doi:10.1093/brain/awv112 BRAIN 2015: 138; 2020–2033 | 2020

BRAIN
A JOURNAL OF NEUROLOGY

Existing Pittsburgh Compound-B positron emission tomography thresholds are too high: statistical and pathological evaluation

Sylvia Villeneuve,[1,*] Gil D. Rabinovici,[1,2,3,*] Brendan I. Cohn-Sheehy,[1,2,3] Cindee Madison,[1] Nagehan Ayakta,[1,3] Pia M. Ghosh,[1,2,3] Renaud La Joie,[1] Samia Kate Arthur-Bentil,[1,2] Jacob W. Vogel,[1] Shawn M. Marks,[1] Manja Lehmann,[1,3] Howard J. Rosen,[3] Bruce Reed,[4,5] John Olichney,[4] Adam L. Boxer,[3] Bruce L. Miller,[3] Ewa Borys,[6] Lee-Way Jin,[7] Eric J. Huang,[8] Lea T. Grinberg,[3,8] Charles DeCarli,[4] William W. Seeley[3,8] and William Jagust[1,2]

*These authors contributed equally to this work.

Amyloid-β, a hallmark of Alzheimer's disease, begins accumulating up to two decades before the onset of dementia, and can be detected *in vivo* applying amyloid-β positron emission tomography tracers such as carbon-11-labelled Pittsburgh compound-B. A variety of thresholds have been applied in the literature to define Pittsburgh compound-B positron emission tomography positivity, but the ability of these thresholds to detect early amyloid-β deposition is unknown, and validation studies comparing Pittsburgh compound-B thresholds to post-mortem amyloid burden are lacking. In this study we first derived thresholds for amyloid positron emission tomography positivity using Pittsburgh compound-B positron emission tomography in 154 cognitively normal older adults with four complementary approaches: (i) reference values from a young control group aged between 20 and 30 years; (ii) a Gaussian mixture model that assigned each subject a probability of being amyloid-β-positive or amyloid-β-negative based on Pittsburgh compound-B index uptake; (iii) a k-means cluster approach that clustered subjects into amyloid-β-positive or amyloid-β-negative based on Pittsburgh compound-B uptake in different brain regions (features); and (iv) an iterative voxel-based analysis that further explored the spatial pattern of early amyloid-β positron emission tomography signal. Next, we tested the sensitivity and specificity of the derived thresholds in 50 individuals who underwent Pittsburgh compound-B positron emission tomography during life and brain autopsy (mean time positron emission tomography to autopsy 3.1 ± 1.8 years). Amyloid at autopsy was classified using Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria, unadjusted for age. The analytic approaches yielded low thresholds (standard uptake value $ratio_{low}$ = 1.21, distribution volume $ratio_{low}$ = 1.08) that represent the earliest detectable Pittsburgh compound-B signal, as well as high thresholds (standard uptake value $ratio_{high}$ = 1.40, distribution volume $ratio_{high}$ = 1.20) that are more conservative in defining Pittsburgh compound-B positron emission tomography positivity. In voxel-wise contrasts, elevated Pittsburgh compound-B retention was first noted in the medial frontal cortex, then the precuneus, lateral frontal and parietal lobes, and finally the lateral temporal lobe. When compared to post-mortem amyloid burden, low proposed thresholds were more sensitive than high thresholds (sensitivities: distribution volume $ratio_{low}$ 81.0%, standard uptake value $ratio_{low}$ 83.3%; distribution volume $ratio_{high}$ 61.9%, standard uptake value $ratio_{high}$ 62.5%) for CERAD moderate-to-frequent neuritic plaques, with similar specificity (distribution volume $ratio_{low}$ 95.8%; standard uptake value $ratio_{low}$, distribution volume $ratio_{high}$ and standard uptake value $ratio_{high}$ 100.0%). A receiver operator characteristic analysis identified optimal distribution volume ratio (1.06) and standard uptake value ratio (1.20) thresholds that were nearly identical to the *a priori* distribution volume $ratio_{low}$ and standard uptake value $ratio_{low}$. In summary, we found that frequently applied thresholds for Pittsburgh compound-B positivity (typically at or above distribution volume $ratio_{high}$ and standard uptake value $ratio_{high}$) are overly stringent in defining amyloid positivity. Lower thresholds in this study resulted in higher sensitivity while not compromising specificity.

Received November 30, 2014. Revised February 13, 2015. Accepted March 3, 2015. Advance Access publication May 7, 2015

1 Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley CA, USA
2 Lawrence Berkeley National Laboratory, Berkeley, California, USA
3 Memory and Aging Centre, Department of Neurology, University of California, San Francisco, USA
4 Department of Neurology, School of Medicine, University of California, Davis, USA
5 Veterans Affairs Northern California Health Care System, Martinez, CA, USA
6 Department of Pathology, Stritch School of Medicine, Loyola University Chicago, USA
7 Department of Pathology and Laboratory Medicine, University of California, Davis, USA
8 Department of Pathology, University of California, San Francisco, USA

Correspondence to: Sylvia Villeneuve,
Helen Wills Neuroscience Institute,
132 Barker Hall,
MC# 3190,
University of California, Berkeley,
CA 94720 USA
E-mail: villeneuve.sylvia@gmail.com

Keywords: Alzheimer's disease; dementia; biomarkers; neurodegeneration; beta-amyloid.

Abbreviations: CERAD = Consortium to Establish a Registry for Alzheimer's Disease; DVR = distribution volume ratios; PiB = Pittsburgh compound-B; SUVR = standardized uptake value ratio

Introduction

PET amyloid imaging has had a profound effect on ageing and dementia research. The first publication of an amyloid-β-selective imaging agent, carbon-11 labelled Pittsburgh compound-B (^{11}C-PiB; Klunk *et al.*, 2004) opened the door to *in vivo* detection of a core aspect of Alzheimer's disease pathology. Soon afterwards, ^{18}F-labelled amyloid imaging agents were developed and commercialized, widely increasing the availability of this technology. The ability to detect and quantify fibrillar brain amyloid-β *in vivo* has helped to establish models of disease pathophysiology and biomarker progression (Jack *et al.*, 2010, 2013; Bateman *et al.*, 2012) and guide the design of clinical trials (Salloway *et al.*, 2014). These studies offer the potential for a more complete understanding of the pathophysiology of Alzheimer's disease, along with the hope of early therapeutic intervention in people who harbour amyloid pathology but do not yet express cognitive decline.

In studies of amyloid imaging, many investigators use an overall measure of radioligand retention in the brain to dichotomize subjects into 'amyloid positive' and 'amyloid negative' categories. However, amyloid-β deposition occurs on a continuum; at present there is no clear *a priori* way to separate individuals who have pathologically relevant amyloid-β deposition from those who do not. Nevertheless, there are important reasons to consider categorical classification of individual subjects. Classification of individuals as amyloid 'positive' or 'negative' is relevant for clinical diagnosis, for inclusion of subjects in anti-amyloid therapeutic trials, and for distinguishing amyloid-β-dependent and amyloid-β-independent changes in cognition and in brain structure and function. Measurements of 1.4–1.5 standardized uptake value ratio (SUVR) units have often been used in the literature to identify amyloid-β-positive subjects using PET scanning with PiB. These thresholds are based on different categorization approaches such as the natural data breakpoints, the upper confidence limit observed in cognitively normal older adults, the lower confidence limits found in patients with clinical Alzheimer's disease dementia, iterative outlier removal, hierarchical clustering or Gaussian mixture modelling (Pike *et al.*, 2007; Aizenstein *et al.*, 2008; Jack *et al.*, 2008, 2012; Hedden *et al.*, 2009; Rowe *et al.*, 2010; Villemagne *et al.*, 2011; Nordberg *et al.*, 2013; Mormino *et al.*, 2014). It is unknown if these thresholds truly allow the earliest possible detection point of pathologically relevant amyloid-β-PET signal. Identifying subjects with amyloid deposition as early as possible is important to truly understand the Alzheimer's disease pathophysiological cascade, and to exclude early amyloid accumulators from studies focusing on normal cognitive ageing and 'suspected non-Alzheimer's disease pathology' (SNAP; Jack *et al.*, 2012).

The main aim of this study was to identify and validate thresholds that detect pathologically relevant PiB signal as early as possible. A secondary aim was to examine the spatial pattern of early amyloid-β PET signal. As a first step, we applied four distinct statistical approaches to define a low threshold for amyloid-β PET positivity based on ^{11}C-PiB data from 154 older adults. Thresholds were defined based on (i) reference values from a young control group; (ii) Gaussian mixture modelling; (iii) a k-means cluster approach; and (iv) an iterative voxel-based analysis. The Gaussian mixture modelling and the cluster analyses also allowed derivation of a higher threshold that might be favoured when reducing the rate of false positives is more important than detecting early PiB-PET signal. The cluster and the voxel-wise analyses further allowed examination of the spatial pattern of early amyloid-β PET signal. Because amyloid-β is hypothesized to start accumulating long before

cognitive impairment is clinically evident, we derived the thresholds based on data acquired in cognitively normal older adults. Thresholds were derived for the two most common methods of PiB-PET quantification: SUVR units, and distribution volume ratios (DVR).

Evaluating sensitivity and specificity of amyloid PET thresholds requires a 'standard of truth' which, for the detection of brain amyloid-β, necessitates pathological examination of the brain (Clark *et al.*, 2011). As a second step, we therefore applied the derived low and high SUVR and DVR thresholds to PiB-PET scans from 50 individuals enrolled in longitudinal studies of ageing and dementia who underwent amyloid imaging and were also followed to autopsy. Classifications based on the proposed thresholds were compared to the burden of amyloid at autopsy as measured by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) scale (Mirra *et al.*, 1991). CERAD is a semi-quantitative scale of amyloid-β neuritic plaques, fibrillar amyloid aggregates considered to be the primary amyloid-β species that bind PiB *in vivo* (Ikonomovic *et al.*, 2008, 2012; Ni *et al.*, 2013).

We hypothesized that the proposed low thresholds would have higher sensitivity for significant amyloid-β burden (defined as CERAD moderate-to-frequent neuritic plaques), whereas the proposed high thresholds would have higher specificity and result in fewer false positives in brains with CERAD absent to sparse neuritic plaques. We also hypothesized that cortical hubs of brain connectivity would be key regions of early PiB signal, supporting the idea that regions of high connectivity are prone to earlier amyloid-β deposition (Buckner *et al.*, 2009).

Materials and methods

Threshold derivation

Participants

Table 1 shows the demographic data of 154 cognitively normal elderly and 18 young adults included in the threshold derivation step. Older adults were from the Berkley Aging Cohort (BAC) and from ongoing studies at the University of California Davis (UCD) Alzheimer's Disease Centre (see Supplementary material for more details). Written informed consent was obtained from participants under protocols approved by the Institutional Review Boards of all participating institutions.

MRI and preprocessing

Structural T_1-weighted magnetic resonance images were obtained on different scanners (Supplementary material). All MPRAGE scans were processed with FreeSurfer version 5.1 (http://surfer.nmr.mgh.harvard.edu/) to derive regions of interest in each subject's native space using the Desikan-Killiany atlas (Desikan *et al.*, 2006). These regions of interest where then used to extract regional cortical PiB values.

PET imaging and preprocessing

All subjects underwent PiB-PET imaging at Lawrence Berkeley National Laboratory (Supplementary material). PiB-PET data were preprocessed using the Statistical Parametric Mapping software package (SPM8; http://www.fil.ion.ucl.ac.uk/spm) using a previously published protocol (Villeneuve *et al.*, 2014). All subjects included in the threshold derivation received dynamic scans. DVRs were generated with Logan graphical analysis, PiB frames corresponding to 35–90 min post-injection and a native-space grey matter cerebellar mask as the reference region (Logan *et al.*, 1996; Price *et al.*, 2005). SUVRs were calculated by dividing the mean uptake 50–70 min post-injection by the grey matter cerebellar mask.

For each subject, both a DVR and a SUVR 'PiB index' were derived from the native-space image by averaging the weighted mean value from Freesurfer-derived regions of interest in frontal, temporal, parietal and posterior cingulate cortex using the Desikan-Killiany atlas (Desikan *et al.*, 2006) (Supplementary Fig. 1A). The PiB index thus includes cortical regions that show a high proclivity for PiB retention in Alzheimer's disease and normal ageing (Price *et al.*, 2005; Rabinovici *et al.*, 2010). Region of interest-specific values were also extracted from 76 regions of interest from the same atlas (Supplementary Table 1). Because of the linear correspondence found between the DVR and the SUVR values, all statistical analyses for the threshold derivation part of the manuscript were performed using DVRs, and a regression line (SUVR = $-0.54 + 1.62 \times$ DVR, R = 0.97) was applied to derive the SUVR cut-offs corresponding to the DVR thresholds.

Statistical analysis

We investigated optimal PiB Index cut-offs to detect early PiB-PET signal using four different methods: (i) a reference group of young adults; (ii) Gaussian mixture modelling analysis; (iii) cluster analysis; and (iv) voxel-wise analysis.

Young adults analysis

There is strong evidence that adults under the age of 30 years are almost invariably free from amyloid-β accumulation (Kok *et al.*, 2009; Fleisher *et al.*, 2012). Thus, the first threshold we used in this study was defined as 2 standard deviations (SD) above the group of 18 young subjects aged between 20 and 30 years. This approach was used in a previous publication from our group (Mormino *et al.*, 2012). In that previous publication, which included 11 of the 18 current young subjects, the cut-off was set at DVR = 1.08.

Gaussian mixture models analysis

A Gaussian mixture model is a probabilistic model assuming that the overall data distribution can be estimated from a mixture of Gaussian distributions. Using that technique, we fit from 1 to 11 Gaussian distributions to our data and used a Bayesian information criterion (BIC) to assess the optimal number of Gaussian distributions represented in our data. We found that the best fit for our data was two Gaussian distributions, which is consistent with a previous report using different data sets (Mormino *et al.*, 2014). Then, each subject was assigned a probability of belonging to each Gaussian distribution. The two cut-offs derived using this technique represent the 90% probability of belonging in the low

Table 1 Participant demographics

	Threshold derivation study		Threshold validation study
	Young adults	Older adults	Older adults
n	18	154	50
Male (%)	8 (44%)	69 (45%)	33 (66%)
Age at PET	23.7 (2.9)	76.0 (6.2)	69.8 (9.6)
Education	13.2 (6.2)	16.1 (2.5)	15.7 (2.9)
MMSE	–	28.9 (1.2)	21.6 (6.8)
ApoE4 (%)	7 (47%)[a]	45 (31%)[b]	13 (27%)[c]
CDR	–	–	1.2 (0.8)
Time from PET to death (years)	–	–	3.1 (1.8)
PiB index DVR	1.01 (.03)	1.12 (.20)	1.14 (.29)[d]
PiB+ (%), DVR cut-off 1.08	0 (0%)	57 (37%)	18 (40%)[d]
PiB+ (%), DVR cut-off 1.20	0 (0%)	34 (22%)	13 (29%)[d]
Clinical diagnosis at PET	Normal (18)	Normal (154)	FTD[e] (28), AD (11), MCI (7), AD/SIVD (1), DLB (1), normal (2)
Primary neuropathological diagnosis	–	–	FTLD[f] (25), AD (6), CVD (6), AD/CVD (3), AD/DLB (3), AD/FTLD (4), TPD (1), AGD (1), no pathological findings (1)
CERAD (absent, sparse, moderate, frequent)	–	–	18, 8, 4, 20

Shown are mean (SD) unless specified otherwise.
[a]Three subjects were not genotyped.
[b]Eight subjects were not genotyped.
[c]Two subjects were not genotyped.
[d]DVR data missing for five subjects.
[e]Clinical syndromes included: corticobasal syndrome (CBS; 8), behavioural-variant FTD (6), FTD and amyotrophic lateral sclerosis (5), non-fluent variant primary progressive aphasia (nfvPPA; 5), nfvPPA/CBS (1), semantic variant PPA (3).
[f]FTLD neuropathological subtypes: FTLD-TDP (12), corticobasal degeneration (7), Pick's disease (3), progressive supranuclear palsy (2), FTLD with non-specific 4-repeat tauopathy (1).
AD = Alzheimer's disease; AGD = argyrophillic grain disease; ApoE = apolipoprotein E; CDR = Clinical Dementia Rating; CVD = cerebrovascular disease; MCI = mild cognitive impairment; DLB = dementia with Lewy bodies; FTD = frontotemporal dementia; FTLD = frontotemporal lobar degeneration; MMSE = Mini-Mental State Examination; SIVD = subcortical ischaemic vascular disease; TPD = tangle-predominant dementia.

(representing the PiB-negative subjects) or the high (representing the PiB-positive subjects) distributions. While admittedly arbitrary, we chose a probability of 90% based on thresholds applied to define abnormal scan results in the literature (Jack *et al.*, 2012). Notably, results were highly similar while using a 95% probability (Supplementary Table 3).

k-Means cluster analysis

This technique is similar to (and not totally independent from) the Gaussian mixture modelling analyses in the sense that it defines how we can cluster, or group, the data together. In this analysis, instead of examining the DVR PiB index we used the PiB DVR values extracted from the 76 regions of interest defined by the Desikan-Killiany atlas. We restricted the analysis to two clusters, one that represents subjects with high amyloid-β deposition and the other representing subjects with low amyloid-β deposition. The two cut-offs derived using this technique represent the 90th percentile of the low cluster (representing the PiB-negative subjects) or the 10th percentile of the high cluster (representing the PiB-positive subjects).

Voxel-wise analysis

In this analysis, we began by ranking all older normal subjects by their DVR PiB index. We then created a reference group of 22 subjects with a mean index of 1.00 to which we compared a series of subsequent groupings of the remaining participants (referred to here as the group of interest). In other words, after selecting the 22 subjects that comprised the reference group, we took the 22 subjects with the next highest PiB index values and performed voxel-wise contrasts between this group of interest and the reference group. We then dropped the subject from the group of interest with the lowest PiB index and added the subject with the next highest value and again performed a voxel-wise comparison between that group of interest of 22 subjects to the same reference group. We continued this process, iteratively creating groups of interest of 22 subjects by dropping the individual with the lowest PiB index and adding the one with the next highest index, such that the 'new' group of interest differed from the previous group by only two subjects. In this way, groups of interest gradually moved up the scale of PiB index DVRs, always using the same initial reference group of 22 subjects (mean DVR = 1.00) for comparison. This procedure was repeated until the subject with the highest DVR was included. A DVR of 1.00 was chosen for the reference group since this DVR value reflects a level of PiB retention in the cortex equivalent to the cerebellum grey matter, and thus indicates no specific tracer retention.

Voxel-wise analyses were performed using SPM8 software. First, the DVR scans were warped to the MDT2 template (Sun *et al.*, 2007). Then, images were smoothed (Gaussian kernel of $10 \times 10 \times 10$ mm), masked to exclude non-grey matter voxels from the analyses, and two-sample *t*-tests were performed. All voxel-wise analyses were corrected for multiple comparisons using a family-wise error (FWE) at $P < 0.05$ and cluster size $k \geqslant 150$.

The DVR threshold was defined as the mean PiB index DVR of the group of interest when a statistically significant signal of elevated PiB retention was first detected. Overall, 66 iterations were used to examine patterns of amyloid-β accumulation. In this model, we assume that increasing DVR represents temporal progression, an assumption that we recognize may not be entirely justified.

Threshold validation

Participants

Table 1 shows the demographic data of the 50 individuals who had both PiB-PET and autopsy. Participants were from the University of California, San Francisco Memory and Aging Center or the UCD Alzheimer's Disease Center (Supplementary material). All but two of these participants were cognitively impaired and therefore not included in the threshold derivation step. Clinical diagnosis was established at a multi-disciplinary conference applying standard research criteria for mild cognitive impairment and dementia syndromes (McKhann *et al.*, 1984, 2011; Roman *et al.*, 1993; Petersen, 2004; McKeith *et al.*, 2005; Albert *et al.*, 2011; Gorno-Tempini *et al.*, 2011; Rascovsky *et al.*, 2011; Armstrong *et al.*, 2013). This report reflects data on 50 individuals who had both PiB-PET and autopsy as of May 2014.

MRI and preprocessing

Structural T_1-weighted MRI were obtained on different scanners (Supplementary material). T_1-weighted images were used only for definition of the cerebellum reference region, using FreeSurfer v5.1 software, and for spatial normalization.

PET imaging and processing

All subjects underwent PET imaging on the same scanners with the same acquisition parameters as for the threshold derivation study. Forty-five of these subjects underwent dynamic PET imaging. The cerebellar grey normalized DVR image was then warped to an anatomical T_1-based template in MNI space using the subjects' T_1 MRIs as reference images. Grey matter segmentation was defined for each subject in template space, applying a probabilistic grey matter mask from the Montreal Neurological Institute (MNI) T_1 template using Statistical Parametric Mapping (SPM8). In five subjects, dynamic data were not obtained, but PiB data were collected from 50–70 min after tracer injection for calculation of SUVRs (in one subject, only data from 55–70 min were available). SUVR and DVR images were subsequently processed identically. All analyses for the threshold validation part of the manuscript were performed using both the DVR and the SUVR values.

Similar to the threshold derivation study, we estimated mean cortical PiB retention using a 'PiB index'. However, due to a high failure rate of FreeSurfer-based segmentation in MRIs derived from dementia subjects, the PiB index in this group was created by combining frontal, temporal, parietal and posterior cingulate regions defined within the Automated Anatomical Labelling (AAL) atlas (Tzourio-Mazoyer *et al.*, 2002) and extracting values in template space. The AAL-defined regions were highly analogous to those defined using the Desikan atlas (Supplementary Fig. 1B), and the two methods yield highly correlated mean DVR values (Supplementary material).

Neuropathological examination

Brain autopsies were performed at the University of California, San Francisco Memory and Aging Center ($n = 38$), University of California Davis ($n = 9$), University of Pennsylvania ($n = 1$), University of California Los Angeles ($n = 1$) and Mayo Clinic Jacksonville ($n = 1$). Pathological assessments were performed using institution-specific protocols, as previously described (Chui *et al.*, 2006; Forman *et al.*, 2006; Grinberg *et al.*, 2013; Magaki *et al.*, 2014; Murray *et al.*, 2014). All autopsies included tissue sampling in regions relevant to the differential diagnosis of dementia based on published consensus criteria (Roman *et al.*, 1993; Newell *et al.*, 1999; McKeith *et al.*, 2005; Mackenzie *et al.*, 2010; Hyman *et al.*, 2012). Tissue staining included haematoxylin and eosin and at least one of the following stains: thioflavin S, modified Bielschowsky, Gallyas silver stain, or immunohistochemistry for amyloid-β (the latter was available in 41/50 cases and is considered equivalent to older stains for the purposes of CERAD staging in updated pathological criteria; Hyman *et al.*, 2012). Neuritic plaque densities were based on the assessment of sections stained with these methods. CERAD scores were based on the highest density of neuritic plaques found at autopsy. Sections were rated unadjusted for age, as follows: 1–5 neuritic plaques in a $\times 100$ field were classified as CERAD-sparse, 6–14 as CERAD-moderate and $\geqslant 15$ as CERAD-frequent (Montine *et al.*, 2012). Immunohistochemistry for hyperphosphorylated tau (MAPT), α-synuclein (SNCA), ubiquitin, and transactive response DNA binding protein 43 (TARDBP) was performed based on institutional protocols. Alzheimer's disease neuropathology was further characterized by Braak stage (Braak and Braak, 1998) and the National Institute on Aging–Reagan criteria (Hyman and Trojanowski, 1997). In some cases neuropathologists had access to the clinical histories and thus may not have been blinded to PiB-PET results.

Autopsy reports were reviewed by an experienced neurologist (G.D.R.) who extracted the primary and contributing neuropathological diagnoses, as well as CERAD, Braak and NIA-Reagan scores for each case. The presence and degree of cortical diffuse plaques and cerebral amyloid angiopathy in brain parenchyma (i.e. not isolated to leptomeninges) was also recorded. Staging applying the updated NIA-AA pathological criteria for Alzheimer's disease (Hyman *et al.*, 2012) were only available for a subset of participants since many of the autopsies preceded publication of these new criteria.

Estimation of sensitivity and specificity

We assessed the relationship between classification of subjects as PiB-PET positive or negative applying the derived low and high thresholds, and compared PET categorization to the classification of subjects by autopsy as positive or negative for significant neuritic plaques. To conform to previous PET-pathology studies (Clark *et al.*, 2012; Curtis *et al.*, 2015), cases with CERAD scores of 'absent' or 'sparse' were categorized as pathologically negative for significant neuritic plaques, while 'moderate' and 'frequent' CERAD scores were considered positive. Sensitivity and specificity for each of these thresholds were estimated by the appropriate observed proportion, and 95% confidence intervals were generated based on the assumption that they follow a binomial distribution. To explore whether the *a priori* thresholds we selected were truly optimal, in a separate exploratory analysis we used receiver-operating characteristic (ROC) curves to determine the PiB thresholds that maximized overall classification accuracy of CERAD moderate-frequent cases versus CERAD absent-sparse cases.

Additional statistical analyses

For both the threshold derivation and validation steps, group differences in continuous variables were examined using Student's *t*-tests. Group differences in dichotomous variables were compared using chi-square or Fisher's exact tests. Statistical analyses were implemented in PASW 21.0 (SPSS Inc.).

Results

Threshold derivation

Young adults

The mean DVR of the young subjects was 1.01 (SD = 0.03), leading to a cut-off of 1.07 that is 2 SD above the mean. Fitting a regression line between DVRs and SUVRs (SUVR = −0.54 + 1.62 × DVR, R = 0.97) showed that a DVR cut-off of 1.07 is equivalent to an SUVR cut-off of 1.19.

Table 2 shows the cut-offs for the four analytic techniques applied in this study.

Gaussian mixture models

The two Gaussian distributions are presented in Fig. 1. Based on the model, each subject was assigned a probability of belonging to either distribution. Two cut-offs were derived: the lower threshold was DVR = 1.09 and corresponds to a 90% probability of belonging in the low amyloid-β distribution; and the higher threshold was DVR = 1.21 and corresponds to a 90% probability of belonging to the high amyloid-β distribution. Applying the DVR versus SUVR regression line identified a low SUVR cut-off of 1.23 and a high SUVR cut-off of 1.42.

Cluster analysis

The two clusters are presented in Fig. 2A, representing subjects with high and low amyloid-β. Each row of each cluster represents one subject and each column represents one significant feature (region of interest), for a total of 12 features. Similar to the Gaussian mixture modelling approach, each subject was assigned a probability of belonging to Cluster A (low amyloid-β) or Cluster B (high amyloid-β). Two cut-offs were derived: the lower threshold was DVR = 1.09 and corresponds to the 90th percentile of Cluster A (low amyloid-β); and the higher threshold was DVR = 1.19 and corresponds to the 10th percentile of Cluster B (high amyloid-β). Fitting a regression line between the DVR and the SUVR data suggest a low SUVR cut-off of 1.23 and a high SUVR cut-off of 1.39.

The 12 features represent the brain regions that best (or most significantly) discriminate two clusters, among the 76 regions of interest tested in the model, are shown in Fig. 2B. The anterior cingulate (left and right) and the precuneus (left and right) stood out as key regions that differentiate the two clusters. The regions of interest representing the 12 significant features, that are mainly symmetrical, as well as their weights in the model are shown in Supplementary Table 2.

Table 2 Cut-offs for capturing early amyloid PET positivity

	2 SD above young (n = 18)	Voxel-wise analysis	Cluster analysis	Gaussian mixture modelling analysis	Optimal cut-off
DVR	1.07	1.07	1.09	1.09	**1.08**
SUVR	1.19	1.19	1.23	1.23	**1.21**

Shown are the optimal cut-offs derived using four independent data analysis techniques as well as the optimal cut-off proposed based on the results of these four analyses. The analyses were performed using the DVR data and a regression line (SUVR = −0.54 + 1.62 × DVR) was fit between the DVR and the SUVR data to calculate the corresponding SUVR cut-offs.

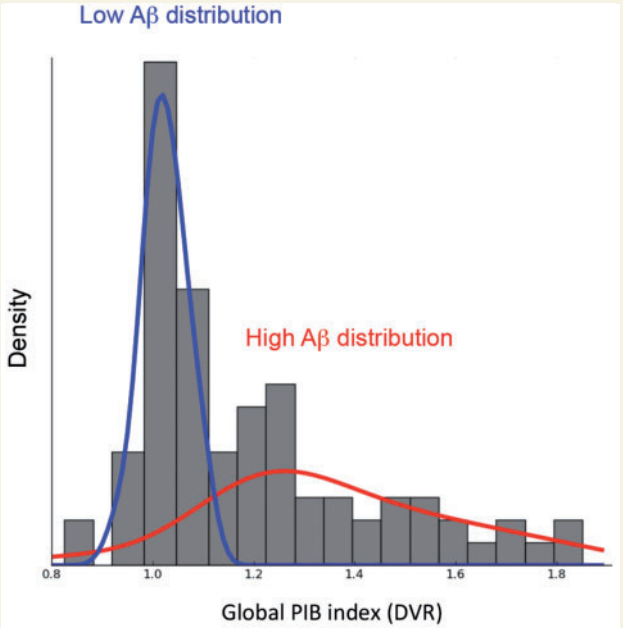

Figure 1 Gaussian mixture model containing two mixtures (distributions) that best fit the PiB index values. The blue line represents the distribution associated with low amyloid-β (Aβ) values while the red line represents the distribution associated with high amyloid-β values. The two Gaussian distributions are superimposed on the subject density histogram for all PiB index values in older subjects.

Voxel-wise analysis

The first significant difference between the reference group (n = 22, mean DVR = 1.00, SD = 0.01) and the groups of interest (n = 22) was found when the group of interest had a mean PiB index DVR of 1.07 (SD = 0.01) (Fig. 3 and Supplementary Video 1). In this contrast, a cluster of significant PiB retention was limited to the medial frontal

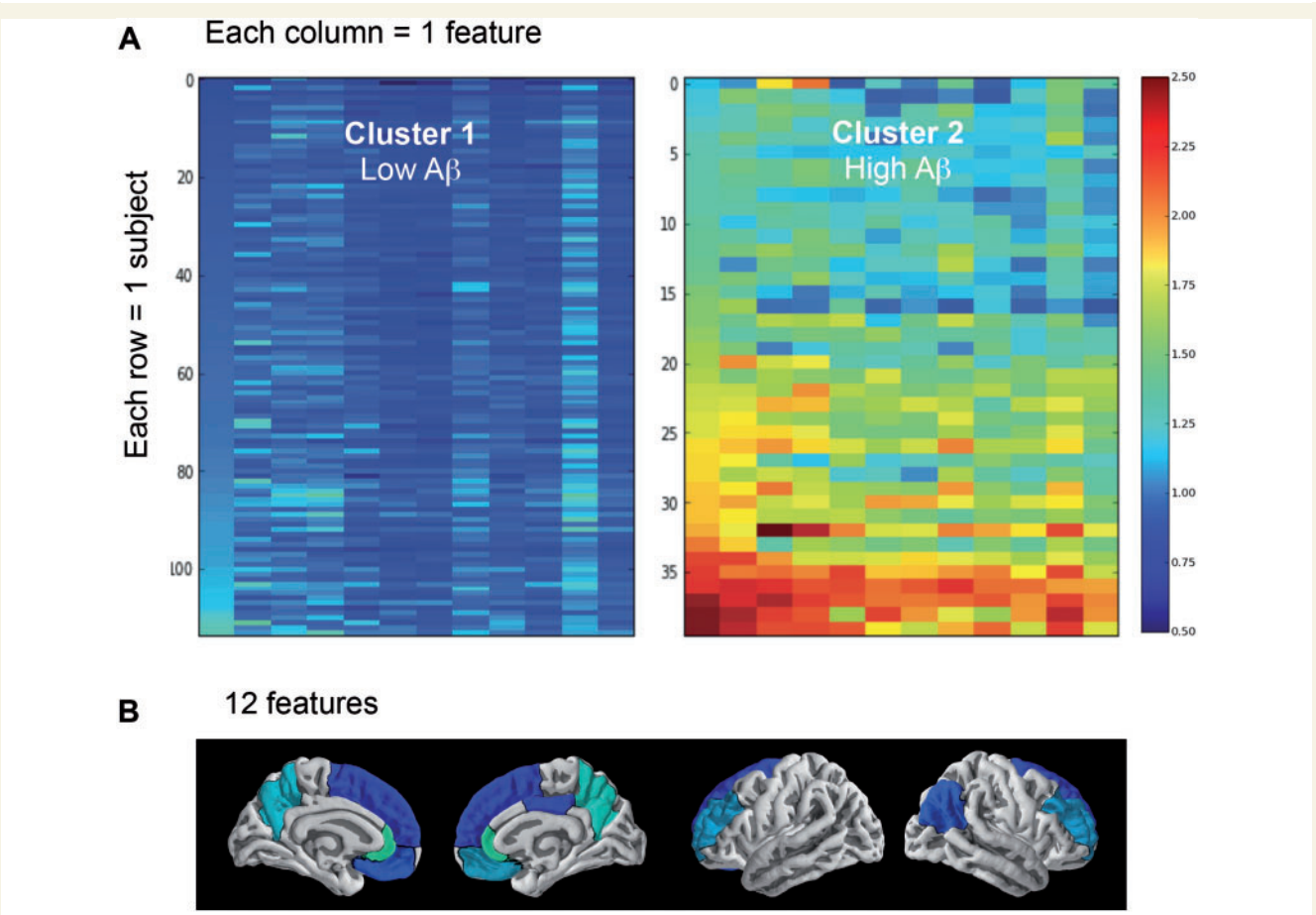

Figure 2 Cluster analysis. Cluster analysis containing two clusters (groups) representing individuals with low and high amyloid burden (**A**), as well as the 12 features (brain regions) that helped identify these two clusters (**B**). Cluster 1 (*left*) represents subjects with low amyloid-β (Aβ) values while Cluster 2 (*right*) represents subjects with high amyloid-β values. Warmer colours are associated with higher DVR values (see the DVR colour scale ranging from 0.5 to 2.5). For each cluster, each row represents one subject and each column represents one of the 12 features that helped identify the two clusters. From left to right, the 12 features are: rostral anterior cingulate left hemisphere (lh); rostral anterior cingulate right hemisphere (rh); precuneus rh; precuneus lh; medial orbitofrontal rh; rostral middle frontal lh; rostral middle frontal rh; inferior parietal rh; medial orbitofrontal lh; superior orbitofrontal rh; posterior cingulate rh; and superior orbitofrontal lh. These 12 features are also projected on a brain with the lighter colours corresponding to the features that have the highest weight in the model (see Supplementary Table 2 for the weight of each feature in the model).

cortex; at very slightly higher levels of mean PiB index DVRs (1.08), significant clusters were found in the precuneus, followed by the lateral frontal and parietal lobes, which appeared when the mean PiB index DVR was near 1.14 (SUVR 1.31), and finally the lateral temporal lobe (Fig. 3 and Supplementary Video 1). The DVR of 1.07 defining the first significant cluster of voxels corresponds to an SUVR of 1.19.

Threshold validation

In the separate autopsy validation cohort, the mean interval between PET scanning and death was 3.1 years (range: 0.2–6.4 years). At autopsy, the most common neuropathological diagnoses were frontotemporal lobar degeneration, Alzheimer's disease (often with mixed pathology) and cerebrovascular disease (Table 1). The distribution of CERAD scores was relatively bimodal, with most subjects classified as CERAD absent or CERAD frequent. Only four subjects fell in the CERAD moderate category.

The overall prevalence of amyloid-β deposits (in the form of neuritic plaques, diffuse plaques or cerebral amyloid angiopathy) at autopsy was 84%, including 64% with neuritic plaques, 80% with diffuse plaques (data missing in four cases) and 50% with cerebral amyloid angiopathy (data missing in six cases). Some degree of amyloid pathology was present at autopsy in 100% of APOE4 carriers versus 77% of non-carriers ($P = 0.09$).

Figure 4 shows examples of PiB-PET scans in subjects representing the spectrum of post-mortem amyloid burden. PiB-PET in the subject with CERAD sparse plaques had DVR and SUVR values well below the low thresholds (Fig. 4). PiB-PET in the subject with CERAD moderate plaques had DVR and SUVR values

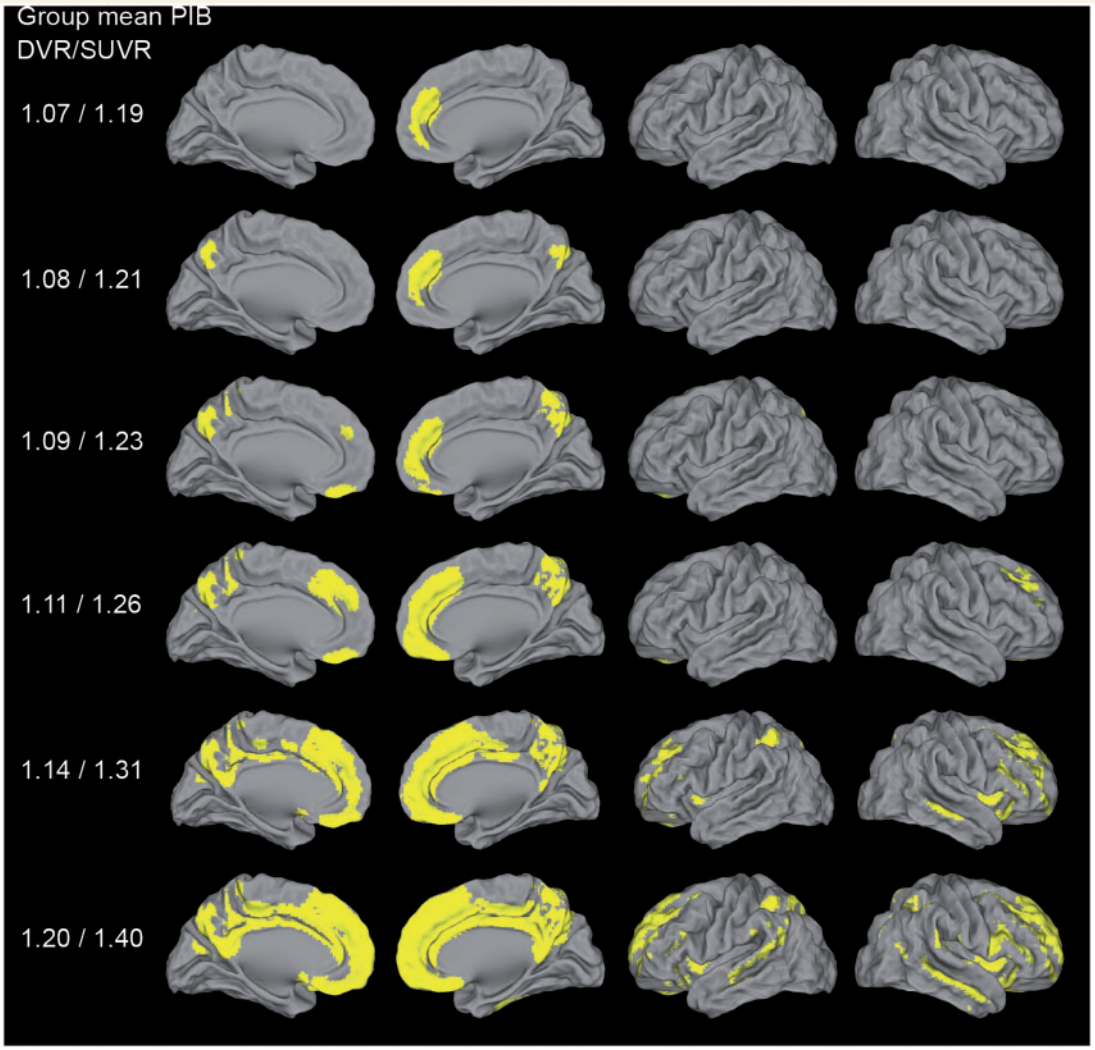

Figure 3 Pattern of early detectable PiB binding in cognitively normal older adults. Each row of each image reflects a voxel-wise contrast of 22 subjects with the mean values for the DVR/SUVR index listed on the left compared to a reference group (n = 22) with a DVR index of 1.00. Significant voxels first appeared when the group mean was DVR = 1.07 (see also Supplementary Video 1 showing all the voxel-wise analyses). Threshold set at $P < 0.05$ after family-wise error correction, k $\geqslant$ 150.

positive by the low thresholds and negative by the high thresholds (Fig. 4). This subject has relatively focal retention of PiB in the left temporal and parietal lobes. PiB-PET in the subject with CERAD frequent neuritic plaques was well above the high DVR and SUVR thresholds (Fig. 4).

Sensitivity and specificity for low and high DVR and SUVR thresholds are shown in Table 3. While DVR_{low} and $SUVR_{low}$ showed reasonable sensitivity for CERAD moderate-to-frequent neuritic plaques (81–83%), DVR_{high} and $SUVR_{high}$ thresholds showed surprisingly low sensitivity (~62%). Conversely, all thresholds showed very high specificity. Even when applying liberal thresholds, there was only one false positive result by DVR_{low}, and none applying $SUVR_{low}$. Overall, DVR and SUVR values were highly correlated (r = 0.98, $P < 0.001$), and both measures showed comparable sensitivity and specificity at both low and high thresholds.

Figure 5 shows the relationships between DVR/SUVR values and CERAD scores. For DVR_{high}, there were eight false negatives (compared to four for DVR_{low}), whereas for $SUVR_{high}$ there were nine false negatives (compared to four for $SUVR_{low}$). Similar plots comparing values to Thal amyloid stages and NIA-AA pathological criteria for Alzheimer's disease in the subset of patients in which these were available are shown in Supplementary Fig. 4. While a direct comparison between the SUVR and DVR methods of PiB quantification is outside the scope of our study, it is worth noting that these methods yielded highly correlated results, and performed comparably in predicting post-mortem amyloid burden.

Supplementary Fig. 3 provides examples of misclassified cases. There was only one false positive case in our series. Visual inspection of the image suggests that the borderline positive DVR value may be due to relatively high PiB retention in white matter, which could contaminate grey

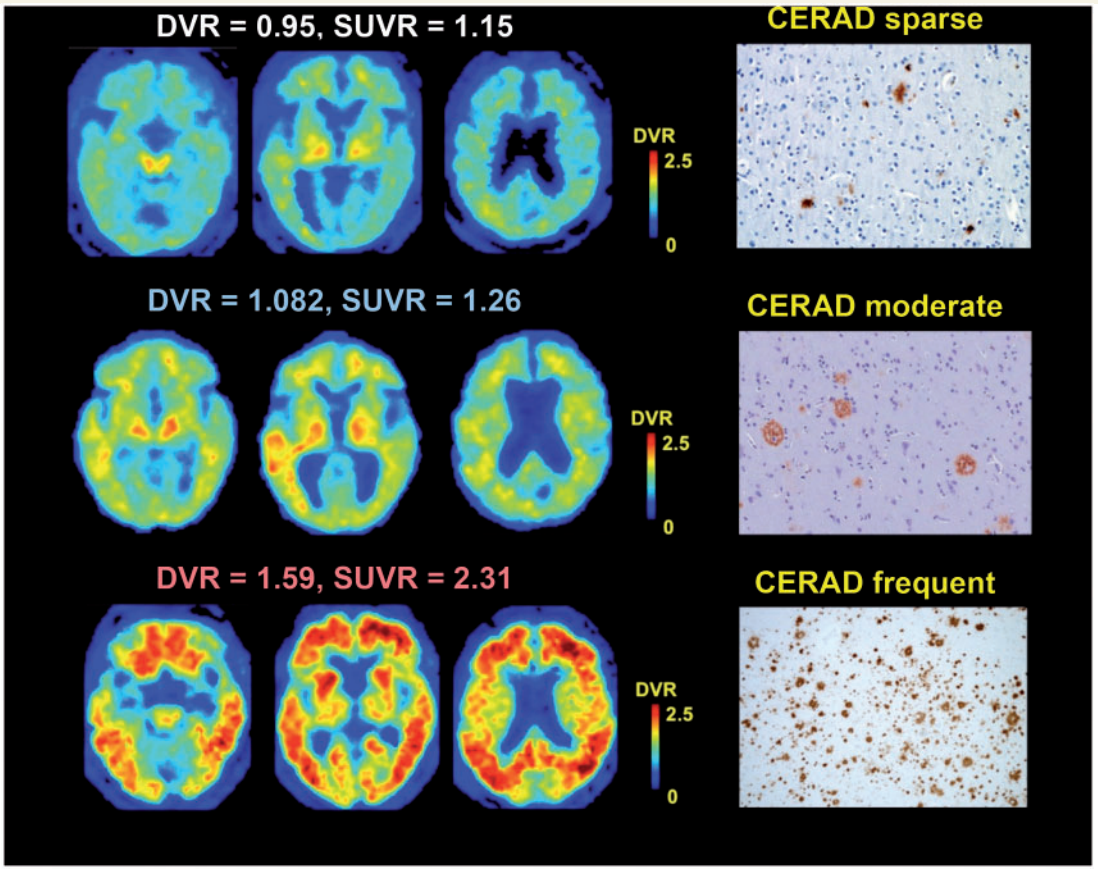

Figure 4 PiB-PET versus post-mortem amyloid. Trans-axial PiB slices from a patient with CERAD sparse (*top row*), moderate (*middle row*) and frequent (*bottom row*) neuritic plaques. PiB-PET trans-axial slices are presented in neurological orientation. Photomicrographs of amyloid-β immunohistochemistry are shown at ×10 (*top* and *bottom rows*) or ×20 (*middle row*) magnification.

Table 3 Sensitivity and specificity for the detection of CERAD moderate-frequent neuritic plaques

	Sensitivity	Specificity
Low thresholds		
DVR = 1.08	**81.0%** (57.4–93.7%)	**95.8%** (76.9–99.8%)
SUVR = 1.21	**83.3%** (61.8–94.5%)	**100.0%** (84.0–100%)
High thresholds		
DVR = 1.20	**61.9%** (38.7–81.0%)	**100.0%** (82.8–100%)
SUVR = 1.40	**62.5%** (40.8–80.4%)	**100.0%** (84.0–100%)

95% confidence intervals shown in parentheses.

matter PiB signal via partial volume effects. Scans were classified as false negatives (by one or both thresholds) for a variety of reasons including focal amyloid accumulation, cortical atrophy and cerebral amyloid angiopathy in the grey matter of the cerebellum (see Supplementary material for more details). PET to pathology intervals were no different in patients that were 'false negatives' by one or more standards versus other patients in the cohort.

The results of the ROC analysis are shown in Supplementary Table 4. The empirically derived optimal DVR (DVR = 1.06, AUC 0.89, sensitivity = 85.7%, specificity = 95.8%) and SUVR (SUVR = 1.20, AUC 0.88, sensitivity = 83.3%, specificity = 100%) thresholds were nearly identical to the *a priori* DVR_{low} and $SUVR_{low}$ thresholds, thus yielding similar sensitivity and specificity values.

Discussion

The selection of threshold values to determine amyloid positivity has important implications for studying mechanisms of Alzheimer's disease pathogenesis and for diagnosis and selection of subjects for clinical trials. Although amyloid-β deposition occurs on a continuum, it is often necessary to dichotomize subjects as amyloid positive or negative. Using different data analysis approaches, we found strong evidence that an optimal threshold for early amyloid-β detection with PiB-PET should be set at a DVR index of 1.08 (or an SUVR of 1.21). Our threshold derivation approaches also provide evidence that a DVR cut-off of 1.20 (SUVR of 1.40) may be valuable when the priority is to minimize false positive subjects. The validity of the low cut-off to detect fibrillar amyloid-β plaque pathology was confirmed by an autopsy study of 50 individuals. In fact, the low DVR threshold of 1.08 was optimal at

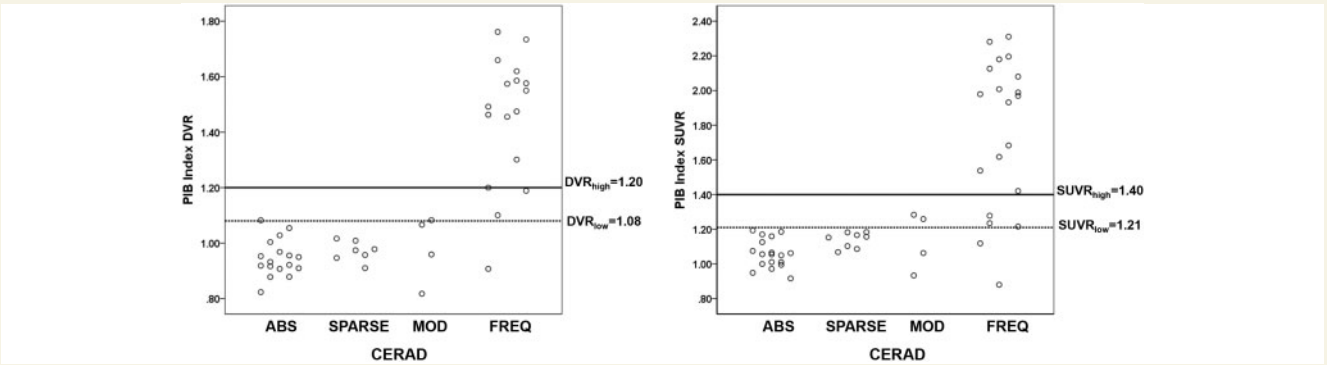

Figure 5 Scatterplots of PiB index DVR (*left*) and SUVR (*right*) versus CERAD rating. Low thresholds are signified by dotted horizontal lines, and high thresholds by solid horizontal lines. ABS = absent; MOD = moderate; FREQ = frequent.

detecting moderate-to-frequent neuritic plaques (specificity = 95.5%; sensitivity = 81%) while the higher DVR threshold was surprisingly insensitive to this burden of amyloid pathology (sensitivity = 61.9%; specificity = 100%). The results in the aggregate suggest that the currently used SUVR thresholds of 1.40 (and higher) and DVR thresholds of 1.20 (and higher) are insensitive and likely misclassify many individuals with substantial amyloid-β neuritic plaque accumulation as amyloid negative.

The threshold derived from a group of young adults highly likely to be free of amyloid-β led to a DVR threshold of 1.07 or a SUVR threshold of 1.19, similar to the threshold defined in a previous study from our group that included a subset of the same subjects (Mormino *et al.*, 2012). The voxel-wise analysis approach also yielded a DVR threshold of 1.07 which represented the lowest mean DVR at which statistical PiB signal was detected. This seemingly low threshold value for DVR index has face validity because regional PiB retention in the cases that would be classified as low positives occurs in medial frontal and medial parietal cortex, brain regions known to show early amyloid-β deposition. Furthermore, both a Gaussian mixture modelling and a cluster analyses suggested a similar DVR cut-off of 1.09 (SUVR = 1.23) to detect early PiB-PET signal. Based on these results, we propose that an optimal threshold for the early detection of amyloid-β with PiB-PET should be a DVR of 1.08, which represents the mean point between 1.07 and 1.09. Based on an independent longitudinal study using a similar processing method (Villemagne *et al.*, 2013), there would be an estimated 7-year time window between our value defining a lower threshold (DVR = 1.08, SUVR = 1.21) and the time when a person reaches an SUVR of 1.40, allowing for potential earlier therapeutic amyloid lowering interventions.

Although a DVR cut-off of 1.08 (SUVR = 1.19) might be optimal to detect early PiB signal, it might not be ideal in all circumstances and should be used with caution. Indeed, a DVR cut-off of 1.08 will likely increase the number of false positive cases due to partial volume effects of white matter binding, or other measurement errors. Selection of thresholds for any given study needs to establish whether false positives or false negatives are more problematic. If a study, or a clinical trial, needs to emphasize specificity over sensitivity our Gaussian mixture modelling and cluster results analyses support the widespread practice of using a DVR of 1.20 (SUVR = 1.40). While this higher threshold might misclassify early accumulators, it will still capture subjects many years before amyloid-β burden approaches a plateau (Villemagne *et al.*, 2011). Labelling individuals who fall below this high threshold as amyloid-β-negative is, however, problematic. In our cognitively normal cohort, 19% of individuals falling below a DVR of 1.20 were classified as positive using the lower threshold. In studies in which it is crucial to detect amyloid-β accumulation as early as possible (e.g. in studies making inferences about neurodegenerative processes that are independent of amyloid-β), low sensitivity can substantially bias conclusions when high thresholds are used.

Relatively few studies have compared PiB-PET results during life with post-mortem amyloid burden. In a small series of six individuals, five cognitively normal and one with dementia, mean cortical DVR values were not strongly associated with neuritic plaque scores, although regional scores, especially in the precuneus, were notable for neuritic plaque accumulation above DVR values of 1.20 (our DVR_{high}) (Sojkova *et al.*, 2011; Driscoll *et al.*, 2012). This sample was too small to estimate sensitivity and specificity. A number of larger studies have assessed thresholds for ^{18}F-labelled amyloid PET tracers in end-of-life populations. In an imaging-pathology series involving 59 subjects studied with the amyloid PET tracer ^{18}F-Florbetapir (Clark *et al.*, 2012), a prespecified threshold SUVR of 1.10 (whole cerebellum reference) had a sensitivity of 97% and specificity of 100% for CERAD moderate-to-frequent neuritic plaques. This threshold is comparable to a PiB SUVR of 1.40 (cerebellar grey reference, or our $SUVR_{high}$) (Landau *et al.*, 2014). It is difficult to directly compare our results to those of ^{18}F-Florbetapir given the many differences not only in tracers but in methods of analysis and subject selection. Recent efforts to standardize and cross-validate amyloid PET values across tracers and analytic methodologies, including the proposed 'Centiloid' scale (Klunk *et al.*,

2015), will help aggregate results from clinicopathological studies of amyloid imaging in an effort to optimize early detection.

In our series, the pathology-defined group with the highest rate of false negatives was the CERAD moderate group, all of whom were classified as negative by the high thresholds. Subjects with intermediate levels of plaque pathology are likely to be the most problematic for detection, and only 8% of our cases (and 15% of cases in the florbetapir series; Clark *et al.*, 2012) were in this category. In contrast, in a large sample of subjects recruited from the community (with normal cognition, mild cognitive impairment, or dementia), intermediate-level Alzheimer's disease pathology (by CERAD and NIA-Reagan criteria) accounted for between 20–30% of cases with Alzheimer's disease as the sole pathological diagnosis (Schneider *et al.*, 2009). Thus, existing studies, which have recruited primarily either end-of-life or tertiary dementia centre populations, seem to undersample a very common intermediate degree of amyloid pathology that is most problematic for image-guided classification. Larger series including more individuals with intermediate pathology will be required to accurately assess imaging-pathological correlations. It is important to note that while a low number of intermediate cases may have biased our study to overestimate sensitivity, our cohort included a larger proportion of individuals with sparse neuritic plaques (16%). Reassuringly, there were no false positives in this group (Fig. 5), which bodes well for the specificity of our proposed low thresholds in future studies.

Many individuals with intermediate DVR/SUVR values (i.e. in between the low and high thresholds) had advanced amyloid pathology at autopsy. Cases in which binding appeared patchy on PET were usually found to have more extensive amyloid deposition at autopsy. Individuals with primarily diffuse plaques or cerebral amyloid angiopathy were not classified as PiB-positive, though PiB binds diffuse plaques and cerebral amyloid angiopathy *in vitro* (Lockhart *et al.*, 2007), and rare 'false positive' PiB scans have been reported in individuals with florid deposition of diffuse plaques or pure cerebral amyloid angiopathy (Kantarci *et al.*, 2012; Ducharme *et al.*, 2013). Overall we conclude that PiB primarily detects relatively advanced amyloid-β neuritic plaque deposition, and that even early PiB signal may merely represent the 'tip of the iceberg' of underlying amyloid pathology. This observation suggests that a negative scan, particularly with widely used high thresholds, does not exclude amyloid pathology.

A secondary goal of this study was to examine the spatial pattern of earliest amyloid-β accumulation detectable with PiB. To capture 'progression' of amyloid-β signal, we ranked subjects based on their PiB Index DVR values and performed iterative voxel-wise contrasts with a low DVR group using a sliding window. We found that early PiB binding is seen in anterior cingulate cortex and other medial frontal regions (Fig. 3 and Supplementary Video 1), consistent with previous reports in preclinical Alzheimer's disease (Sperling *et al.*, 2009). The anterior cingulate and precuneus were also the regions showing the highest discriminant power in in the cluster analyses (Fig. 2). Anterior cingulate and precuneus are highly inter-connected cortical hubs, and early PiB signal in these regions reinforces the hypothesis that regions of high connectivity are prone to earlier amyloid-β deposition (Buckner *et al.*, 2009). Overall, our cross-sectional analysis suggests that amyloid-β may spread from medial to lateral frontal and parietal regions, later involving the lateral temporal lobe. By the time a person reaches the widely used SUVR thresholds of 1.40 and 1.50, amyloid-β is already widespread across most of the association cortex, relatively sparing the occipital lobes and unimodal processing regions. Of course these longitudinal inferences are based on cross-sectional data and should thus be interpreted with caution. Furthermore, as PiB mainly binds to the fibrillar form of amyloid-β, we cannot exclude the possibility that other forms of amyloid-β may have different spatial patterns.

The major limitation of all studies trying to set a threshold for amyloid-β positivity is that these thresholds are highly dependent on the specific imaging methodology used. This includes the radiotracer used, the time period following injection when the images are acquired, the reference region, and whether partial volume correction is employed. For instance, previous studies showed that applying partial volume correction increased PiB values by 10 to 20% in normal controls (Rabinovici *et al.*, 2010; Villemagne *et al.*, 2011) while choosing the whole cerebellum as the reference region decreased the values by ~6% (unpublished data). As our four analytical methods gave almost identical thresholds, one can argue that all of the techniques used in this study were equivalent. Therefore, using a group of young adults might be an efficient way to set a low threshold. Our autopsy validation cohort consisted mainly of patients with dementia, many of whom had significant cortical atrophy. This could potentially lower overall PiB retention values due to partial volume effects, thus underestimating sensitivity and overestimating specificity of low thresholds when they are applied to cognitively normal individuals who will likely have less brain atrophy. Many autopsies were performed before publication of the updated NIA-AA pathological criteria for Alzheimer's disease, and Thal amyloid stages were available only in a subset of cases. Our mean PET to autopsy interval was relatively long compared to previous studies. Although longer PET to post-mortem intervals have been proposed as a possible explanation for false negative cases (presumably due to continued amyloid-β aggregation between the scan and death) (Clark *et al.*, 2012), we found no evidence of this in our series, with no difference in the PET-to-autopsy interval between false negative cases and the rest of the cohort. PiB index was measured using different processing pipelines in the derivation and validation cohorts. However, these processing streams yield highly correlated values, and the robustness of the

low thresholds across two processing platforms strengthens our results. We also did not include the striatum in the PiB index—this region may be sensitive to early amyloid-β accumulation in autosomal dominant forms of Alzheimer's disease (Bateman *et al.*, 2012). Finally, our study was designed to define and test thresholds for mean cortical PiB retention. Approaches that examine region-specific binding should be explored, as they may detect early signal prior to more global values becoming positive.

In summary we found that the low PiB DVR = 1.08 (or SUVR = 1.21) thresholds showed higher sensitivity than the higher thresholds typically applied in the literature (DVR = 1.20; SUVR = 1.40) without compromising specificity. Further PET to pathology correlative studies are needed to validate these findings, and to define optimal thresholds for other amyloid-β tracers and image analysis approaches.

Acknowledgements

We would like to acknowledge Suzanne Baker, Mustafa Janabi, Kris Norton and James O'Neil for their support with PET scanning; Dennis Dickson, Mario Mendez, John Trojanowski and Harry Vinters for patient referrals and autopsies; and Helaine St-Amant and Angie Yi for help with data preparation and helpful discussions throughout the project.

Funding

This work was supported by National Institute on Aging grants K23-AG031861 and R01-AG045611 to G.D.R., P01-AG1972403 to B.LM. and W.W.S, P50-AG023501 to B.L.M., G.D.R and W.W.S., R01-AG032306 and K24-AG045333 to H.J.R, P01-AG12435, P30-AG10129, R01-AG021028 and R01-AG031563 to C.D., R01-AG031563 to B.R.; R01AG038791 to A.L.B., R01-AG034570 to WJJ; the Consortium for Frontotemporal Dementia Research to B.L.M. and W.W.S; the Tau Consortium to W.W.S., G.D.R and W.J.J.; John Douglas French Alzheimer's Foundation to G.D.R. and B.L.M.; State of California Department of Health Services Alzheimer's Disease Research Center of California grant 04-33516 to B.L.M; Hellman Family Foundation award to G.D.R.; and Canadian Institutes of Health Research post-doctoral fellowship to S.V.

Supplementary material

Supplementary material is available at *Brain* online.

References

Aizenstein HJ, Nebes RD, Saxton JA, Price JC, Mathis CA, Tsopelas ND, et al. Frequent amyloid deposition without significant cognitive impairment among the elderly. Arch Neurol 2008; 65: 1509–17.

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 270–9.

Armstrong MJ, Litvan I, Lang AE, Bak TH, Bhatia KP, Borroni B, et al. Criteria for the diagnosis of corticobasal degeneration. Neurology 2013; 80: 496–503.

Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med. 2012; 367: 795–804.

Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med 2012; 367: 795–804.

Braak H, Braak E. Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. Neurobiol Aging 1998; 18 (4 Suppl 1): S85–S8.

Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu H, Hedden T, et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J Neurosci 2009; 29: 1860–73.

Chui HC, Zarow C, Mack WJ, Ellis WG, Zheng L, Jagust WJ, et al. Cognitive impact of subcortical vascular and Alzheimer's disease pathology. Ann Neurol 2006; 60: 677–87.

Clark CM, Pontecorvo MJ, Beach TG, Bedell BJ, Coleman RE, Doraiswamy PM, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study. Lancet Neurol 2012; 11: 669–78.

Clark CM, Schneider JA, Bedell BJ, Beach TG, Bilker WB, Mintun MA, et al. Use of florbetapir-PET for imaging beta-amyloid pathology. JAMA 2011; 305: 275–83.

Curtis C, Gamez JE, Singh U, Sadowsky CH, Villena T, Sabbagh MN, et al. Phase 3 trial of flutemetamol labeled with radioactive fluorine 18 imaging and neuritic plaque density. JAMA Neurol 2015; 72: 287–94.

Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 2006; 31: 968–80.

Driscoll I, Troncoso JC, Rudow G, Sojkova J, Pletnikova O, Zhou Y, et al. Correspondence between *in vivo* (11)C-PiB-PET amyloid imaging and postmortem, region-matched assessment of plaques. Acta Neuropathol 2012; 124: 823–31.

Ducharme S, Guiot MC, Nikelski J, Chertkow H. Does a positive Pittsburgh Compound B scan in a patient with dementia equal Alzheimer disease? JAMA Neurol 2013; 70: 912–4.

Fleisher AS, Chen K, Quiroz YT, Jakimovich LJ, Gomez MG, Langois CM, et al. Florbetapir PET analysis of amyloid-beta deposition in the presenilin 1 E280A autosomal dominant Alzheimer's disease kindred: a cross-sectional study. Lancet Neurol 2012; 11: 1057–65.

Forman MS, Farmer J, Johnson JK, Clark CM, Arnold SE, Coslett HB, et al. Frontotemporal dementia: clinicopathological correlations. Ann Neurol 2006; 59: 952–62.

Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. Neurology 2011; 76: 1006–14.

Grinberg LT, Wang X, Wang C, Sohn PD, Theofilas P, Sidhu M, et al. Argyrophilic grain disease differs from other tauopathies by lacking tau acetylation. Acta Neuropathol 2013; 125: 581–93.

Hedden T, Van Dijk KR, Becker JA, Mehta A, Sperling RA, Johnson KA, et al. Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. J Neurosci 2009; 29: 12686–94.

Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimers Dement 2012; 8: 1–13.

Hyman BT, Trojanowski JQ. Consensus recommendations for the postmortem diagnosis of Alzheimer disease from the National Institute on Aging and the Reagan Institute Working Group on diagnostic criteria for the neuropathological assessment of Alzheimer disease. J Neuropathol Exp Neurol 1997; 56: 1095–7.

Ikonomovic MD, Abrahamson EE, Price JC, Hamilton RL, Mathis CA, Paljug WR, et al. Early AD pathology in a [C-11]PiB-negative case: a PiB-amyloid imaging, biochemical, and immunohistochemical study. Acta Neuropathol 2012; 123: 433–47.

Ikonomovic MD, Klunk WE, Abrahamson EE, Mathis CA, Price JC, Tsopelas ND, et al. Post-mortem correlates of *in vivo* PiB-PET amyloid imaging in a typical case of Alzheimer's disease. Brain 2008; 131 (Pt 6): 1630–45.

Jack CR, Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol 2013; 12: 207–16.

Jack CR, Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010; 9: 119–28.

Jack CR, Jr, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, et al. An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. Ann Neurol 2012; 71: 765–75.

Jack CR, Jr, Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, et al. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain 2008; 131 (Pt 3): 665–80.

Jack CR, Jr, Wiste HJ, Lesnick TG, Weigand SD, Knopman DS, Vemuri P, et al. Brain beta-amyloid load approaches a plateau. Neurology 2013; 80: 890–6.

Kantarci K, Yang C, Schneider JA, Senjem ML, Reyes DA, Lowe VJ, et al. Antemortem amyloid imaging and beta-amyloid pathology in a case with dementia with Lewy bodies. Neurobiol Aging 2012; 33: 878–85.

Klunk WE, Engler H, Nordberg A, Wang Y, Blomqvist G, Holt DP, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol 2004; 55: 306–19.

Klunk WE, Koeppe RA, Price JC, Benzinger TL, Devous MD, Sr., Jagust WJ, et al. The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement 2015; 11: 1–15, e4.

Kok E, Haikonen S, Luoto T, Huhtala H, Goebeler S, Haapasalo H, et al. Apolipoprotein E-dependent accumulation of Alzheimer disease-related lesions begins in middle age. Ann Neurol 2009; 65: 650–7.

Landau SM, Thomas BA, Thurfjell L, Schmidt M, Margolin R, Mintun M, et al. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. Eur J Nucl Med Mol Imaging 2014; 41: 1398–407.

Lockhart A, Lamb JR, Osredkar T, Sue LI, Joyce JN, Ye L, et al. PIB is a non-specific imaging marker of amyloid-beta (Abeta) peptide-related cerebral amyloidosis. Brain 2007; 130 (Pt 10): 2607–15.

Logan J, Fowler JS, Volkow ND, Wang GJ, Ding YS, Alexoff DL. Distribution volume ratios without blood sampling from graphical analysis of PET data. J Cereb Blood Flow Metabol 1996; 16: 834–40.

Mackenzie IR, Neumann M, Bigio EH, Cairns NJ, Alafuzoff I, Kril J, et al. Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. Acta Neuropathol 2010; 119: 1–4.

Magaki S, Yong WH, Khanlou N, Tung S, Vinters HV. Comorbidity in dementia: update of an ongoing autopsy study. J Am Geriatr Soc 2014; 62: 1722–8.

McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. Neurology 2005; 65: 1863–72.

McKhann G, Drachman DA, Folstein M, Katzman R, Price D, Stadlan EM. 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 1984; 34: 939–44.

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 263–9.

Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 1991; 41: 479–86.

Montine TJ, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Dickson DW, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. Acta Neuropathol 2012; 123: 1–11.

Mormino EC, Betensky RA, Hedden T, Schultz AP, Ward A, Huijbers W, et al. Amyloid and APOE {varepsilon}4 interact to influence short-term decline in preclinical Alzheimer disease. Neurology 2014; 82: 1760–7.

Mormino EC, Brandel MG, Madison CM, Rabinovici GD, Marks S, Baker SL, et al. Not quite PIB-positive, not quite PIB-negative: slight PIB elevations in elderly normal control subjects are biologically relevant. NeuroImage 2012; 59: 1152–60.

Murray ME, Cannon A, Graff-Radford NR, Liesinger AM, Rutherford NJ, Ross OA, et al. Differential clinicopathologic and genetic features of late-onset amnestic dementias. Acta Neuropathol 2014; 128: 411–21.

Newell KL, Hyman BT, Growdon JH, Hedley-Whyte ET. Application of the National Institute on Aging (NIA)-Reagan Institute criteria for the neuropathological diagnosis of Alzheimer disease. J Neuropathol Exp Neurol 1999; 58: 1147–55.

Ni R, Gillberg PG, Bergfors A, Marutle A, Nordberg A. Amyloid tracers detect multiple binding sites in Alzheimer's disease brain tissue. Brain 2013; 136 (Pt 7): 2217–27.

Nordberg A, Carter SF, Rinne J, Drzezga A, Brooks DJ, Vandenberghe R, et al. A European multicentre PET study of fibrillar amyloid in Alzheimer's disease. Eur J Nucl Med Mol Imaging. 2013; 40: 104–14.

Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004; 256: 183–94.

Pike KE, Savage G, Villemagne VL, Ng S, Moss SA, Maruff P, et al. Beta-amyloid imaging and memory in non-demented individuals: evidence for preclinical Alzheimer's disease. Brain 2007; 130 (Pt 11): 2837–44.

Price JC, Klunk WE, Lopresti BJ, Lu X, Hoge JA, Ziolko SK, et al. Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh Compound-B. J Cereb Blood Flow Metab 2005; 25: 1528–47.

Rabinovici GD, Furst AJ, Alkalay A, Racine CA, O'Neil JP, Janabi M, et al. Increased metabolic vulnerability in early-onset Alzheimer's disease is not related to amyloid burden. Brain 2010; 133 (Pt 2): 512–28.

Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 2011; 134 (Pt 9): 2456–77.

Roman GC, Tatemichi TK, Erkinjuntii T. Vascular dementia: diagnostic criteria for research studies. Neurology 1993; 43: 250–60.

Rowe CC, Ellis KA, Rimajova M, Bourgeat P, Pike KE, Jones G, et al. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiology of aging 2010; 31: 1275–83.

Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. N Engl J Med 2014; 370: 322–33.

Schneider JA, Arvanitakis Z, Leurgans SE, Bennett DA. The neuropathology of probable Alzheimer disease and mild cognitive impairment. Ann Neurol 2009; 66: 200–8.

Sojkova J, Driscoll I, Iacono D, Zhou Y, Codispoti KE, Kraut MA, et al. *In vivo* fibrillar beta-amyloid detected using [11C]PiB positron emission tomography and neuropathologic assessment in older adults. Arch Neurol 2011; 68: 232–40.

Sperling RA, Laviolette PS, O'Keefe K, O'Brien J, Rentz DM, Pihlajamaki M, et al. Amyloid deposition is associated with impaired default network function in older persons without dementia. Neuron 2009; 63: 178–88.

Sun FT, Schriber RA, Greenia JM, He J, Gitcho A, Jagust WJ. Automated template-based PET region of interest analyses in the aging brain. NeuroImage 2007; 34: 608–17.

The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. Neurobiol Aging 1997; 18 (4 Suppl): S1–2.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 2002; 15: 273–89.

Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol 2013; 12: 357–67.

Villemagne VL, Pike KE, Chetelat G, Ellis KA, Mulligan RS, Bourgeat P, et al. Longitudinal assessment of Abeta and cognition in aging and Alzheimer disease. Ann Neurol 2011; 69: 181–92.

Villeneuve S, Reed BR, Wirth M, Haase CM, Madison CM, Ayakta N, et al. Cortical thickness mediates the effect of beta-amyloid on episodic memory. Neurology 2014; 82: 761–7.