognitive evaluations of persons with MCI, the assessment of memory recall following a delay (long-term memory) is usually emphasized by clinicians over the assessment of immediate (short-term) memory. However, whether Aβ is better correlated with immediate or with delayed memory recall is not known. This point is of clinical relevance as if Aβ shows an association of a similar magnitude with immediate and with delayed memory measures, scores on immediate recall trials of memory measures may be sufficient to allow clinicians make predictions regarding the presence of Aβ.

Among the most commonly used measures of memory in neuropsychological evaluations of older persons are tests assessing word-list recall, paragraph recall, and memory for visuospatial information. In the current study, we investigated the relationship between individual cognitive measures and Aβ burden, hippocampal atrophy, and WMH and evaluated the extent to which such measures can be used in clinical settings to increase confidence in the diagnosis of MCI due to AD.

MATERIALS AND METHODS

Participants

Forty-five older adults who were referred from local memory clinics and who met consensus criteria for MCI at the time the study was undertaken [24] enrolled in the current study. Specifically, all participants or their next of kin reported a history of cognitive decline, while remaining relatively independent in their activities of daily living. The presence of cognitive impairment and relative functional independence were subsequently confirmed by a neuropsychological assessment (described below). On the basis of the neuropsychological assessment, four participants were classified as having a non-amnestic MCI. Because individuals with non-amnestic MCI are less likely to progress to AD [25], these four individuals were excluded from the current analyses, which focused on the remaining 41 patients. The following additional criteria had to be met for inclusion in the study: Participants had to be at least 60 years old and to have had at least 7 years of formal schooling. They were also required to communicate fluently in English and to have no contraindications to undergo an MRI scan. A further inclusion criterion was the availability of a reliable next-of-kin who also agreed to participate in the study and provide a collateral history. Exclusion criteria included the presence of dementia, a score lower than 23 on the Mini-Mental Status Examination (MMSE) [MMSE; 26], and being unable to give informed consent. In addition, participants were excluded if they presented with other conditions that may impair their cognition and independence, including other neurological (stroke, multiple sclerosis, epilepsy, moderate-severe traumatic brain injury), psychiatric (psychotic symptoms, bipolar disorder), or substance use conditions (e.g., drug and alcohol dependence). Participants who were on stable low doses of psychotropic medication were not excluded from the study; however, participants were excluded if they were already prescribed acetylcholinesterase inhibitors or memantine.

The study was approved by the Austin Health ethical review board, and all participants provided written informed consent prior to being screened for participation.

Neuropsychological testing

All participants underwent a neuropsychological evaluation by a neuropsychologist who was blinded to the imaging results. Apart from the addition of the MMSE and the Clinical Dementia Rating scale (CDR) [CDR; 27], the neuropsychological examination was identical to that described in Pike et al. [15], and included: Boston Naming Test (BNT-30 item version) [BNT-30 item version; 28], Digit Span and Digit Symbol-Coding subtests from the Wechsler Adult Intelligence Scale-Third edition [WAIS-III; 29], California Verbal Learning Test-Second edition (CVLT-II) [CVLT-II; 30], Logical Memory (LM) immediate and delayed recall from the Wechsler Memory Scale-Third edition[1] [WMS-III, Story A only; 31, 1997], Rey

[1] For consistency with the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Australian Imaging and Biomarkers Study of Ageing (AIBL), the protocol for administration of story recall was modified and included story A only, which was only read once. Recall was tested immediately and following a 25 min delay.

Complex Figure Test (RCFT) [RCFT; 32], and the Verbal Fluency subtest of the Delis–Kaplan Executive Function System [D-KEFS; 33]. A composite episodic memory score was calculated by computing the average of the z scores for the RCFT, CVLT-II, and LM long delayed recall trials. A composite non-memory score was calculated by computing the average of the z scores for Letter and Category Fluency, BNT, Digit Span, Digit Symbol-Coding, and RCFT copy. The neuropsychological evaluation was conducted up to 4 weeks prior to the ^{18}F-Florbetaben PET scan.

Image acquisition

MRI

MR imaging was performed prior to the PET scan and consisted of a 3D T1-weighted magnetization prepared rapid gradient echo (MPRAGE) sequence and a fluid attenuated inversion recovery (FLAIR) sequence. MRI was done for screening and subsequent co-registration with the PET images.

^{18}F-Florbetaben Imaging

Labeling was done in the Austin Health Centre for PET, as previously described [34]. Mean specific activity at the time of injection for MCI was 60 ± 29 GBq/μmol. Imaging was performed with a 3D GSO Phillips Allegro PET camera. A 2-min transmission scan using a rotating 137Cs source was done for attenuation correction immediately prior to scanning. Each participant received on average 286 ± 19 MBq of ^{18}F-Florbetaben IV over 38 ± 17 s. Images obtained between 90–110 min post injection were analyzed. Images were reconstructed using a 3D RAMLA algorithm.

Image analysis

All image analyses were conducted by persons blinded to the clinical status and cognitive test results of participants.

Hippocampal volumes were obtained by a commercial, fully automated volumetric measurement program (NeuroQuant®) that was applied to the 3D MP-RAGE MRI data [35]. For cross validation, individual hippocampal volumes were compared with volumes obtained by manual extraction using the VBM tool from SPM5, which was performed in 38 of the amnestic MCI participants with suitable MRI data ($r = 0.77$) [36]. Preprocessing of the FLAIR images was performed to correct for bias field effects and remove noise using anisotropic diffusion prior to manual segmentation of deep WMH. Manual segmentation of the WMH was performed using the MRICRO software by P.R. The total WMH volume in each MCI subject was calculated, as well as the number of individual lesions.

PET images were processed with a semi-automatic volume of interest (VOI) method. This method used a preset template of narrow cortical VOI that were applied to either the spatially normalized ^{18}F-Florbetaben scan or via placement on the subject's spatially normalized co-registered MRI by a single operator (VLV) blind to the subject's clinical status. Minor manual adjustments on the MRI were made to ensure that overlap with white matter and cerebrospinal fluid was minimized. Spatial normalization and co-registration of the PET and MRI images was performed using SPM8 [37]. Mean radioactivity values were obtained from VOI for cortical, subcortical, and cerebellar regions. The cerebellar cortical VOI were placed taking care to avoid cerebellar white matter. No correction for partial volume effect was applied to the PET data.

The standardized uptake value (SUV), defined as the decay-corrected brain radioactivity concentration normalized for injected dose and body weight, was calculated for all regions. These were then used to derive the SUV ratio (SUVR), which was referenced to cerebellar cortex. Neocortical Aβ burden was expressed as the average SUVR of the area-weighted mean for the following cortical ROIs: frontal (consisting of dorsolateral prefrontal, ventrolateral prefrontal, and orbitofrontal regions), superior parietal, lateral temporal, lateral occipital, and anterior and posterior cingulate. In order to identify a SUVR 'cut-off', a hierarchical cluster analysis of the neocortical SUVR of ^{18}F-Florbetaben scans in healthy control participants was performed as previously described [38], yielding a cut-off for "high" or "low" neocortical SUVR of ≥ 1.45.

Statistical analyses

Data was first screened for missing data and the presence of outliers, which showed that LM scores were missing for one participant, and that RCFT delayed recall score was missing for another participant. Therefore, the composite memory score for these two participants was based on two memory measures rather than three. Standard scores on each of the neuropsychological tests were calculated using the means and standard deviations on these same

measures from a carefully selected cohort of 45 healthy aged-matched older participants taken from the Australian Imaging Biomarkers and Lifestyle (AIBL) Study of Ageing [39]. This control group comprised individuals who all met the following criteria: negative Aβ scans, no global or hippocampal atrophy, MMSE >28, CDR = 0, CDR Sum of Boxes = 0, mean Geriatric Depression Scale = 1 ± 1.4, and mean abdominal circumference = 92 ± 13. Thus the reference group consisted of individuals with no indicators of disease that are likely to impact on cognitive function. Pearson's correlation coefficients were calculated to assess the degree of linear relationship between neuropsychology test scores and neocortical SUVR, hippocampal volume, and WMH. Step-wise hierarchical regression models were built with neocortical SUVR as the dependent variable, and scores on memory measures as the predictor variables. To assess the relationship between Aβ burden and memory beyond the impact of other variables, in all models, age, education, MMSE, and hippocampal volume entered the model first. Receiver operating characteristic (ROC) analyses were performed to determine cut-off for each memory test associated with various sensitivity and specificity values to detect high or low Aβ burden. Data are presented as mean $\pm$ SD, unless otherwise stated.

RESULTS

Table 1 shows participants' means (SE), standard scores, and range on demographic and neuropsychological measures, as well as their association with neocortical SUVR, hippocampal volume, and WMH. As expected, this sample of amnestic MCI patients was impaired on all measures of learning and memory as reflected in the standard scores. With respect to non-memory measures, with the exception of scores on the BNT, performance on these tests was generally preserved (Table 1). Age was moderately associated with hippocampal volume and WMH, whereas neither of the demographical variables was related to neocortical SUVR values. Examination of the linear relationship between performance on cognitive measures and neocortical SUVR revealed moderate to strong relationships between the composite memory score, as well as each of the individual measures of episodic memory and neocortical SUVR (Fig. 1). While the association between hippocampal volume and the composite memory score remained significant after controlling for the effects of age on hippocampal volume, the delayed recall of a figure (RCFT) was the only memory measure that correlated with hippocampal volume. Cognitive measures other than memory generally showed a weak association with

Table 1
Means, SE, range, and standard scores on demographic and neuropsychological measures, together with linear correlation coefficients with Aβ SUVR, WMH, and with hippocampal volume

	Mean (SE)	Standard (z) score	Range	SUVR (Pearson *r*)	WMH (Pearson *r*)	Hipp vol (Pearson *r*)
Age	72.5 (1.0)	–	60–85	0.00	0.36*	–0.45**
Education (y)	13.5 (0.6)	–	7–25	–0.13	–0.19	–0.26
% Male	61	–	–	0.09	–0.07	–0.18
MMSE	27.3 (0.3)	–2.7	24–30	–0.38**	0.17	0.20
CVLT-II total	34.0 (1.6)	–2.3	17–61	–0.41**	–0.16	0.23
CVLT-II short delay	5.2 (0.6)	–2.5	0–15	–0.33*	–0.08	0.26
CVLT-II long delay	4.9 (0.6)	–3.4	0–15	–0.39**	–0.13	0.29
RCFT copy	29.6 (1.0)	–0.1	2–36	–0.10	–0.33*	–0.05
RCFT immediate	11.1 (1.0)	–1.4	0–31	–0.35*	–0.24	0.28
RCFT delay	10.4 (1.0)	–1.4	0–28	–0.42**	–0.28	0.47**
LM immediate	7.8 (0.7)	–1.3	0–17	–0.51**	0.1	0.14
LM delay	5.2 (0.7)	–1.9	0–13	–0.58**	0.09	0.09
(Letter fluency- F,A,S)	36.3 (1.6)	–0.4	9–67	0.14	–0.13	0.04
BNT	25.3 (0.5)	–1.7	16–30	0.07	–0.40**	–0.06
Category fluency	34.0 (1.5)	–0.8	17–60	0.1	–0.11	0.06
Digit span	16.4 (0.4)	–0.6	9–21	–0.00	–0.00	–0.37*
Digit-symbol coding	46.8 (2.1)	–0.9	21–76	–0.09	–0.21	–0.03
Composite memory	–	–2.5	–4.1–(–0.3)	–0.60**	–0.14	0.32*
Composite non-memory	–	–0.9	–3.0–0.5	–0.00	–0.38**	–0.09

*$p<0.05$, **$p<0.01$; Z scores calculated using a normative sample of 45 healthy older adults from the AIBL study. See Statistical Analysis section for details of this sample. Correlations shown between standardized uptake value ratio (SUVR) and cognitive measures are uncorrected, whereas correlations between white matter hyperintensities (WMH) and hippocampal volume with cognitive measures are corrected for age; BNT, Boston Naming Test; CVLT-II, California Verbal Learning Test-Second edition; LM, Logical Memory immediate and delayed recall; MMSE, Mini-Mental Status Examination; RCFT, Rey Complex Figure Test.

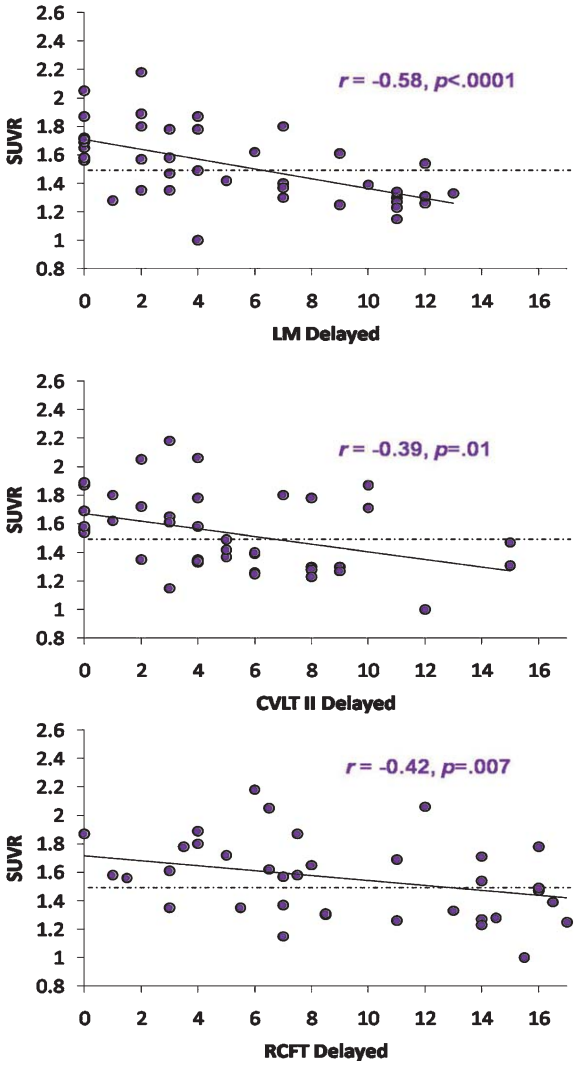

Fig. 1. Linear relationships between standardized uptake value ratio (SUVR) and three measure of delayed memory recall. High and low Aβ are separated by the dotted line (SUVR = 1.45).

either SUVR or hippocampal volume. However, the composite non-memory score was significantly associated with WMH.

As can be seen in Table 1 and Fig. 1, LM II showed the strongest association with neocortical SUVR, while showing the lowest association with hippocampal volume. To explore the extent to which the three memory measures independently explained variance in neocortical SUVR scores, three hierarchical regressions models were built. In each of these models, age and education entered the model first, followed by the MMSE score as a measure of global cognition. Hippocampal volume entered the models next. The three models subsequently differed in the order in which the three memory measures entered them. In each of these models, two memory measures entered the model together, with the third memory measure entering the model last. Table 2 shows a summary of these procedures. As can be seen, age and education explained minimal variance in neocortical SUVR scores, and although the change in the amount of explained variance when MMSE score was added to the model was significant, the model as a whole remained non-significant. As can be further noted, the addition of hippocampal volume to the model at the next step did not explain additional variance in SUVR. The point at which the model reached significance in the prediction of Aβ varied with the step in which LM II score entered the model. Specifically, the first regression model did not reach significance when both the CVLT-II and the RCFT entered the model, but when LM II was added to the model in the next step, it reached significance. In the other two models, once LM II entered the regression in the fourth step, no additional improvement in the amount of variance explained was achieved with the addition of the remaining memory measure in the final step.

As expected, for each memory measure, the relationship between scores on the immediate and the delayed recall trials was very strong ($r = 0.86$, $r = 0.94$, and $r = 0.88$ for the CVLT-II, RCFT, and LM, respectively). It is therefore possible that the associations that were observed between neocortical SUVR and the delayed recall trials of the three memory measures merely reflected the scores obtained on immediate recall of these measures. To explore this possibility, the relationship between neocortical SUVR and the delayed recall trials of the three memory measures was examined again while controlling for immediate recall scores in partial correlation analyses. Interestingly, while the relationship between neocortical SUVR and delayed recall on both the CVLT-II and RCFT became weak and non-significant ($r = –0.22$, *ns*; $r = –0.09$, *ns* respectively), the relationship with the delayed recall of LM remained significant ($r = –0.32$, $p = 0.04$).

Consistent with other studies, 56.1% ($n = 23$) of persons with amnestic MCI in the current study presented with high Aβ burden (neocortical SUVR >1.4). There were no differences between participants with high or low Aβ burden in terms of age, education, hippocampal volume, or WMH. MCI participants with high Aβ burden obtained a significantly lower composite memory score than did participants with low Aβ burden ($t(39) = 5.0$, $p < 0.001$, $d = 1.6$). Examination of group differences on the specific memory measures

Table 2
Results of three hierarchical regression models with Aβ as the dependent variable. After age, education, Mini-Mental Status Examination (MMSE), and hippocampal volume entered the model as predictors in each model, the three memory tests entered the model. In each model, a different memory test entered the model last

Step		B	β	R	R^2	ΔR^2	p (model)
1	Age	0.001	0.02	0.13	0.02	0.02	ns
	Education	–0.01	–0.14				
2	MMSE	–0.05	–0.38**	0.40	0.16	0.14*	ns
3	Hippocampal volume	–0.06	–0.12	0.41	0.17	0.01	ns
Regression Model 1: LM delay entered last							
4	CVLT delay	–0.02	–0.23	0.54	0.29	0.12	0.05
	RCFT delay	–0.01	–0.26				
5	LM delay	–0.03	–0.47**	0.65	0.42	0.13**	0.01
Regression Model 2: CVLT-II delay entered last							
4	RCFT delay	–0.00	–0.13	0.63	0.40	0.23**	0.006
	LM delay	–0.03	–0.50**				
5	CVLT-II delay	–0.01	–0.16	0.65	0.42	0.02	0.008
Regression Model 3: RCFT delay entered last							
4	CVLT delay	–0.01	–0.18	0.64	0.41	0.24**	0.004
	LM delay	–0.03	–0.50**				
5	RCFT delay	–0.00	–0.08	0.65	0.42	0.00	0.008

*p<0.05, **p<0.01; CVLT-II, California Verbal Learning Test-Second edition; LM, Logical Memory delayed recall; RCFT, Rey Complex Figure Test.

showed that MCI participants with high Aβ burden scored lower than those with low Aβ burden on the CVLT-II delayed recall ($t(39) = 2.7$, $p < 0.01$, $d = 0.9$, RCFT delayed recall, $t(39) = 3.2$, $p < 0.01$, $d = 0.89$, and LM II, $t(38) = 4.8$, $p < 0.001$, $d = 1.6$). No group differences were observed on the non-memory composite score, or on any of the specific non-memory measures.

To explore the sensitivity and specificity of different memory test cut-off scores to detect a positive versus negative scans, ROC analyses were performed. A visual inspection of the ROC curve depicted in Fig. 2 shows that LM II was overall more accurate than RCFT or CVLT-II in predicting whether a given scan revealed high or low, Aβ burden and this was confirmed by calculation of the area under the curve (Table 3). Table 3 also provides cut-off and classification accuracy values associated with 90% sensitivity to detect a high Aβ scan and with 90% specificity to detect a low Aβ scan. When the degree of test sensitivity to detect a high Aβ scan was set high (90%), CVLT-II was associated with greater specificity (i.e., fewer false positives) than the other two measures. However, when the degree of test specificity to detect low Aβ scans was set high (90%), LM II was associated with greater sensitivity relative to the other two measures. As can also be seen in Table 3, at 90% specificity, LM II score could be used to correctly classify 80% of participants as having high or low Aβ scan.

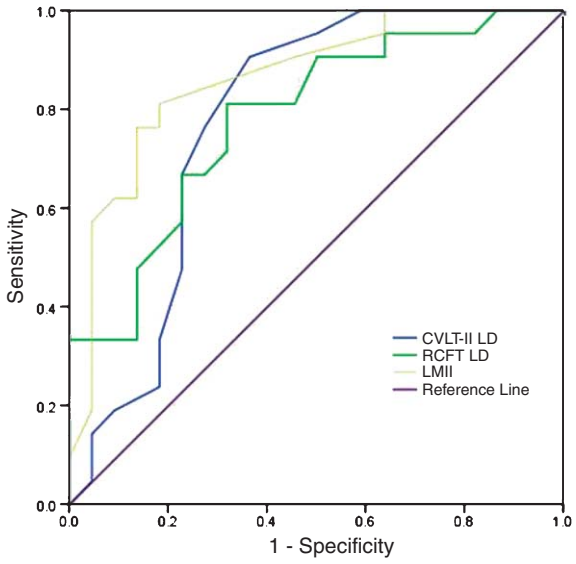

Fig. 2. ROC curve showing sensitivity and specificity of three memory measures to detect high Aβ scan.

Table 3
Cut-off values on three memory measures associated with 90% sensitivity to detect a positive Aβ scan and with 90% specificity to detect a negative Aβ scan

	CVLT-II	RCFT	LM II
Area under the curve	0.79	0.77	0.86
90% sensitivity			
Cut-off	≤4	≤6	≤2
Specificity	61%	36%	54%
Classification accuracy	73%	60%	70%
90% specificity			
Cut-off	≥10	≥16	≥8
Sensitivity	11%	33%	72%
Classification accuracy	54%	62%	80%

CVLT-II, California Verbal Learning Test-Second edition; LM, Logical Memory delayed recall; RCFT, Rey Complex Figure Test.

DISCUSSION

The current study confirms the well-established finding that among persons meeting criteria amnestic MCI, some display an 'AD-like' pattern of Aβ burden, but a substantial number display a 'Control-like' pattern of Aβ deposition. The current study also showed that in persons with amnestic MCI, episodic memory function is related to Aβ burden when treated both as a continuous measure and as a dichotomous variable, whereas other cognitive domains appear to be independent of Aβ burden. The significant association between Aβ burden and episodic memory was observed for all three memory measures used in the current study (story recall, word-list recall, and recall of a complex figure), as well as for a composite memory score which was derived from these three measures. In contrast to the findings of Mormino et al. [21] and Chételat et al. [23], the association between the memory composite and global Aβ burden remained significant after controlling for hippocampal volume. Examination of the association between Aβ burden and the three memory measures showed that Aβ burden was most strongly associated with delayed recall of a story. Furthermore, while both word-list and figure recall were moderately associated with hippocampal volume, story recall was not significantly associated with hippocampal volume. The independent relationship between Aβ burden and story recall was confirmed in regression analyses that showed that it was the only memory measure that accounted for additional variance in Aβ burden beyond that explained by demographic factors, hippocampal volume, global cognitive status, and other memory measures. Interestingly, while the association between Aβ burden and the delayed recall trial of both the word-list and the figure recall diminished when the immediate recall trial of these tests was controlled, the delayed recall of a story remained significantly related to Aβ even after controlling for the immediate recall trial. Consistent with results from other studies, performance on several non-memory measures was associated with WMH reflecting cerebrovascular pathology [14].

These findings have both theoretical and clinical implications. The relationship found between Aβ burden and scores on the delayed recall of a story had three characteristics not shared by the relationship between Aβ burden and the other memory measures. First, at $r = -0.58$, the association between Aβ burden and story recall was the strongest linear relationship found between Aβ burden and any other individual tests in the current study. Second, story recall was *selectively* associated with Aβ burden, but not with hippocampal volume, age, or education. In contrast, figure recall was associated equally strongly with both Aβ burden and hippocampal volume, and although the association between delayed word-list recall and the hippocampal volume failed to reach significance after accounting for age, this association was more than twice larger than the association between story recall and hippocampal volume. Finally, the association between Aβ burden and the delayed recall trial of a story was the only association that remained significant after controlling for scores on the immediate recall trial of the test.

The precise factors that give rise to these findings are unclear, and an independent replication of these findings is required to increase confidence in these observations. The strong, selective association between story recall and Aβ burden in this study may implicate particular neuroanatomical regions that are both important for performance on this task, and in which high Aβ levels are found. Whether the relationship between story recall and Aβ burden varied by cortical region is yet to be determined. Given the multiple comparisons that such analysis would entail and the associated risk of inflated error, this needs to be addressed in a separate study using a larger sample. Alternatively, this association may reflect the non-specific and broad range of cognitive abilities, including attention, comprehension, semantic knowledge and working memory, that are required for successful performance on story recall. The lack of association between story recall and hippocampal volume in the current study is particularly surprising given the well-established link between memory performance and hippocampal volume [40]. Indeed,

hippocampal volume was significantly associated with the composite memory score—an association that was most likely driven by the associations between word-list and figure recall with hippocampal volume. The reason for the lack of association between story recall and hippocampal volume is unclear but may be in part related to differences in the cognitive demands of the three memory measures used in this study. Story and word-list recall are similar in that they both involve an initial auditory presentation of information that places significant demands on the attention and working memory systems, whereas in the case of figure recall, participants are first presented with the figure and asked to draw a copy of it while the figure remains in sight. In all three tasks participants are not told that they will be asked to recall the information following a delay. However, whereas the story and figure to be recalled are only presented once during the initial exposure, the word-list is repeated to participants several times to explore learning processes. In this respect, initial and subsequent performance on both story and figure recall is more vulnerable to momentary fluctuations in attention, whereas the word-list paradigm involving repeated exposure to the same information gives people the opportunity to overcome transient attentional difficulties. Delayed recall of a story after a single exposure is therefore sometimes viewed as an unreliable measure of memory, mostly reflecting performance on the immediate recall of the story. Conversely, it is possible that the immediate recall of a story in persons with amnestic MCI benefits from the presentation of contextual information, whereas by the delayed recall trial, performance is more impaired due to considerable memory loss. The information in a word-list task, however, is not presented within a context, and immediate recall is moderated by semantic strategies, which are typically impaired in preclinical AD. Hence it is possible that initial performance as well as delayed recall would be impaired. Indeed, in the current study, the mean score on the immediate and delayed trial of the word-list task was virtually unchanged, whereas the mean delayed recall of a story was lower than the immediate recall trial (see Table 1), suggesting that overall, the time delay resulted in more forgetting in the case of the story than in the case of the word-list. Furthermore, as already noted, the association between Aβ burden and story recall was the only association that remained significant after controlling for the immediate recall trial. This implies that the initial recall of the story alone cannot fully account for the observed relationship between Aβ burden and delayed recall of a story, whereas in respect to both word-list and figure recall, controlling for the immediate recall trial led to elimination of the relationship between Aβ burden and the delayed recall trial of these measures. Taken together, these observations suggest that delayed recall per se plays a role in the relationship between story recall and Aβ burden but not in the relationship between Aβ burden and either word-list or figure recall.

If replicated, the results of the current study may also have implications for the choice of memory measures to be used in clinical settings to which persons with MCI are likely to present. Traditionally, when screening for AD or MCI, clinicians' choice of memory measures had been driven primarily by the need to reliably detect an objective deficit in episodic memory, and by the capacity of tests to demonstrate the signature of early AD-related memory problems, namely, impaired learning and rapid forgetting with little benefit from cueing in the context of preserved immediate memory ability. These continue to be important considerations in the selection of tests and assessment procedures, and therefore the assessment of immediate and delayed recall, learning as well as recognition will remain necessary components of the cognitive evaluation. However, with increasing recognition of the important role of *in vivo* biomarker detection, particularly Aβ, in determining the etiology underlying the clinical phenotype, the extent to which a memory measure can reliably indicate high Aβ burden may also be considered in the choice of memory measures used clinically.

In research settings, the relationship between Aβ burden and memory is usually explored using a composite memory score, combining scores from several tests of memory. Although this approach is generally more likely to yield more reliable memory estimates than any individual test score, cognitive assessments in clinical settings are often very limited in time. Therefore, comparing the capacity of different memory assessment measures to reflect Aβ burden is of clinical relevance. Although recall of a story in the current study showed the strongest and most independent linear relationship with Aβ burden, in clinical settings, in which Aβ burden is more likely to be treated as a dichotomous variable (i.e., high or low), the choice of memory measure is likely to be determined by the preferred levels of sensitivity and specificity. There is a known trade-off between the sensitivity and specificity of most measures and decisions regarding the preferred sensitivity of a test are generally made in the context of the perceived 'cost' of lower specificity (i.e.,

false positives). Given that Aβ is increasingly viewed as a necessary but insufficient factor in the development of AD, a high degree of specificity may arguably be preferred (i.e., no person is erroneously identified as carrying a high Aβ burden), in which case the current study suggests that story recall may be the measure of choice (the sensitivity associated with at least 90% specificity was 72% for LM II, 33% for RCFT, and only 11% for CVLT-II). If, however, maximum sensitivity is preferred, for example to maximize participation in a clinical trial open only to persons with a high Aβ burden, the results of the current study suggest that CVLT-II may be preferable (the specificity associated with at least 90% sensitivity was 57% for LM II, 53% for RCFT, and 64% for CVLT-II). Importantly, assessment of episodic memory by means of story recall has two important advantages over assessment with measures of word-list recall that are pertinent in clinical settings. First, recall of a story is a brief test requiring minimal assessment time, whereas word-list recall measures generally take longer to administer. Second, the administration of a story recall test requires minimal expertise in administration and scoring and can thus by given in a wide range of clinical settings by individuals without specialist training. Word-list learning tasks, in contrast, require greater expertise in administration, scoring and interpretation, and are therefore generally administered by neuropsychologists.

Several factors limited the ability to draw firm conclusions based on the results of the current study. First, the relatively small sample size and deviations from normality in the distribution of some measures may have implications for the stability of some of the findings. However, the strong effects found, particularly in respect to the relationship between Aβ burden and story recall, are likely to be robust in the face of departure from normality. Indeed, when the main correlations were carried out under ordinal data assumptions (e.g., using a Spearman correlation coefficient), the pattern and size of correlations was unchanged (data not shown). Plans to replicate these findings in a large independent sample from the AIBL Study are currently underway. The study is also limited by the memory measures that were included in the neuropsychological assessment. While the three measures used in the current study are in very common use in both clinical and research settings, inclusion of memory measures that were designed to target specific memory processes implicated in early and preclinical AD, such as paired-associate learning, is of interest. Also of note is that while we have used specific memory tests (Logical Memory, California Verbal Learning Test-Second edition, and the Rey Complex Figure Test), we have discussed the results using the more general terms: story, word-list, and figure recall. This is because the findings arguably reflect certain memory processes and testing principles rather than specific instruments. Nevertheless, it is of interest to see whether similar results are found using alternative word-list, story and figure recall tests. Similarly, in the current study, Aβ burden was assessed using ^{18}F-Florberaben. Whether or not the current findings will be replicated using a different Aβ radiotracer remains to be seen. Our plan to repeat these analyses with data from participants from the AIBL study will allow us to ascertain whether similar patterns will emerge when Aβ burden is assessed with Pittsburgh Compound B [PiB; 41]. Further, examination of the associations between specific memory measures and Aβ burden in healthy older adults, as well as in other memory-impaired populations without Aβ pathology is necessary to determine the extent to which these findings are specific to individuals with amnestic MCI. Finally, longitudinal analyses of the relevance of the various memory measures in the prediction of further cognitive decline and conversion to dementia in this sample are currently underway.

In conclusion, the current study found that Aβ burden is associated with episodic memory measures in persons with MCI, and that while this association was partly mediated by hippocampal volume in the case of figure and word-list recall, story recall was selectively associated with Aβ burden. Confidence in making a diagnosis of MCI due to AD in clinical settings may be enhanced when using memory measures that have been shown to be strongly and independently associated with Aβ burden.

ACKNOWLEDGMENTS

Funding support for this study was provided by Bayer Schering Pharma as part of the Imaging Aβ in MCI with ^{18}F-Florbetaben project. The funding source had no direct role in the design and conduct of the study, collection, analysis, interpretation of data, or preparation and review of this manuscript. The following people are acknowledged for their support: Greg Savage, Kerryn Pike, Narelle Langdon, Paul Yates, Svetlana Pejoska, Denise El-Sheikh, Rachel Mulligan, and Gareth Jones. We are also grateful to all the individuals who participated in the current study.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=1486).

REFERENCES

[1] 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**, 939-944.

[2] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[3] Dubois B, Feldman HH, Jacova C, Cummings JL, DeKosky ST, Barberger-Gateau P, Delacourte A, Frisoni G, Fox NC, Galasko D (2010) Revising the definition of Alzheimer's disease: A new lexicon. *Lancet Neurol* **9**, 1118-1127.

[4] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and Alzheimer's Association workgroup. *Alzheimers Dement* **7**, 270-279.

[5] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR, Kaye J, Montine TJ (2011) Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging and the Alzheimer's Association workgroup. *Alzheimers Dement* **7**, 280-292.

[6] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging and the Alzheimer's Association workgroup. *Alzheimers Dement* **7**, 263-269.

[7] Price JL, Morris JC (1999) Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Ann Neurol* **45**, 358-368.

[8] Rowe CC, Ng S, Ackermann U, Gong SJ, Pike K, Savage G, Cowie TF, Dickinson KL, Maruff P, Darby D, Smith C, Woodward M, Merory J, Tochon-Danguy H, O'Keefe G, Klunk WE, Mathis CA, Price JC, Masters CL, Villemagne VL (2007) Imaging beta-amyloid burden in aging and dementia. *Neurology* **68**, 1718-1725.

[9] Sperling RA, LaViolette PS, O'Keefe K, O'Brien J, Rentz DM, Pihlajamaki M, Marshall G, Hyman BT, Selkoe DJ, Hedden T (2009) Amyloid deposition is associated with impaired default network function in older persons without dementia. *Neuron* **63**, 178-188.

[10] Wolk DA, Price JC, Saxton JA, Snitz BE, James JA, Lopez OL, Aizenstein HJ, Cohen AD, Weissfeld LA, Mathis CA (2009) Amyloid imaging in mild cognitive impairment subtypes. *Ann Neurol* **65**, 557-568.

[11] Villemagne VL, Pike KE, Chételat G, Ellis KA, Mulligan RS, Bourgeat P, Ackermann U, Jones G, Szoeke C, Salvado O (2011) Longitudinal assessment of A and cognition in aging and Alzheimer disease. *Ann Neurol* **69**, 181-192.

[12] Villemagne VL, Fodero-Tavoletti MT, Pike KE, Cappai R, Masters CL, Rowe CC (2008) The ART of loss: Abeta imaging in the evaluation of Alzheimer's disease and other dementias. *Mol Neurobiol* **38**, 1-15.

[13] Cummings BJ, Pike CJ, Shankle R, Cotman CW (1996) Beta-amyloid deposition and other measures of neuropathology predict cognitive status in Alzheimer's disease. *Neurobiol Aging* **17**, 921-933.

[14] Nordahl CW, Ranganath C, Yonelinas AP, DeCarli C, Reed BR, Jagust WJ (2005) Different mechanisms of episodic memory failure in mild cognitive impairment. *Neuropsychologia* **43**, 1688-1697.

[15] Pike KE, Savage G, Villemagne VL, Ng S, Moss SA, Maruff P, Mathis CA, Klunk WE, Masters CL, Rowe CC (2007) Beta-amyloid imaging and memory in non-demented individuals: Evidence for preclinical Alzheimer's disease. *Brain* **130**, 2837-2844.

[16] Aizenstein HJ, Nebes RD, Saxton JA, Price JC, Mathis CA, Tsopelas ND, Ziolko SK, James JA, Snitz BE, Houck PR, Bi W, Cohen AD, Lopresti BJ, DeKosky ST, Halligan EM, Klunk WE (2008) Frequent amyloid deposition without significant cognitive impairment among the elderly. *Arch Neurol* **65**, 1509-1517.

[17] Pike KE, Ellis KA, Villemagne VL, Good N, Chételat G, Ames D, Szoeke C, Laws SM (2011) Cognition and beta-amyloid in preclinical Alzheimer's disease: Data from the AIBL study. *Neuropsychologia* **49**, 2384-2390.

[18] Storandt M, Mintun MA, Head D, Morris JC (2009) Cognitive decline and brain volume loss as signatures of cerebral amyloid-β peptide deposition identified with Pittsburgh compound B: Cognitive decline associated with Aβ deposition. *Arch Neurol* **66**, 1476.

[19] Rentz DM, Locascio JJ, Becker JA, Moran EK, Eng E, Buckner RL, Sperling RA, Johnson KA (2010) Cognition, reserve, and amyloid deposition in normal aging. *Ann Neurol* **67**, 353-364.

[20] Jack CR Jr, Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, Knopman DS, Boeve BF, Klunk WE, Mathis CA, Petersen RC (2008) 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. *Brain* **131**, 665-680.

[21] Mormino E, Kluth J, Madison C, Rabinovici G, Baker S, Miller B, Koeppe R, Mathis C, Weiner M, Jagust W (2009) Episodic memory loss is related to hippocampal-mediated -amyloid deposition in elderly subjects. *Brain* **132**, 1310.

[22] Bennett DA, Schneider JA, Wilson RS, Bienias JL, Arnold SE (2004) Neurofibrillary tangles mediate the association of amyloid load with clinical Alzheimer disease and level of cognitive function. *Arch Neurol* **61**, 378-384.

[23] Chételat G, Villemagne VL, Pike KE, Ellis KA, Bourgeat P, Jones G, O'Keefe GJ, Salvado O, Szoeke C, Martins RN (2011) Independent contribution of temporal -amyloid deposition to memory decline in the pre-dementia phase of Alzheimer's disease. *Brain* **134**, 798.

[24] Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med* **256**, 183-194.

[25] Busse A, Hensel A, Guhne U, Angermeyer MC, Riedel-Heller SG (2006) Mild cognitive impairment: Long-term course of four clinical subtypes. *Neurology* **67**, 2176-2185.

[26] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[27] Morris JC (1993) The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* **43**, 2412-2414.

[28] Saxton J, Ratcliff G, Munro CA, Coffey EC, Becker JT, Fried L, Kuller L (2000) Normative data on the Boston Naming Test and two equivalent 30-item short forms. *Clin Neuropsychol* **14**, 526-534.

[29] Wechsler D (1997) *WAIS-III Administration and Scoring Manual*, The Psychological Corporation, San Antonio.

[30] Delis D, Kramer J, Kaplan E, Ober B (2000) *CVLT-II: California Verbal Learning Test Second Edition Adult Version*, The Psychological Corporation, San Antonio.

[31] Wechsler D (1997) *Wechsler Memory Scale - Third Edition*, The Psychological Corporation, San Antonio.

[32] Meyers J, Meyers K (1995) *Rey Complex Figure Test and Recognition Trial*. Psychological Assessment Resources Inc., Odessa.

[33] Delis D, Kaplan E, Kramer J (2001) *D-KEFS executive function system*, The Psychological Corporation, San Antonio.

[34] Rowe CC, Ackerman U, Browne W, Mulligan R, Pike KL, O'Keefe G, Tochon-Danguy H, Chan G, Berlangieri SU, Jones G, Dickinson-Rowe KL, Kung HP, Zhang W, Kung MP, Skovronsky D, Dyrks T, Holl G, Krause S, Friebe M, Lehman L, Lindemann S, Dinkelborg LM, Masters CL, Villemagne VL (2008) Imaging of amyloid beta in Alzheimer's disease with 18F-BAY94-9172, a novel PET tracer: Proof of mechanism. *Lancet Neurol* **7**, 129-135.

[35] Kovacevic S, Rafii MS, Brewer JB (2009) High-throughput, fully-automated volumetry for prediction of MMSE and CDR decline in mild cognitive impairment. *Alzheimer Dis Assoc Disord* **23**, 139.

[36] Mevel K, Desgranges B, Baron JC, Landeau B, de La Sayette V, Viader F, Eustache F, Chételat G (2011) Which SPM method should be used to extract hippocampal measures in early Alzheimer's disease? *J Neuroimage* **21**, 310-316.

[37] Good CD, Johnsrude IS, Ashburner J, Henson RN, Friston KJ, Frackowiak RS (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* **14**, 21-36.

[38] Villemagne VL, Ong K, Mulligan RS, Holl G, Pejoska S, Jones G, O'Keefe G, Ackerman U, Tochon-Danguy H, Chan JG, Reininger CB, Fels L, Putz B, Rohde B, Masters CL, Rowe CC (2011) Amyloid imaging with 18F-florbetaben in Alzheimer disease and other dementias. *J Nucl Med* **52**, 1210-1217.

[39] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P (2009) The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatrics* **21**, 672-687.

[40] Bobinski M, de Leon MJ, Tarnawski M, Wegiel J, Reisberg B, Miller DC, Wisniewski HM (1998) Neuronal and volume loss in CA1 of the hippocampal formation uniquely predicts duration and severity of Alzheimer disease. *Brain Res* **805**, 267-269.

[41] Klunk WE, Engler H, Nordberg A, Wang Y, Blomqvist G, Holt DP, Bergstrom M, Savitcheva I, Huang GF, Estrada S, Ausen B, Debnath ML, Barletta J, Price JC, Sandell J, Lopresti BJ, Wall A, Koivisto P, Antoni G, Mathis CA, Langstrom B (2004) Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol* **55**, 306-319.