Use in diagnosis of Alzheimer’s disease
Chapter 7

Molecular imaging in the diagnosis of Alzheimer’s disease: visual interpretation of $[^{11}\text{C}]$PIB and $[^{18}\text{F}]$FDDNP PET images

Nelleke Tolboom, Wiesje M. van der Flier, Jolanda Boverhoff, Maqsood Yaqub, Mike Wattjes, Pieter G. Rajmakers, Frederik Barkhof, Philip Scheltens, Karl Herholz, Adriaan A. Lammertsma, Bart N.M. van Berckel

Submitted
Abstract

Objective: To evaluate visual interpretation of $[^{11}C]$PIB and $[^{18}F]$FDDNP PET images as potential supportive diagnostic markers for Alzheimer’s disease (AD).

Methods: Twenty-one AD patients and 20 controls were included. Parametric $[^{11}C]$PIB and $[^{18}F]$FDDNP BP$_{ND}$ images were visually rated as ‘AD’ or ‘normal’. Data were compared with ratings of $[^{18}F]$FDG PET images and MRI derived medial temporal lobe atrophy (MTA) scores. Inter-rater agreement and agreement with clinical diagnosis was assessed for all imaging modalities. In addition, cut-off values for quantitative global $[^{11}C]$PIB and $[^{18}F]$FDDNP BP$_{ND}$ were determined. Visual ratings were compared with dichotomised quantitative values.

Results: Agreement between readers was excellent for $[^{11}C]$PIB, $[^{18}F]$FDDNP and MTA (Cohen’s kappa $\kappa \geq 0.85$) and moderate for $[^{18}F]$FDG ($\kappa = 0.56$). Highest sensitivity was found for $[^{11}C]$PIB and $[^{18}F]$FDG (both 1.0). Highest specificity was found for MTA (0.90) and $[^{11}C]$PIB (0.85). $[^{18}F]$FDDNP had lowest sensitivity and specificity (0.67 and 0.53, respectively). Cut-off for quantitative $[^{11}C]$PIB BP$_{ND}$ was 0.54 (sensitivity and specificity both 0.95) and for $[^{18}F]$FDDNP BP$_{ND}$ 0.07 (sensitivity 0.80, specificity 0.73). Agreement between quantitative analyses and visual ratings was excellent for $[^{11}C]$PIB ($\kappa = 0.85$) and fair for $[^{18}F]$FDDNP ($\kappa = 0.40$). For both younger (<65 years) and older subjects (≥65), the same modalities performed best with respect to sensitivity and specificity.

Conclusion: Visual interpretation of $[^{11}C]$PIB images was easy and accurate, showing promise as a supportive diagnostic marker for AD. Moreover, $[^{11}C]$PIB showed the best combination of sensitivity and specificity. Visual interpretation of $[^{18}F]$FDDNP images was insufficient. Quantitative analysis of $[^{18}F]$FDDNP data showed considerable higher diagnostic value than visual analysis.
**Introduction**

The current diagnosis of Alzheimer’s disease (AD) primarily depends on clinical evaluation (1). Auxiliary investigations may have additional value, as they can increase sensitivity for identifying AD, especially at early stages (2). This is particularly true for those that are easy to use in clinical practice, for instance by markers that enable visual identification of patterns suggestive of AD. Current guidelines recommend neuroimaging at least once during diagnostic work-up (3). Two established methods are medial temporal lobe atrophy (MTA) using magnetic resonance imaging (MRI) (4) and decreased temporoparietal cerebral glucose metabolism using [18F]FDG positron emission tomography (PET) (5). Visual assessment of MTA reliably separates AD patients from normal age matched controls, with a sensitivity and specificity greater than 85% (4,6,7). Visual interpretation of [18F]FDG PET has a sensitivity greater than 85% and a specificity between 70 and 90% for the detection of AD (8,9). Nevertheless, both markers are also affected by normal aging (10-12), which complicates visual assessment. Furthermore, MTA and reduced cerebral glucose metabolism are only indirect measures of disease. Methods that visualise AD pathology directly could therefore improve the diagnostic evaluation of patients with AD. Recently, several PET tracers have been developed for visualising AD pathology directly. Of these ligands, [11C]PIB (Pittsburgh Compound-B) (13) and [18F]FDDNP (2-(1-{6-[(2-[F-18]fluoroethyl)(methyl)amino]-2-aphthyl} ethylidene) malononitrile) (14) have been used most widely. Both tracers have shown the ability to distinguish AD patients from controls at a group level (13-16). However, little is known about accuracy and reliability of visual interpretation of images acquired using these new biomarkers. To date, only one study has compared visual interpretation of [18F]FDG and [11C]PIB images, reporting better diagnostic accuracy for [11C]PIB (17). For [18F]FDDNP, no reports on visual interpretation of images have been published.

The aim of the present study was to evaluate visual interpretation of [11C]PIB and [18F]FDDNP images as potential supportive diagnostic markers for AD. Both diagnostic accuracy and inter-observer variability of [11C]PIB and [18F]FDDNP were evaluated. Results were compared with those of visual assessment of [18F]FDG and with MTA. In addition, visual interpretation of [11C]PIB and [18F]FDDNP images was compared with quantitative assessment of global binding.

**Methods**

**Subjects**

Twenty-one AD patients (mean±SD 63±6 years) and 20 controls (67±6 years) were included in this study. Global and regional [11C]PIB and [18F]FDDNP binding in a largely overlapping sample have been reported elsewhere (16). [11C]PIB and MRI scans were available for all subjects. [18F]FDDNP PET scans were not available for 5 AD patients and 5 controls and [18F]FDG scans were not available for 5 controls.
All patients received a standard dementia screening that included medical history, physical and neurological examinations, screening laboratory tests, and brain MRI. Clinical diagnosis was established by consensus in a multidisciplinary team, without knowledge of PET results. All AD patients met NINCDS-ADRDA criteria for probable AD (1). Controls were recruited through advertisements in newspapers, without mentioning the specific reason of the PET research, and underwent the same diagnostic procedures. Exclusion criteria were history of major psychiatric or neurological (other than AD) illness and use of non-steroidal anti-inflammatory drugs (NSAIDS), as the latter have been reported to compete with [18F]FDDNP for binding to Aβ fibrils in vitro and to Aβ plaques ex vivo (18). Additional exclusion criteria for controls were subjective memory complaints or clinically significant abnormalities on the MRI scan (as determined by a neuroradiologist). Written informed consent was obtained from all subjects after a complete written and verbal description of the study. The study was approved by the Medical Ethics Review Committee of the VU University Medical Centre Amsterdam.

**Positron emission tomography**

Using an ECAT EXACT HR+ scanner (19) (Siemens/CTI, Knoxville, USA), dynamic 90 minutes [11C]PIB and [18F]FDDNP PET scans were performed as described previously (16). All subjects who were imaged using both ligands underwent scanning on the same day, except for 4 subjects who had an average interval of 5½ weeks between scans due to radiosynthesis failure on the planned date. In the same week as [11C]PIB and/or [18F]FDDNP scanning, static [18F]FDG scans were acquired. Following transmission scanning, a static 15 minutes emission scan was performed 45 minutes after intravenous injection of 185 MBq of [18F]FDG. Subjects were in resting state with eyes closed and ears unplugged, and scanning took place in a dimly lit room with minimal background noise.

**Magnetic resonance imaging**

All subjects underwent a structural MRI scan using a 1.5 T Sonata scanner (Siemens Medical Solutions, Erlangen, Germany). The scan protocol included a coronal T1-weighted 3D MPRAGE (magnetization prepared rapid acquisition gradient echo; slice thickness 1.5 mm, 160 slices; matrix size 256x256; voxel size 1.0x1.0x1.5 mm; echo time 3.97 ms; repetition time 2700 ms; inversion time 950 ms; flip angle 8°). Oblique/coronal (perpendicular to the long axis of the hippocampus), sagittal and axial 3mm formats were obtained from these data.

**Image analysis**

Quantitative analysis of [11C]PIB and [18F]FDDNP PET scans was performed as described previously (16). In short, parametric images of BPND were generated by applying a basis function implementation of the “two-step” simplified reference tissue model (RPM2) with cerebellar grey matter as reference tissue. The following grey matter ROIs were defined using PVE lab (20): cerebellum, frontal (volume weighted average of orbital frontal, medial inferior frontal and
superior frontal), parietal, temporal (volume weighted average of superior temporal and medial inferior temporal) cortex, medial temporal lobe (MTL) (volume weighted average of enthorinal cortex and hippocampus), and posterior cingulate. A global cortical ROI was defined, based on a volume weighted BPND average of the aforementioned ROIs, excluding for cerebellum. To aid visual assessment of $^{18}$F-FDG, images were analysed using a previously validated automated discrimination tool (PMOD Alzheimer discrimination tool) (21). This discrimination analysis was based on a comparison of age corrected measured $^{18}$F-FDG uptake in each voxel with the predicted uptake derived from a dataset of healthy elderly subjects. Resulting deviations were expressed as t-values.

**Visual readings**

All scans were presented to the readers in a randomised order. All $^{11}$C-PIB, $^{18}$F-FDDNP and $^{18}$F-FDG scans were rated separately by 2 readers (BvB and PR for $^{11}$C-PIB and $^{18}$F-FDDNP; BvB and KH for $^{18}$F-FDG) and classified as either AD or normal (i.e. unlikely to be AD). For rating of $^{18}$F-FDG, readers had access to results of the Alzheimer discrimination tool (21), but the final decision was based on their own assessment. Readers of PET scans were nuclear medicine physicians (BB, PR) and a neurologist (KH) with expertise in PET imaging and they were blinded to clinical information. Results of individual readers were compared and discrepancies were discussed to establish consensus.

Atrophy rating on T1 weighted MRI was separately performed by 2 trained neuroradiologists (FB and MW), who were blinded to clinical information, but had knowledge of the age of the subject. MTA was rated visually on the oblique/coronal images using a scale ranging from 0 (no atrophy) to 4 (severe atrophy) (4) Averaged MTA scores were dichotomised according to age. MTA scores were considered as abnormal depending on age (years), i.e. ≥1 for subjects <65, ≥ 1.5 for subjects ≥65 and < 75, and ≥ 2 for subjects ≥ 75 years. MTA ratings of the most experienced reader (FB) were used for further analysis.

**Statistical analysis**

Data are presented as mean±SD, unless otherwise stated. Frequency distributions for gender were compared using the Chi-square test. Differences between the two groups were assessed using analysis of variance (ANOVA) with age and gender as covariates. To assess consistency of visual interpretations, Cohen's Kappa (κ) for agreement between readers was calculated for all imaging modalities. A kappa <0.20 denotes poor agreement, 0.20≤κ<0.40 fair, 0.40≤κ<0.60 moderate, 0.60≤κ<0.80 substantial, and >0.80 excellent agreement (22). Sensitivity, specificity, likelihood ratios and accuracy were calculated for visual interpretation of $^{11}$C-PIB, $^{18}$F-FDDNP and $^{18}$F-FDG images, and for MTA. Clinical diagnosis was used as reference criterion. Calculations were performed as follows: sensitivity = AD patients with abnormal images divided by total number of AD patients; specificity = controls with normal images divided by total number of control subjects; positive likelihood ratio LR+ = sensitivity/(1-specificity); negative likelihood ratio LR- = (1-sensitivity)/specificity; accuracy=
AD patients with abnormal images plus controls with normal images divided by the total number of individuals investigated.

Next, agreement between visual interpretation of \([^{11}C]PIB\) and \([^{18}F]FDDNP\) images and corresponding quantitative values of global binding potential (BP\(_{ND}\)) was assessed. To this end, receiver-operating-characteristic (ROC) curves were generated to determine optimal cut-off values for global \([^{11}C]PIB\) and \([^{18}F]FDDNP\) BP\(_{ND}\). Optimal cut-off values were defined as those values that yielded at least 80% sensitivity, with accompanying specificity for detecting AD \((23)\), using clinical diagnosis as reference criterion. Based on these ROC defined cut-off values, quantitative BP\(_{ND}\) values of all subjects were classified as ‘normal’ or ‘abnormal’ and agreement between quantitative and visual assessments was assessed using Cohen’s Kappa.

Finally, to assess the effects of age on visual interpretation, agreement between readers, agreement with clinical diagnosis and agreement with quantitative values were also performed after the whole sample was divided into younger (<65 years) and older subjects (≥65 years). This cut-off was chosen as it is regarded to be the division between early and late onset of AD \((24)\).

**Results**

Demographic and clinical characteristics according to diagnostic group are presented in Table 1. The two groups did not differ with respect to age or sex, although AD patients tended to be slightly younger than healthy controls. As expected, AD patients had lower average MMSE scores, higher MTA scores, and higher global cortical \([^{11}C]PIB\) and \([^{18}F]FDDNP\) BP\(_{ND}\) values.

**Agreement between readers**

Agreement between visual readings was excellent for \([^{11}C]PIB\) \((\kappa=0.85)\), \([^{18}F]FDDNP\) \((\kappa=0.87)\) and MTA \((\kappa=0.90)\), and moderate for \([^{18}F]FDG\) \((\kappa=0.56)\).

Figure 1 shows examples of normal and AD-like images for all modalities.

**Table 1:** Demographic and clinical characteristics of the two diagnostic groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Controls</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (whole sample)</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Age, year</td>
<td>67±6</td>
<td>63±6</td>
</tr>
<tr>
<td>Sex, f/m</td>
<td>7/13</td>
<td>7/14</td>
</tr>
<tr>
<td>MMSE</td>
<td>29±1</td>
<td>23±2**</td>
</tr>
<tr>
<td>MTA</td>
<td>0.5±0.6</td>
<td>1.5±0.9**</td>
</tr>
<tr>
<td>Global ([^{11}C]PIB) BP(_{ND})</td>
<td>0.13±0.22</td>
<td>0.84±0.20**</td>
</tr>
<tr>
<td>Global ([^{18}F]FDDNP) BP(_{ND})</td>
<td>0.06±0.03</td>
<td>0.09±0.02*</td>
</tr>
</tbody>
</table>

Data are presented as mean ± SD, where appropriate.

AD, Alzheimer’s disease; f, female; m, male; MMSE, mini-mental-state examination; MTA, medial temporal lobe atrophy BP\(_{ND}\); binding potential. MTA, medial temporal lobe atrophy

\(*p<0.05\)

\(**p<0.01\)
Agreement with clinical diagnosis
Sensitivity, specificity, LR+, LR- and accuracy for visual interpretation of all imaging modalities are presented in Table 2. $[^{11}C]PIB$ and $[^{18}F]FDG$ had the highest sensitivity (all patients were identified correctly). Lowest sensitivity was obtained for $[^{18}F]FDDNP$ and MTA. Highest specificity was found for MTA and $[^{11}C]PIB$. $[^{18}F]FDG$ and $[^{18}F]FDDNP$ had low specificity. Sensitivity and specificity were reflected in LR as highest LR+ for MTA and best LR- for both $[^{11}C]PIB$ and $[^{18}F]FDG$. Highest accuracy was found for $[^{11}C]PIB$ whilst it was lowest for $[^{18}F]FDDNP$. MTA and $[^{18}F]FDG$ has similar accuracy.

Figure 1: Examples of normal (left panel, healthy control) and AD-like (right panel, AD patient) images for $[^{11}C]PIB$, $[^{18}F]FDDNP$, $[^{18}F]FDG$ and MTA on MRI. All scans were acquired in the same subjects. Uptake is represented by the colour scale for the PET images. Increased $[^{11}C]PIB$ and $[^{18}F]FDDNP$ uptake is seen mainly in frontal regions on sagittal images. Decreased $[^{18}F]FDG$ is seen mainly in parietotemporal regions on the axial image. MTA is best seen on coronal images.
Chapter 7

Agreement with quantitative analysis

Figure 2 shows ROC curves for both ligands. The $[^{11}C]$PIB cut-off value was determined at a BPND of 0.54. The area under the curve (AUC) was 0.96 (95% confidence interval (CI) 0.90-1.03), with both sensitivity and specificity of 0.95. These values differed only slightly from sensitivity and specificity of visual ratings. For global $[^{18}F]$FDDNP binding the optimal cut-off value was found to be at a BPND of 0.07 resulting in a sensitivity of 0.80, a specificity of 0.73 and an AUC (95% confidence interval 0.86-0.97). Table 2 displays sensitivity, specificity, likelihood ratios and accuracy for visual ratings across AD patients and controls.

Table 2: Sensitivity, specificity, likelihood ratios and accuracy for visual ratings across AD patients and controls.

<table>
<thead>
<tr>
<th>Visual rating</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR+</th>
<th>LR-</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[^{11}C]$PIB</td>
<td>1.0 (21/21)</td>
<td>0.85 (17/20)</td>
<td>6.7</td>
<td>0</td>
<td>0.93</td>
</tr>
<tr>
<td>$[^{18}F]$FDDNP</td>
<td>0.67 (10/15)</td>
<td>0.53 (8/15)</td>
<td>1.5</td>
<td>0.6</td>
<td>0.60</td>
</tr>
<tr>
<td>$[^{18}F]$FDG</td>
<td>1.00 (21/21)</td>
<td>0.56 (9/16)</td>
<td>2.3</td>
<td>0</td>
<td>0.81</td>
</tr>
<tr>
<td>MTA</td>
<td>0.71 (15/21)</td>
<td>0.90 (18/20)</td>
<td>7.1</td>
<td>0.3</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Young (<65 years)

<table>
<thead>
<tr>
<th>Visual rating</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR+</th>
<th>LR-</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[^{11}C]$PIB</td>
<td>1.00 (13/13)</td>
<td>1.00 (8/8)</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
</tr>
<tr>
<td>$[^{18}F]$FDDNP</td>
<td>0.75 (6/8)</td>
<td>0.60 (3/5)</td>
<td>1.9</td>
<td>0.4</td>
<td>0.69</td>
</tr>
<tr>
<td>$[^{18}F]$FDG</td>
<td>1.00 (13/13)</td>
<td>0.67 (4/6)</td>
<td>3.0</td>
<td>-</td>
<td>0.89</td>
</tr>
<tr>
<td>MTA</td>
<td>0.69 (9/13)</td>
<td>1.00 (8/8)</td>
<td>-</td>
<td>0.3</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Old (≥65 years)

<table>
<thead>
<tr>
<th>Visual rating</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR+</th>
<th>LR-</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[^{11}C]$PIB</td>
<td>1.00 (8/8)</td>
<td>0.75 (9/12)</td>
<td>4</td>
<td>-</td>
<td>0.85</td>
</tr>
<tr>
<td>$[^{18}F]$FDDNP</td>
<td>0.57 (4/7)</td>
<td>0.50 (5/10)</td>
<td>1.1</td>
<td>0.9</td>
<td>0.53</td>
</tr>
<tr>
<td>$[^{18}F]$FDG</td>
<td>1.00 (8/8)</td>
<td>0.50 (5/10)</td>
<td>2</td>
<td>-</td>
<td>0.72</td>
</tr>
<tr>
<td>MTA</td>
<td>0.75 (6/8)</td>
<td>0.83 (10/12)</td>
<td>4.4</td>
<td>0.3</td>
<td>0.80</td>
</tr>
</tbody>
</table>

MTA; medial temporal lobe atrophy, LR+; positive likelihood ratio, LR-; negative likelihood ratio
Numbers between brackets represent numbers of patients
If sensitivity =1, LR- =0, if specificity=1, LR+ =0

Figure 2: Receiver-operator-characteristic (ROC) curves for global $[^{11}C]$PIB and $[^{18}F]$FDDNP BPND. Cut-off values were determined based on the highest sum of sensitivity and specificity with a minimal sensitivity of 80%. Solid line represents $[^{11}C]$PIB and dashed line $[^{18}F]$FDDNP.
Agreement of quantitative assessment between both ligands is illustrated in Figure 3, where the reference lines indicate cut-off values of both tracers. For $[^{11}\text{C}]$PIB, all AD patients showed high global $\text{BP}_{\text{ND}}$, whilst only one control subject showed binding above the cut-off value. For $[^{18}\text{F}]$FDDNP, 3 AD patients showed global binding below and 4 controls global binding above the cut-off value. Agreement between quantitative analysis and visual rating was excellent for $[^{11}\text{C}]$PIB ($\kappa=0.85$) and fair for $[^{18}\text{F}]$FDDNP ($\kappa=0.40$).

**Young versus old**

In an additional analysis, the sample was divided into younger (<65 years) and older (≥65 years) subjects. The younger group consisted of 13 AD patients (mean±SD 59±3 years, 5 female (F)) and 8 controls (60±3 years, 4 F), whilst the older group included 8 AD patients (69±3 years, 2 F) and 12 controls (71±4 years, 3 F). In both groups, AD patients had lower average MMSE scores, higher MTA scores and higher global cortical $[^{11}\text{C}]$PIB and $[^{18}\text{F}]$FDDNP $\text{BP}_{\text{ND}}$ values than controls. For both younger and older subjects, sensitivity was comparable to those of the total sample except for sensitivity of $[^{18}\text{F}]$FDDNP, which was lower in older subjects (Table 2). For all imaging modalities, specificity and accuracy were lower in older subjects.

For visual assessment of $[^{11}\text{C}]$PIB, age did not affect agreement with quantitative values, as it was excellent in both younger ($\kappa=0.90$) and older ($\kappa=0.85$) subjects. For $[^{18}\text{F}]$FDDNP, however, substantial agreement with quantitative values was found for younger subjects, but poor agreement for older subjects ($\kappa=0.68$ and $\kappa=0.18$, respectively).

**Figure 3**: Agreement between quantitative assessment of global $[^{11}\text{C}]$PIB and $[^{18}\text{F}]$FDDNP $\text{BP}_{\text{ND}}$ with $[^{18}\text{F}]$FDDNP binding on the x-axis and $[^{11}\text{C}]$PIB binding on the y-axis. Reference lines represent cut-off values. The right upper quadrant contains subjects with relatively high binding for both ligands, whilst the left lower quadrant displays subjects with relatively low binding for both ligands. AD patients are represented by circles and controls by squares.
Chapter 7

**Discrepancies between modalities**

Examples of contrasting ratings between modalities are illustrated in Figure 4. Visual interpretation of $[^{11}C]PIB$ classified three older control subjects as AD. Concurrent ratings of $[^{18}F]FDDNP$, $[^{18}F]FDG$ and MTA were abnormal for one subject, whilst ratings of scans of the other two varied. Moreover, agreement between visual ratings of $[^{11}C]PIB$ and $[^{18}F]FDDNP$ and dichotomised status using cut-off values for global binding varied. For subject A, $[^{11}C]PIB$ and $[^{18}F]FDDNP$ were visually rated as abnormal and global $BP_{ND}$ of both ligands was above cut-off ($[^{11}C]PIB$ $BP_{ND}$=0.91, $[^{18}F]FDDNP$ $BP_{ND}$=0.10). For subject B, $[^{11}C]PIB$ was visually rated as AD, but global $[^{11}C]PIB$ $BP_{ND}$ (0.47) was slightly below cut-off. $[^{18}F]FDDNP$ was rated as AD, both visually and quantitatively ($BP_{ND}$ 0.08). For subject C all imaging modalities were visually rated as abnormal. The same was true for global $[^{18}F]FDDNP$ $BP_{ND}$ (0.09). In contrast, global $[^{11}C]PIB$ $BP_{ND}$ was below the cut-off value (0.26).

**Figure 4:** $[^{11}C]PIB$, $[^{18}F]FDDNP$, $[^{18}F]FDG$ and MRI of three controls visually rated as having increased $[^{11}C]PIB$ uptake. Column 1: 73 year old male with $[^{11}C]PIB$ and $[^{18}F]FDDNP$ visually rated as abnormal. $[^{18}F]FDG$ and MTA were considered normal (MTA scores of right and left hippocampus = 1). Column 2: 74 year old female with $[^{11}C]PIB$ and $[^{18}F]FDG$ visually rated as abnormal. $[^{18}F]FDDNP$ and MTA were considered normal (MTA scores of right hippocampus = 1 and left hippocampus = 0). Column 3: 80 year old male with all imaging modalities visually rated as abnormal (MTA scores of right and left hippocampus = 2).
Discussion

This study evaluated visual assessment of [11C]PIB and [18F]FDDNP images as potential supportive diagnostic markers for AD. Visual interpretation of [11C]PIB images showed a high diagnostic accuracy combined with high inter-observer agreement. Furthermore, agreement between visual and quantitative assessment of [11C]PIB was high. Moreover, in the present cohort, visual rating of [11C]PIB for identification of AD performed equally well as the combination of [18F]FDG (high sensitivity) and MTA (high specificity). Diagnostic accuracy of [11C]PIB was even higher than that of other markers. Visual rating of [18F]FDDNP images for identification of AD had the lowest sensitivity, specificity and accuracy. Additionally, agreement with quantitative assessment was only fair.

The low accuracy of visual interpretation of [18F]FDDNP may be due to the fact that the difference in [18F]FDDNP BPND between AD patients and controls is subtle (14-16). Quantitative assessment of [18F]FDDNP scans was considerably better in identifying AD and could therefore still be a potential tool in diagnosing dementia. This is especially true as [18F]FDDNP is the only PET ligand available to date, which also has affinity for tangles (25). As such, [18F]FDDNP binding may provide an in vivo measure of tau. This is supported by the correlation with tau levels in cerebrospinal fluid (26). However, at present [18F]FDDNP does not discriminate AD from controls with equal accuracy as CSF levels of tau (27), potentially due to the high level of nonspecific binding. Increasing the specific to nonspecific binding ratio by using ligands that are chemically and structurally related to [18F]FDDNP may greatly enhance clinical impact.

Age may be an important factor in the accuracy of imaging biomarkers for diagnosing AD. In this study, a cut-off of 65 years was chosen, as this is generally regarded as the age for separating early from late onset of AD (24). Although underlying AD pathology does not differ between these groups, increasing age is thought to lead to, amongst others, reduction of cerebral glucose metabolism, reduction of hippocampal volume and increased deposition of neurofibrillary tangles (10-12,28). This could have an effect on the performance of diagnostic markers. However, for both young and old subjects, the same modalities performed best with respect to sensitivity and specificity. Diagnostic accuracy did differ slightly. For younger subjects [18F]FDG performed better than MTA, whilst for older subjects this was the other way round due to a decrease in accuracy of [18F]FDG.

For [18F]FDDNP, an age effect did seem to have complicated interpretation of scans. In older subjects only five out of ten AD patients were visually identified as AD and agreement with quantitative values of [18F]FDDNP was poor for older subjects. This was in contrast to the substantial agreement found for younger subjects. A possible explanation is an age related increase in binding in regions not used for calculation of ‘global binding’ (e.g. occipital cortex, thalamus, striatum, white matter), which might have affected visual readings. This suggests that, with more practice, agreement between visual assessment and quantitative values in older subjects may improve.

Visual interpretation of [11C]PIB images in older subjects rated three out of twelve cognitively
healthy elderly as abnormal. This is in line with post mortem studies reporting presence of AD pathology in around 30% of cognitively healthy elderly subjects (28) and with other studies reporting increased [11C]PIB uptake in cognitively normal elderly (29-33). These subjects may be preclinical AD patients, as deposition of pathology is thought to start a decade before cognitive impairment arises (34). Alternatively, deposition of amyloid could have a more benign character, not leading to clinical AD. To verify whether these subjects are indeed preclinical cases will require longitudinal follow up studies. Assuming that these subjects are indeed preclinical rather than false positive cases, specificity for the other imaging modalities would likely also have been affected, resulting in falsely low specificity estimates (35).

A limitation of the present study is that readings of imaging modalities were dichotomised. Especially [18F]FDG provides a graded answer and forcing it into just two alternatives does not lead to use of it's full potential. Therefore the diagnostic accuracy of [18F]FDG in the present study might be lower that the diagnostic value in clinical practice. Moreover, in order to assess agreement of visual ratings with quantitative values of [11C]PIB and [18F]FDDNP, global tracer binding was also dichotomised using cut-off values. For [11C]PIB, this cut-off value was in line with previously determined values for global binding (32,36). Deposition of AD pathology in the brain is thought to be a continuum and increases can be subtle and/or isolated to certain regions before widespread deposition (31,37). Therefore, using a global cut-off value could underestimate the number of subjects with increased binding. However, as the purpose of the study was to compare visual interpretation of both ligands (based on the entire image) with corresponding quantitative values, a global estimate was considered to be the most appropriate method. The present findings are in line with a previously published study (17) evaluating visual interpretation of [11C]PIB and [18F]FDG, reporting similar agreement between readers for both modalities and similar sensitivity and specificity for [11C]PIB. Although, in the present study, sensitivity in the older subjects was higher for [18F]FDG, average age of the older subjects was also lower.

These findings contribute to the discussion on new supportive biomarkers for diagnosing AD (2). High diagnostic accuracy of visual interpretation of [11C]PIB images in the present study suggests that it has potential in both refuting and confirming the diagnosis of AD. Additional use of quantitative [18F]FDDNP assessment seems limited in differentiating AD patients from controls, but it could still have value in the differential diagnosis of rare forms of dementia (38), as [18F]FDDNP binding might be an in vivo marker of tau. Increased [11C]PIB load in the brain has also been reported in other neurodegenerative diseases (39,41) and cognitively healthy elderly (31). Therefore, in clinical practice a positive [11C]PIB scan should be combined with other supportive markers for AD.


