Performance Characteristics of Amyloid PET with Florbetapir F 18 in Patients with Alzheimer’s Disease and Cognitively Normal Subjects

Abhinay D. Joshi 1, Michael J. Pontecorvo 1, Christopher M. Clark 1,2, Alan P. Carpenter 1, Danna L. Jennings 3, Carl H. Sadowsky 4, Lee P. Adler 5, Karel D. Kovnat 5, John P. Seibyl 1,6,7, Anupa Arora 1, Krishnendu Saha 1, Jason D. Burns 1, Mark J. Lowrey 1, Mark A. Mintum 1, Daniel M. Skovronsky 1,2, and the Florbetapir F 18 Study Investigators

1Avid Radiopharmaceuticals, Inc., Philadelphia, Pennsylvania; 2Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania; 3Molecular NeuroImaging, LLC, New Haven, Connecticut; 4Nova SE University, Ft. Lauderdale, Florida; 5Adler Institute for Advanced Imaging, Jenkintown, Pennsylvania; 6Institute for Neurodegenerative Disorders, New Haven, Connecticut; and 7Yale University School of Medicine, New Haven, Connecticut

The objectives of this study were to examine the effective dose range and the test–retest reliability of florbetapir F 18 using, first, visual assessment by independent raters masked to clinical information and, second, semiautomated quantitative measures of cortical target area to cerebellum standardized uptake value ratios (SUVR) as primary outcome measures. Visual ratings of PET image quality and tracer retention or SUVr. The difference in the visual quality across the 111-MBq (3-mCi) and 370-MBq (10-mCi) dose in the separate set of subjects (AD, n = 10; YHCs, n = 10), test–retest reliability was evaluated by comparing intrasubject visual read ratings and SUVRs for 2 PET images acquired within 4 wk of each other. Results: There were no meaningful differences between the 111-MBq (3-mCi) and 370-MBq (10-mCi) dose in the visual rating or SUVR. The difference in the visual quality across 111 and 370 MBq showed a trend toward lower image quality, but no statistical significance was achieved (t-test; t1 = −1.617, P = 0.12) in this relatively small sample of subjects. At both dose levels, visual ratings of amyloid burden identified 100% of AD subjects as Aβ-positive and 100% of YHCs as Aβ-negative. Mean intrasubject test–retest variability for cortical average SUVRs with the cerebellum as a reference over the 50- to 70-min period was 2.4% ± 1.41% for AD subjects and 1.5% ± 0.84% for controls. The overall SUVR test–retest correlation coefficient was 0.99. The overall k-statistic for test–retest agreement for Aβ classification of the masked reads was 0.89 (95% confidence interval, 0.69–1.0). Conclusion: Florbetapir F 18 appears to have a wide effective dose range and a high test–retest reliability for both quantitative (SUVR) values and visual assessment of the ligand. These imaging performance properties provide important technical information on the use of florbetapir F 18 and PET to detect cerebral amyloid aggregates.

Key Words: Aβ; florbetapir; PET; biomarkers; test–retest; dose ranging

DOI: 10.2967/jnumed.111.090340

Despite the introduction of standardized criteria for the clinical diagnosis of Alzheimer’s disease (AD) (1,2), a degree of uncertainty nevertheless commonly persists (3). The consequence of this uncertainty is reflected in the misdiagnosis rate documented in postmortem studies (4–8). With the anticipated development of pathologically targeted treatments and the goal of treating during the earliest symptomatic stage, there is a growing consensus regarding the advantage of incorporating valid biomarkers of AD pathology in the assessment of individuals with late-life cognitive impairment. This is reflected in the suggested revision of AD clinical diagnostic criteria revisions proposed by the International Working Group for New Research Criteria for the Diagnosis of AD, the National Institute on Aging–Alzheimer’s Association, and the American Psychiatric Association (3,9–11). Although there has been longstanding work on the development and evaluation of biomarkers that identify the consequences of AD pathology, such as MRI-defined regional brain atrophy (12) and cerebral spinal fluid measures of β-amyloid and τ-proteins, the ability to directly measure a disease-defining pathology using molecular imaging has only recently become possible. The initial studies using 11C-Pittsburgh compound B were the first to demonstrate the feasibility of this approach (13). However, the 20-min half-life of 11C limits the use of this molecular imaging ligand to specialized research centers. The ability to incorporate amyloid PET in the routine practice of medicine has led to the search for 18F ligands with performance characteristics appropriate for use in community health-care facilities (14–17).
Florbetapir F 18 (\(^{18}\)F-AV-45) is an \(^{18}\)F amyloid PET ligand with rapid brain uptake and rapid washout from gray tissues not containing amyloid, high affinity to aggregated \(\beta\)-amyloid, a short imaging time, good separation between the amyloid retention and background signal, and a long, stable pseudoequilibrium permitting flexibility in imaging acquisition time (18). The current study was designed to examine the following 2 key imaging performance properties: appropriate radiotracer dose for optimal quality imaging and test–retest reliability of amyloid PET with florbetapir F 18.

**MATERIALS AND METHODS**

Data were collected in 2 separate protocols, 1 to evaluate the dose-range flexibility and 1 to assess test–retest variability. In both protocols, subjects were recruited from the community by the principal investigators at 6 participating sites. Protocols were approved by the institutional review board, and written informed consent was obtained for each subject enrolled. AD participants felt to be too cognitively impaired to provide truly informed consent indicated their willingness to participate with informed consent provided by their designated decision maker, usually their spouse or an adult child. Volunteers qualifying for the AD arm were 50 y or older, met criteria of the National Institute of Neurological and Communicative Diseases and Stroke and Alzheimer’s Disease and Related Disorders Association for probable AD, and had a Mini Mental State Examination score between 10 and 24 at the time of enrollment. In addition, each AD patient had to have a knowledgeable coparticipant (usually a spouse or adult child) who was willing and able to provide information about the evolution of their cognitive impairment and decline in functional abilities.

Control subjects were neurologically healthy. As amyloid deposition has been described in cognitively normal subjects over 55 y of age (19,20), control subjects in this study had to be between the ages of 35 and 55 y (inclusive) and free of cognitive impairment by history and examination.

**Clinical Assessment**

At the time of enrollment, the subject’s diagnostic category was confirmed using routine clinical methods. In addition, each participant completed a standard battery of cognitive tests and functional assessment measures including a Mini Mental State Examination, the cognitive component of the Alzheimer’s Disease Assessment Scale, an immediate and delayed Wechsler Logical Memory assessment, and a category (animals) verbal fluency measure. Safety assessments included vital signs, an electrocardiogram, and standard blood and urine tests. Subjects who showed evidence of any non-AD neurodegenerative or psychiatric disorder on clinical examination or MRI (6 mo previous) or clinically significant medical comorbidity that might pose a safety risk or interfere with interpretation of the scan were excluded from the study.

**Florbetapir F 18 PET Acquisition**

Subjects participating in the low versus standard dose comparison study were randomized to receive an intravenous injection of either 111 MBq (3 mCi) or 370 MBq (10 mCi) of florbetapir F 18 in a volume of 10 mL or less. Of the 20 subjects enrolled in this study, 9 (5 with AD and 4 control subjects) received a single injection of 111 MBq (3 mCi), and 11 (4 with AD and 7 control subjects) received 370 MBq (10 mCi) of florbetapir F 18. Imaging began immediately after administration of florbetapir F 18 and with a dynamic series of images captured for 90 min (frame sequence, 2 \(\times\) 2.5, 1 \(\times\) 5, 4 \(\times\) 10, 4 \(\times\) 5, and 2 \(\times\) 10 min). The imaging data were reconstructed using iterative reconstruction (4 iterations, 16 subsets), with a 128 \(\times\) 128 image size, 2.67 \(\times\) 2.67 mm pixel size, 2-mm slice thickness, and 5-mm postreconstruction gaussian filter.

Twenty participants, 10 AD patients and 10 controls, were enrolled in the test–retest study. Each participant was imaged on 2 separate days not more than 4 wk apart and received a single intravenous dose of 370 MBq (10 mCi) of florbetapir F 18, followed 50 min later by a 20-min PET session (50–70 min after injection). At 2 of the 3 study centers, images were acquired in 5-min frames, allowing comparison of test–retest reproducibility for a 10-min (50–60 min after injection) and a 20-min image acquisition. At the third center, only a single 20-min scan was obtained. All the imaging data were reconstructed using iterative reconstruction (4 iterations, 16 subsets), with a 128 \(\times\) 128 image size, 2.67 \(\times\) 2.67 mm pixel size, 2-mm slice thickness, and 5-mm postreconstruction gaussian filter.

**Image Analysis**

PET scans from each study were visually evaluated by 3 board-certified nuclear medicine physicians masked to the subject’s clinical and demographic information (4 readers were used, with 2 overlapping each study). In the dose-comparison study, images from each subject were summed over the 50- to 55-min and 55- to 60-min time frames to get an average 10-min (50–60 min after injection) image. IMViewer, an image display package from the IMVista Corp., was used for reading images. For each image, the masked reader rated the evaluable on the basis of patient positioning, presence of the entire brain, noise, and resolution. Evaluative images were then rated for image quality on a 1–5 scale (5 was excellent and 1 was poor) on the basis of the pixilation, noise, and resolution in all 3 orientations (transverse, coronal, and sagittal). The mean quality rating was calculated for each case from the quality assessment of 3 readers, and these mean values were used for between-dose statistical analysis. To characterize overall tracer accumulation, the reader performed intensity thresholding of the image such that white matter in the pons or cerebellum began to become prominent. Relative distribution of activity in white versus gray matter was examined in the transverse orientation at the level of thalamus. Younger healthy controls (YHCs) showed an x-shaped white-matter pattern, and subjects with AD showed a loss of this gray–white distinction, with increased uptake in gray matter. In coronal and sagittal positions at the level of precuneus, the \(\beta\)-amyloid–positive images showed accumulation along the midline and surface of the cortex higher than cerebellum gray, and \(\beta\)-amyloid–negative images showed accumulation consistent with cerebellum gray. In the test–retest study, images were given a binary rating based on their assessment of the overall degree of florbetapir F 18 retention in the cortical gray matter. In addition, all images were normalized to a standard template in Talairach space using statistical parametric mapping (SPM) software (21), and cortical-to-cerebellum standardized uptake value ratios (SUVRs) were calculated for 6 target cortical regions of interest (frontal, temporal, parietal, anterior cingulate, posterior cingulate, and precuneus), using the whole cerebellum as the reference region. The unweighted average cortical SUVRs was calculated by combining the individual values for each of the cortical regions (22,23).
For the dose-comparison study, Student t tests were used to compare the mean SUVr at a representative imaging time (50–60 min) across the 2 dose levels, separately for AD patients and YHCs. A Student t test was also used to compare the image quality rating between the 2 doses. In the test–retest study, an intraclass correlation coefficient was calculated to evaluate the relationship between test and retest SUVs. The mean intrasubject test–retest variability (%) across the control and AD groups was also calculated using cortical average SUVr (relative to cerebellum) as the absolute value: [(retest − test)/test], expressed as a percentage. The intraclass κ-statistic (24) was used to assess the agreement of the 2 reading results (test and retest) for each reader.

For each study, consistency of rating between readers was summarized with the Fleiss κ-statistic (25).

RESULTS

Table 1 provides the demographic and clinical features of each of the cohorts participating in these 2 studies. The only notable finding is a larger proportion of women in the AD group of the test–retest study than in the dose-comparison study.

Comparison of 111 MBq (3 mCi) to 370 MBq (10 mCi) of Florbetapir F 18

In this study, all 20 subjects had valid PET scan data and were included in the evaluable population. However, 1 AD subject (322-003) in the 370-MBq group did not complete all of the planned imaging times because the PET scan was interrupted as a result of a power failure. All available data for this subject were included in the analysis of the imaging data.

At both 111- and 370-MBq dose levels, SUVs and SUVrs for target cortical brain regions were greater in AD subjects than in control subjects. The difference between AD patients and control subjects was similar for both dose levels. For both dose levels, the time–activity curves show a clear separation between the activity in cortical target areas and the cerebellum in AD subjects but not in the control subjects, beginning around 15 min after injection. Washout approached asymptote by 30 min, and there was little change in SUVs in either AD subjects or controls between 30 and 90 min after injection (Fig. 1). In AD subjects, mean cortical average SUVs at the representative imaging time (50–60 min after injection) for the 111- and the 370-MBq doses were 1.78 ± 0.22 and 1.66 ± 0.29, respectively. In control subjects, mean cortical average SUVrs for the 111- and 370-MBq doses were 0.97 ± 0.05 and 0.99 ± 0.09, respectively. These SUVr results for the 111- and 370-MBq dose groups were not significantly different (groupwise t test; \( t_{15} = 1.14, P = 0.27 \)).

Although all images (50–60 min) were rated as evaluable on visual assessment, some images from the 111-MBq group were relatively count-starved, resulting in modestly reduced image quality (Fig. 2). Image quality ratings (mean of 3 readers for each case) were slightly lower for the 111- than for the 370-MBq group (overall mean, 3.44 ± 0.65 and 4.00 ± 0.63 for 111- and 370-MBq groups, respectively). However, this trend was not sufficient to reach statistical significance (\( t_1 = -1.62, P = 0.12 \)) and did not appear to affect the ability to read the scans. All AD subjects, regardless of the dose administered, were rated as β-amyloid–positive, and all controls were rated as β-amyloid–negative, using a binary (forced-choice) score (Fleiss \( \kappa = 1.00 [95\% confidence interval, 0.74–1.00] \)).

Test–Retest Reliability of Florbetapir F 18

For this study, all 20 subjects underwent a valid PET scan and were included in the evaluable population; thus, all the imaging data were analyzed.

Results were highly repeatable between test and retest days for AD patients and control subjects. SUVrs for the cortical average relative to cerebellum of test and retest 20-min scans were 1.42 ± 0.25 and 1.41 ± 0.27 for AD subjects and 1.00 ± 0.06 and 1.01 ± 0.06 for control subjects. These are mean SUVrs across all AD subjects for test (1.42) and retest (1.41). Similarly for controls, mean test SUVr was 1.00 and retest was 1.01. Similar results were obtained for individual cortical regions (Table 2). The intraclass correlation coefficients for cortical average SUVr were 0.99 for AD subjects and 0.96 for control subjects and in excess of 0.99 (Fig. 2) for the 2 groups combined. For those subjects who had individual frames to allow reformatting the PET images, the intraclass correlation coefficients for AD subjects and for control subjects were comparable for the 50- to 60-min and the 50- to 70-min periods.

Intrasubject test–retest variability for cortical average SUVr (cerebellum) over the 50- to 70-min period was 2.40% ± 1.41% for AD subjects and 1.50% ± 0.84% for

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics: Both Studies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Demographic</th>
<th>111 MBq (3 mCi)</th>
<th>370 MBq (10 mCi)</th>
<th>Test–retest</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD (n = 5)</td>
<td>Control (n = 4)</td>
<td>AD (n = 4)</td>
<td>Control (n = 7)</td>
</tr>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>74.8</td>
<td>48.0</td>
<td>76.8</td>
</tr>
<tr>
<td>Range</td>
<td>58–88</td>
<td>38–52</td>
<td>65–90</td>
</tr>
<tr>
<td><strong>Percentage women</strong></td>
<td>40</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td><strong>Mini Mental State Examination</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>18.4</td>
<td>29.8</td>
<td>12.5</td>
</tr>
</tbody>
</table>
controls. Similar results were observed for the 50- to 60-min period. In a subset of control subjects with available data at both time periods, the test–retest variability for cortical average SUVr was 1.50% ± 0.84%, 1.43% ± 0.92%, respectively, for the 50- to 70-min and 50- to 60-min periods. In the subset of AD subjects with available data at both time periods, the variability was 1.54% ± 0.81% and 1.93% ± 1.20%, respectively, for the 50- to 70-min and 50- to 60-min periods. Thus, there was no additional gain in reliability and repeatability with longer imaging times.

There was a high level of agreement between the visual read classification (β-amyloid–positive or β-amyloid–negative) for the test and retest images. Each reader assigned the same classification to the test and retest images on 19 of 20 image pairs (95%), resulting in a mean intrareader κ-value of 0.88 ± 0.015. Agreement between readers across the 40 images (test and retest combined) was also high, with a Fleiss κ of 0.76 (95% confidence interval, 0.58–0.94). Importantly, all cases of disagreement occurred for scans from clinically diagnosed AD subjects. Images from YHCs were read as amyloid-negative (β-amyloid–negative) in 100% of cases.

Performance of Combined YHC Group

The mean SUVr for YHCs (age, <55 y; n = 21) was evaluated by combining the retest data from the test–retest study and the 111- and 370-MBq data from the dose-comparison study. Across all subjects, the mean SUVr was 0.987 ± 0.066. A similar value, 0.982 ± 0.073, was obtained using only those subjects younger than 50 y (n = 13). By either method, the confidence limit for the upper 5% of the distribution (mean ± 1.65 SD) was an SUVr of approximately 1.10.

DISCUSSION

Florbetapir F 18 is one of a new generation of 18F-labeled PET ligands with the potential to assess the presence and density of aggregated β-amyloid in a living patient (14–17).
We report here 2 studies that bear on the performance characteristics of florbetapir F 18 PET.

The primary objectives of the dose-ranging study were to compare the performance characteristics of 111 MBq (3 mCi) and 370 MBq (10 mCi) of florbetapir F 18 in patients with AD and YHC subjects. The qualitative and quantitative results for the 111-MBq (3 mCi) and 370-MBq (10 mCi) dose levels were similar. At both dose levels, increased tracer retention could be seen in the cortex relative to the cerebellum as early as 30 min after injection, SUVrs approached asymptote 30 min after injection, and there was little difference in the values from 50 to 90 min after injection. The visual quality of PET images was rated slightly (no significant difference) better for the 370-MBq dose group than for the 111-MBq dose group. Nonetheless, this difference in image quality did not affect the ability of the masked reader to identify high and low β-amyloid ligand retention. On the basis of all data, however, the 370-MBq dose would provide more robust imaging than the 111-MBq dose. Because radiation exposure at the 370-MBq level is within an acceptable range, this dose has become the recommended standard. However, these results indicate that lower dose levels might be effective in some situations.

The primary objective of a test–retest study was to evaluate reproducibility of florbetapir F 18 when used in younger healthy subjects and patients with AD. Both semiautomated image analysis and the visual read results were highly reliable. The percentage difference between test and retest images was less than 3% for the cortical average SUVr tested at both 50- to 60-min and 50- to 70-min acquisitions. There was a very high intraclass correlation (0.99) for the test–retest results at both the 50- to 60-min and 50- to 70-min postinjection times. Thus, a 10-min florbetapir F 18 scan 50 min after injection appears sufficient for evaluating the degree of amyloid pathology.

Reliability for several other amyloid PET ligands has previously been reported. 11C-Pittsburgh compound B provides good test–retest reproducibility with an intraclass correlation of 0.96 (26). Recently, a study with another 18F compound for amyloid imaging, 18F-flutemetamol, showed high test–retest replicability, with an intraclass correlation of 0.96 in a group of patients and controls (27). Our study documents that florbetapir F 18 also provides high test–retest results, similar to those of 11C-Pittsburgh compound B and 18F-flutemetamol.

Binary (β-amyloid–positive or β-amyloid–negative) visual interpretation of the florbetapir F 18 PET images also proved highly reliable both across readers (Fleiss κ = 1.0 in the dose-comparison study and 0.75 in the test–retest study) and within patients for test–retest images (intrareader κ > 0.86 for all 3 readers). These κ-values are all within the range that might be considered substantial–to–almost-perfect agreement (28).

The confidence limit for the upper 5% of the SUVr distribution based on the YHC group from both studies was estimated to be approximately 1.10. This value is therefore proposed as a threshold for quantitative classification of β-amyloid–positive and β-amyloid–negative cases. The usefulness of this threshold was confirmed in a subsequent imaging-to-histopathology correlation study (29).

### TABLE 2

<table>
<thead>
<tr>
<th>Region</th>
<th>AD patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Retest</td>
</tr>
<tr>
<td>Average of the following 6 regions*</td>
<td>1.42 ± 0.24</td>
<td>1.41 ± 0.27</td>
</tr>
<tr>
<td>Frontal</td>
<td>1.24 ± 0.25</td>
<td>1.24 ± 0.28</td>
</tr>
<tr>
<td>Anterior cingulate</td>
<td>1.45 ± 0.31</td>
<td>1.44 ± 0.32</td>
</tr>
<tr>
<td>Temporal</td>
<td>1.40 ± 0.22</td>
<td>1.40 ± 0.26</td>
</tr>
<tr>
<td>Parietal</td>
<td>1.26 ± 0.23</td>
<td>1.25 ± 0.26</td>
</tr>
<tr>
<td>Precuneus</td>
<td>1.63 ± 0.34</td>
<td>1.62 ± 0.35</td>
</tr>
<tr>
<td>Posterior cingulate</td>
<td>1.51 ± 0.28</td>
<td>1.50 ± 0.28</td>
</tr>
</tbody>
</table>

*Average SUVr values for test–retest in AD patients and healthy controls are similar. Data are mean ± SD.

FIGURE 3. Correlation of cortical average-to-cerebellum SUVr test and retest for all AD patients and healthy controls (HC).
One limitation of the dose-comparison study is that it was performed in 2 different cohorts of subjects, rather than as a within-subject study. One risk in such a small between-subject design is that interpretation can be confounded by a few anomalous subjects (e.g., an amyloid-positive healthy control or an amyloid-negative subject clinically diagnosed as AD). This risk was mitigated in part by the use of YHCs rather than older controls, and in fact both doses appeared effective in separating AD from YHC by both visual read and quantitative (SUVr) analysis.

A second limitation of both studies is that the use of YHCs rather than older controls or subjects with mild cognitive impairment may have reduced sensitivity for detecting some differences between doses and variation in test–retest performance. In view of accumulating evidence (30,31) clearly showing that amyloid deposition takes place years before the onset of dementia, subjects with mild cognitive impairment and clinically healthy elderly can have significant amyloid levels both on PET scans and at autopsy. Recent studies suggest that 5%–10% of older healthy control or subjects with mild cognitive impairment have borderline positive levels of amyloid (27,32), and it is theoretically possible that these subjects would be more difficult to detect with a low dose of tracer or might produce more variable SUVs or visual interpretation than the YHCs used in this study.

CONCLUSION

Florbetapir F 18 showed a wide effective dose range and high test–retest reliability for both visual assessment of ligand retention and SUVr values. Both the dose-ranging study and the test–retest study indicate that florbetapir F 18 is a robust imaging tool well suited for use as a biomarker for AD pathology in standard community PET facilities.

DISCLOSURE STATEMENT

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

ACKNOWLEDGMENTS

Special thanks are given to Nathaniel Lim, Tyler Benedum, Eric Chellquist, Wei Zhang, Cassie Bowen, Caitlin Pearson, Raul Jackson, Andrew Kielt, Jim Dolan, and Paul Kennedy for their support in the manufacturing of the florbetapir F 18 injection. This study was funded by Avid Radiopharmaceuticals Inc., Philadelphia, Pennsylvania. Abhinay D. Joshi, Michael J. Pontecorvo, Christopher M. Clark, Alan P. Carpenter, Mark A. Mintun, Daniel M. Skovronsky, Mark J. Lowrey, Jason D. Burns, and Anupa Arora are Avid employees. Lee P. Adler, Danina L. Jennings, and Carl H. Sadowsky have no relevant financial disclosures other than the clinical study contract with Avid for this work. John P. Seibyl has no relevant disclosures. Krishnendu Saha is a former employee of Avid. Karel D. Kovnat has no relevant disclosures. No other potential conflict of interest relevant to this article was reported.

REFERENCES


