Impact of Training Method on the Robustness of the Visual Assessment of ^{18}F-Florbetaben PET Scans: Results from a Phase-3 Study

John Seibyl[1], Ana M. Catafau[2], Henryk Barthel[3], Kenji Ishii[4], Christopher C. Rowe[5], James B. Leverenz[6], Bernardino Ghetti[7], James W. Ironside[8], Masaki Takao[9,10], Hiroyasu Akatsu[11,12], Shigeo Murayama[13,14], Santiago Bullich[2], Andre Mueller[2], Norman Koglin[2], Walter J. Schulz-Schaeffer[15], Anja Hoffmann[16], Marwan N. Sabbagh[17], Andrew W. Stephens[2], and Osama Sabri[3]

[1]Molecular Neuroimaging LLC, New Haven, Connecticut; [2]Piramal Imaging GmbH, Berlin, Germany; [3]Leipzig University, Leipzig, Germany; [4]Department of Positron Medical Center, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan; [5]Austin Health, University of Melbourne, Melbourne, Victoria, Australia; [6]Virginia-Puget Sound Health Care System and University of Washington, Seattle, Washington; [7]Indiana University School of Medicine, Indianapolis, Indiana; [8]University of Edinburgh, Edinburgh, Scotland; [9]Mihara Memorial Hospital, Isesaki, Japan; [10]Department of Neurology, Saitama International Medical Center, Saitama Medical University, Saitama, Japan; [11]Fukushimura Hospital, Toyohashi, Japan; [12]Departments of Community-Based Medicine and Neurology, Nagoya City University Graduate School of Medical Sciences, Nagoya City, Aichi, Japan; [13]Department of Neurology and Neuropathology (the Brain Bank for Aging Research), Tokyo Metropolitan Geriatric Hospital, Tokyo, Japan; [14]Institute of Gerontology, Tokyo, Japan; [15]Georg-August University Göttingen, Göttingen, Germany; [16]Bayer Pharma AG, Berlin, Germany; and [17]Alzheimer's and Memory Disorders Division, Department of Neurology, Barrow Neurological Institute, Phoenix, Arizona

Training for accurate image interpretation is essential for the clinical use of β-amyloid PET imaging, but the role of interpreter training and the accuracy of the algorithm for routine visual assessment of florbetaben PET scans are unclear. The aim of this study was to test the robustness of the visual assessment method for florbetaben scans, comparing efficacy readouts across different interpreters and training methods and against a histopathology standard of truth (SoT). **Methods:** Analysis was based on data from an international open-label, nonrandomized, multicenter phase-3 study in patients with or without dementia (ClinicalTrials.gov: NCT01020838). Florbetaben scans were assessed visually and quantitatively, and results were compared with amyloid plaque scores. For visual assessment, either in-person training ($n = 3$ expert interpreters) or an electronic training method ($n = 5$ naïve interpreters) was used. Brain samples from participants who died during the study were used to determine the histopathologic SoT using Bielschowsky silver staining (BSS) and immunohistochemistry for β-amyloid plaques. **Results:** Data were available from 82 patients who died and underwent postmortem histopathology. When visual assessment results were compared with BSS + immunohistochemistry as SoT, median sensitivity was 98.2% for the in-person–trained interpreters and 96.4% for the e-trained interpreters, and median specificity was 92.3% and 88.5%, respectively. Median accuracy was 95.1% and 91.5%, respectively. On the basis of BSS only as the SoT, median sensitivity was 98.1% and 96.2%, respectively; median specificity was 80.0% and 76.7%, respectively; and median accuracy was 91.5% and 86.6%, respectively. Interinterpreter agreement (Fleiss κ) was excellent (0.89) for in-person–trained interpreters and very good (0.71) for e-trained interpreters. Median intrainterpreter agreement was 0.9 for both in-person–trained and e-trained interpreters. Visual and quantitative assessments were concordant in 88.9% of scans for in-person–trained interpreters and in 87.7% of scans for e-trained interpreters. **Conclusion:** Visual assessment of florbetaben images was robust in challenging scans from elderly end-of-life individuals. Sensitivity, specificity, and interinterpreter agreement were high, independent of expertise and training method. Visual assessment was accurate and reliable for detection of plaques using BSS and immunohistochemistry and well correlated with quantitative assessments.

Key Words: Alzheimer's disease; florbetaben; positron-emission tomography

Florbetaben is an ^{18}F-labeled β-amyloid tracer developed for PET. In 2014, florbetaben was approved by the U.S. Food and Drug Administration, the European Medicines Agency, and the Korean Ministry of Food and Drug Safety to detect or exclude the presence of neuritic β-amyloid plaques in the brain (1,2). In a phase-3 study, high sensitivity and specificity were demonstrated for florbetaben in the detection of β-amyloid aggregates comparing PET with postmortem histopathology (3).

The current clinical method for interpretation of florbetaben β-amyloid PET scans is visual assessment (4). Training for accurate image interpretation is a key issue for β-amyloid PET imaging—all interpreters require training and several different training tools are available. The type of interpreter training and the accuracy for visual assessment of florbetaben PET scans are unclear. To test the adequacy of the training comparing visual assessment against β-amyloid histopathology, scans from end-of-life patients are required. It is known that pronounced atrophy and other brain abnormalities can compromise image interpretation in end-of-life individuals. Indeed, severe structural brain abnormalities were present

in the participants in the histopathology cohort, leading to challenging scans. If acceptable images can be obtained under these conditions, improved interpretations can be expected in the clinical setting with less technically challenging scans. Moreover, a technically difficult dataset would be likely to expose differences between training approaches (electronic training tool vs. in-person trainer) that might inform best training practices.

The aim of the present study was to test the robustness of the visual assessment method for florbetaben scans, comparing efficacy readouts (sensitivity, specificity, and κ-values) across different interpreters and training methods, against a histopathology standard of truth (SoT) in end-of-life patients. The study also compared visual and quantitative assessments of florbetaben PET, because quantification is commonly used in research and may be implemented in future clinical routine.

MATERIALS AND METHODS

Study Design and Population

This analysis was based on an international open-label, nonrandomized, multicenter phase-3 study (ClinicalTrials.gov: NCT01020838). The study was conducted in accordance with the Declaration of Helsinki, and approvals from regulatory authorities and ethics committees were obtained.

Participants were recruited from 15 centers (including dementia clinics with brain-bank experience, hospices, private practices, and dementia self-help groups) in Australia, Europe, Asia, and North America and examined between February 2010 and August 2013. Eligible subjects were nondemented individuals ($n = 9$) and patients with Alzheimer disease (AD; $n = 60$), dementia with Lewy bodies ($n = 4$), or other dementias ($n = 9$). Key exclusion criteria were cerebral large-vessel disease, brain tumors, and cardiovascular instability requiring intensive care or therapeutic intervention. All participants (or their legal representatives) provided written informed consent to undergo brain MRI and PET scanning with florbetaben and to donate their brain for postmortem examination. Details of the study methods have been presented previously (3).

Brain Image Data Acquisition

PET images were acquired 90–110 min after intravenous injection of 300 MBq ± 20% florbetaben (5) according to a standardized acquisition and image-processing protocol established during a technical visit to each center. Three-dimensional volumetric T1-weighted brain MR images (e.g., magnetization prepared rapid gradient echo or spoiled gradient recalled sequences) were collected.

Study Design

Florbetaben scans were assessed visually and quantitatively and results compared with β-amyloid presence/absence in pathology. For visual assessment, either in-person training ($n = 3$ expert interpreters) or an electronic training (e-training) method ($n = 5$ naïve interpreters) was used. Composite SUVRs were determined (6), and receiver-operating-characteristic curve analysis was used to ascertain the optimal threshold for the sensitivity/specificity calculations. The composite SUVR providing the highest sum of sensitivity and specificity was selected as a cutoff value.

Visual Assessment Method

The in-person expert training and electronic training modules were identical in approach and content (6,7). The training emphasized normal white matter anatomy using structural MRI and coregistered florbetaben PET images to appreciate white matter–gray matter boundaries because a positive scan demonstrates extension of radiotracer uptake beyond the cortical white matter to adjacent gray matter in key brain regions. Specifically, interpreters used a regional cortical tracer uptake scoring system (RCTU) (1, no tracer uptake; 2, moderate tracer uptake; 3, pronounced tracer uptake) in 4 brain areas: lateral temporal cortex, frontal cortex, posterior cingulate cortex/precuneus, and parietal cortex (Fig. 1). The resulting scores condense into a binary interpretation (score 1, negative; score 2 or 3, positive). An RCTU score of 1 in each brain region led to a brain amyloid plaque load (BAPL) score of 1, and an RCTU score of 2 in any brain region and no score 3 led to a BAPL score of 2. An RCTU score of 3 in any of the 4 brain regions led to a BAPL of 3.

No access to other scan orientations (i.e., coronal, sagittal) and no reorientation or structural information from CT or MRI were available to interpreters. Interpreters viewed scans in gray scale only. There were some minor differences between the training, as the trainees could not ask questions of an expert interpreter during the review with the electronic training tool.

All images were assessed by 8 interpreters: 3 in-person–trained experts and 5 naïve interpreters trained using an e-training tool. All interpreters were nuclear medicine physicians. The expert interpreter was defined as having direct experience with PET amyloid scans. The naïve interpreter had no experience with the visual assessment method and was not involved in any pivotal study using an ^{18}F-labeled amyloid imaging agent.

Quantitative Assessment Method

Brain PET image quantification was performed using a standardized volume-of-interest template applied to the spatially normalized
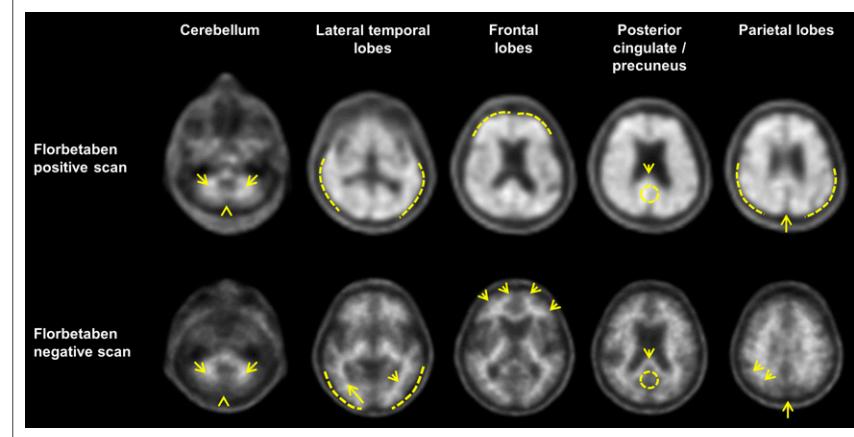

FIGURE 1. Visual patterns of abnormality and normality taught to interpreters using electronic training tool or during in-person training. (Cerebellum) Contrast between white matter (arrows) and gray matter is seen in both negative and positive scans. Extracerebral tracer uptake in scalp and in posterior sagittal sinus (arrowhead) can be seen. (Lateral temporal lobes) Positive scan shows plumped, smooth appearance of outer border of brain (dashed line) from tracer uptake in gray matter. Spiculated or mountainous appearance of white matter (arrows) characterizes negative scan. (Frontal lobes) Positive scan shows that tracer uptake has plumped, smooth appearance due to gray matter signal (dashed line). Spiculated appearance of white matter in frontal lobes (arrows) is seen in negative scan. (Posterior cingulate/precuneus) Adjacent to splenium (arrow), region appears as hypointense hole (circle) in negative scan, whereas this hole is absent (circle) in positive scan. (Parietal lobes) In positive scan, midline between parietal lobes is thinner. Cortical areas are filled up and show smoother appearance as uptake extends to outer rim. In negative scan, midline between parietal lobes can be easily identified (long arrow); white matter has spiculated appearance (short arrow) with less uptake to outer rim (dashed line).

Sensitivity, Specificity, and Accuracy of Visual Florbetaben PET Scan Interpretations for 8 Interpreters ($n = 82$ Scans)

Training	Interpreter	BSS Sensitivity	BSS Specificity	BSS Accuracy	BSS/IHC Sensitivity	BSS/IHC Specificity	BSS/IHC Accuracy
In-person	1	96.2 (90.9–100.0)	83.3 (70.0–96.7)	91.5 (85.4–97.5)	94.6 (88.7–100.0)	92.3 (82.1–100.0)	93.9 (88.7–99.1)
	2	98.1 (94.3–100.0)	76.7 (61.5–91.8)	90.2 (83.8–96.7)	98.2 (94.7–100.0)	88.5 (76.2–100.0)	95.1 (90.5–99.8)
	3	98.1 (94.3–100.0)	80.0 (65.7–94.3)	91.5 (85.4–97.5)	98.2 (94.7–100.0)	92.3 (82.1–100.0)	96.3 (92.3–100.0)
Electronic media	1	94.2 (87.9–100.0)	80.0 (65.7–94.3)	89.0 (82.3–95.8)	92.9 (86.1–99.6)	88.5 (76.2–100.0)	91.5 (85.4–97.5)
	2	98.1 (94.3–100.0)	46.7 (28.8–64.5)	79.3 (70.5–88.0)	98.2 (94.7–100.0)	53.8 (34.7–73.0)	84.1 (76.2–92.1)
	3	90.4 (82.4–98.4)	80.0 (65.7–94.3)	86.6 (79.2–94.0)	91.1 (83.6–98.5)	92.3 (82.1–100.0)	91.5 (85.4–97.5)
	4	96.2 (90.9–100.0)	76.7 (61.5–91.8)	89.0 (82.3–95.8)	96.4 (91.6–100.0)	88.5 (76.2–100.0)	93.9 (88.7–99.1)
	5	100.0	56.7- (38.9–74.4)	84.1 (76.2–92.1)	100.0	65.4 (47.1–83.7)	89.0 (82.3–95.8)
Statistics	P	0.64	0.28	0.03	0.76	0.21	0.05

IHC = immunohistochemistry; Statistics = independent Mann–Whitney tests.
Data in parentheses are 95% CIs.

gray matter PET image based on a gray/white/cerebrospinal fluid segmentation of the participant's T1-weighted volumetric MRI (6). A region-of-interest template (6) sampled the lateral temporal, frontal, anterior and posterior cingulate gyrus/precuneus, and parietal lobes, which were averaged to determine composite SUVRs calculated using the cerebellar cortex as the reference tissue. Eighty-one of the 82 brains were evaluated in this fashion; in 1 scan the segmentation process failed because of poor technical quality of the MRI.

Pathology SoT

Brain samples from 82 participants who died during the study were used to determine the histopathologic SoT. Six brain regions were examined with both Bielschowsky silver staining (BSS) and immunohistochemistry for the β-amyloid protein: middle frontal gyrus, occipital cortex, hippocampus/parahippocampal gyrus, anterior cingulate cortex, posterior cingulate cortex/precuneus, and cerebellar cortex.

The presence of amyloid plaques was assessed by a blinded histopathology consensus panel of 3 expert neuropathologists using 2 different methods: BSS and immunohistochemistry for β-amyloid. For the analysis presented here, only neuritic plaques and cored plaques were considered. Neuritic plaque density (as detected by BSS) was assessed according to the Consortium for Establishing a Registry for Alzheimer Disease criteria (8), providing a semiquantitative score with

the categories absent, sparse, moderate, or frequent. The same semiquantitative categories were used to score the number of cored plaques detected by β-amyloid immunohistochemistry. β-amyloid was regarded as present in a given brain region when sufficient neuritic or cored plaques were present to achieve a score of moderate or frequent.

Importantly, BSS is not specific for β-amyloid deposits and also has some technical limitations that may hinder the identification of some neuritic plaques in AD (9). The combination of BSS and immunohistochemistry for β-amyloid is recommended in current neuropathologic guidelines for assessment of AD pathology (10,11). Therefore, both BSS and BSS + immunohistochemistry data were used for further analyses.

Statistical Analysis

Sensitivity, specificity, and accuracy were evaluated by comparing visual assessments with the histopathologic SoT. Independent Mann–Whitney tests were performed to compare the reported parameters between training groups. The reliability of visual assessment was evaluated by Cohen and Fleiss κ. To investigate the intrainterpreter agreement, a random subsample of images (22 for the in-person–trained group and 20 for the e-trained group) was reinterpreted by all interpreters. The association between quantitative and visual assessments was evaluated with a χ^2 test.

TABLE 2
Interinterpreter Cohen κ-Values*

In-person	In-person training 1	2	3	Electronic media training	1	2	3	4	5
1	x	0.93 (0.85–1.00)	0.89 (0.79–0.98)	1	x	0.57 (0.38–0.75)	0.87 (0.76–0.97)	0.91 (0.82–1.00)	0.63 (0.45–0.80)
2		x	0.86 (0.75–0.97)	2		x	0.53 (0.34–0.71)	0.66 (0.48–0.83)	0.75 (0.59–0.92)
3			x	3			x	0.86 (0.76–0.97)	0.58 (0.41–0.76)
				4				x	0.71 (0.56–0.87)
				5					x

*κ-values in range of 0.81–1.00 are considered excellent, 0.61–0.80 very good, and 0.41–0.60 moderate.
Data in parentheses are 95% CIs.

TABLE 3
Intrainterpreter Cohen κ-Values

Training	Interpreter	Cohen κ
In-person (n = 22)	1	0.90 (0.72–1.00)
	2	0.90 (0.71–1.00)
	3	0.79 (0.51–1.00)
Electronic media (n = 20)	1	1.00 (1.00–1.00)
	2	0.66 (0.30–1.00)
	3	0.90 (0.71–1.00)
	4	0.90 (0.70–1.00)
	5	1.00 (1.00–1.00)

Data in parentheses are 95% CIs.

RESULTS

Study Population and Postmortem β-Amyloid Histopathology

In total, 205 end-of-life individuals underwent florbetaben PET imaging (52% men; mean age ± SD, 76.9 ± 11 y [range, 48–98 y]). As of August 2013, 82 participants had died and undergone autopsy and postmortem histopathology (clinical diagnoses: AD, n = 60; dementia with Lewy bodies, n = 4; other dementias, n = 9; nondemented, n = 9). Comparison of clinical diagnosis of AD and β-amyloid pathology (BSS + immunohistochemistry as SoT) revealed a discordance in some patients with AD, with 13 of the 60 patients with a clinical diagnosis of probable AD found to be β-amyloid–negative. Of the 4 subjects with dementia with Lewy bodies as clinical diagnosis, 2 were β-amyloid–positive. Seven of the 9 subjects clinically classified as other dementia were β-amyloid–negative. Four of the 9 nondemented healthy elderly subjects were β-amyloid–positive. In total, of the 82 brains examined by histopathology, 56 were β-amyloid–positive and 26 were β-amyloid–negative with BSS + immunohistochemistry as SoT. When BSS was used alone, 52 brains were β-amyloid–positive and 30 β-amyloid–negative.

Sensitivity and Specificity of Visual Assessment

Of the 82 PET scans assessed in this study, 81 were classified equally by both training groups. With BSS + immunohistochemistry as SoT, a median sensitivity of 98.2% (range, 94.6%–98.2%) was obtained for the 3 in-person–trained interpreters and 96.4% (range, 91.1%–100%) for the 5 e-trained interpreters. With BSS alone, median sensitivity was 98.1% (range, 96.2–98.1) and 96.2%
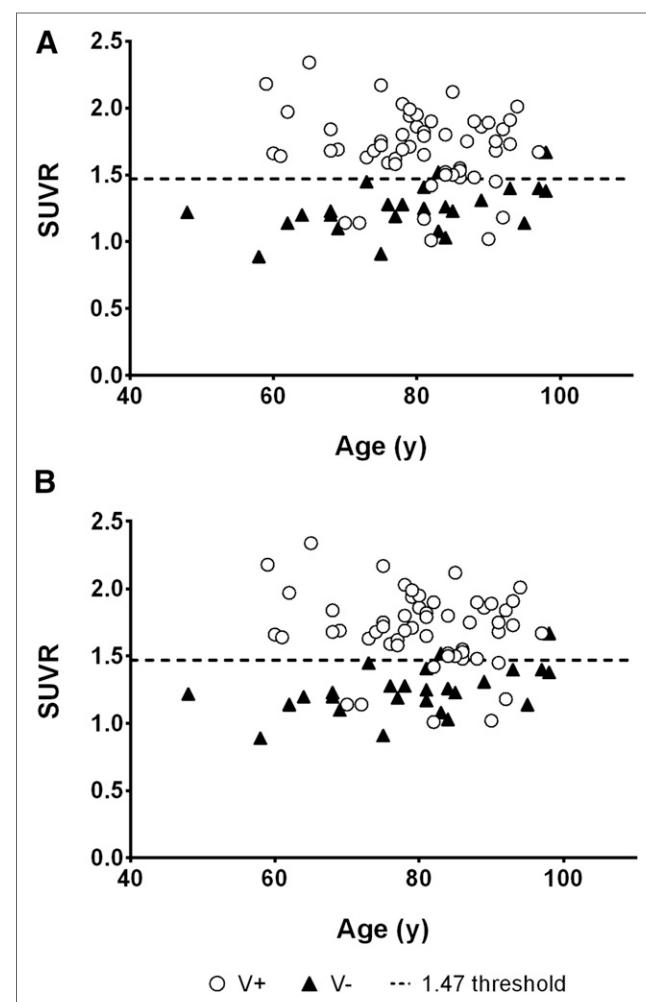
FIGURE 2. Comparison of visual and quantitative image assessment. Visually assessed images were based on most results classified into V+ (visual-positive scan) and V− (visual-negative scan). Assessment data are plotted related to composite SUVR and subject age (y). Most interpretation data of electronically trained interpreters (A) and most interpretation data of in-person–trained interpreters (B). Electronically trained interpreters (A) and in-person–trained interpreters (B) are indicated separately.

(range, 90.4%–100%), respectively (Table 1). With BSS + immunohistochemistry, median specificity was 92.3% (range, 88.5%–92.3%) for the in-person–trained interpreters and 88.5% (range, 53.9%–92.3%) for the e-trained interpreters, whereas BSS alone led to 80.0% (range,

TABLE 4
Electronically Trained Visual Assessment in Comparison to Composite SUVR Quantification

	Quantitative assessment		
Visual assessment	Positive	Negative	Total
Positive	48	8	56
Negative	2	23	25
Total	50	31	81

TABLE 5
In-Person Visual Assessment in Comparison to Composite SUVR Quantification

	Quantitative assessment		
Visual assessment	Positive	Negative	Total
Positive	48	7	55
Negative	2	24	26
Total	50	31	81

TABLE 6
Comparison of Visual Versus Semiquantitative Analysis for Discordant Cases*

Case	Age (y)	Visual in-person trained	Agreement	Visual e-trained	Agreement	Quantification	SUVR	PET evaluation	Histopathology (BSS + IHC)	Clinical diagnosis
3	82	Positive	100%	Positive	100%	Negative	1.01	Marked atrophy	Positive	AD
5	90	Positive	100%	Positive	100%	Negative	1.02	Marked atrophy	Positive	AD
39	82	Positive	100%	Positive	100%	Negative	1.42	Marked atrophy	Positive	AD
40	70	Positive	100%	Positive	100%	Negative	1.14	—	Positive	Other dementia
43	92	Positive	100%	Positive	80%	Negative	1.18	Marked atrophy	Positive	AD
44	72	Positive	100%	Positive	100%	Negative	1.14	Marked atrophy	Positive	AD
45	91	Positive	100%	Positive	100%	Negative	1.45	Marked atrophy	Positive	AD
75	81	Negative	67%	Positive	80%	Negative	1.17	—	Negative	AD
22	98	Negative	67%	Negative	80%	Positive	1.67	—	Positive	AD
67	83	Negative	100%	Negative	80%	Positive	1.52	Marked atrophy	Negative	Other dementia

*Discordant cases are those visually interpreted as positive but quantified as negative or visually interpreted as negative and quantified as positive.

Agreement = Interinterpreter agreement within training group; PET evaluation = additional PET assessment conducted after initial study; IHC = immunohistochemistry.

76.7%–83.3%) and 76.7% (range, 46.7%–80%), respectively (Table 1). Median accuracy with BSS + immunohistochemistry was 95.1% (range, 93.9%–96.3%) for in-person–trained interpreters and 91.5% (range, 84.1%–93.9%) for e-trained interpreters. BSS alone as SoT had a median accuracy of 91.5% (range, 90.2%–91.5%) for in-person–trained interpreters and 86.6% (79.3%–89%) for e-trained interpreters (Table 1). Independent Mann–Whitney tests were performed to compare the reported parameters between training groups. No statistical differences were found for sensitivity and specificity, independent of the SoT. Accuracy was higher for expert in-person–trained interpreters than for the e-trained interpreters ($P = 0.03$) (Table 1). Individual interpreter results of visual assessments are provided in Supplemental Table 1 (supplemental materials are available at http://jnm.snmjournals.org).

Inter- and Intrainterpreter Agreement

Interinterpreter agreement (Fleiss κ) was 0.89 (considered excellent; 95% confidence interval [CI], 0.82–0.97) for expert in-person–trained interpreters and 0.71 (considered very good; 95% CI, 0.62–0.81) for naïve e-trained interpreters (Table 2). Median intrainterpreter agreement was 0.9 (range, 0.79–0.90) for expert in-person–trained interpreters and 0.9 (range, 0.66–1.00) for e-trained interpreters (Table 3).

Comparison of Visual and Quantitative Assessments

Eighty-one of 82 images were evaluated with both quantitative and visual methods by the in-person–trained interpreters and e-trained interpreters. Receiver-operating-characteristic curve analysis of composite gray matter SUVRs resulted in an optimal cutoff of 1.47 with a sensitivity of 85.7% (95% CI, 73.8%–93.6%) and a specificity of 92.0% (95% CI, 74.0%–99.0%). Seventy-one (87.7%) scans were classified as positive or negative by both the e-trained interpreters and the quantitative approaches, with the remaining 10 (12.3%) scans showing a discrepancy between the 2 approaches. Both methods (visual assessment by 5 e-trained interpreters and SUVR quantification) were significantly and strongly related ($\chi^2 = 44.19$, $P < 0.0001$, κ = 0.73; Table 4; Fig. 2A). Similarly, for

in-person–trained interpreters, 72 (88.9%) assessments were concordant and 9 (11.1%) were discordant with quantitation, with a significant relationship between methods ($\chi^2 = 47.33$, $P < 0.0001$, κ = 0.76; Table 5; Fig. 2B).

Further details of the 10 discordant cases are summarized in Table 6. Concordance with the pathology results was found with visual assessment in 9 cases by in-person–trained interpreters and in 8 cases by e-trained interpreters, but with the quantitative assessment only 2 cases matched the pathology results. Marked atrophy was found in 7 cases, all with visual assessment matching pathology results (6 positive and 1 negative case). However, for the 7 marked atrophy cases none of the quantitative assessment results matched with pathology, showing SUVRs less than 1.47 in the 6 positive cases and an SUVR of 1.52 in the negative case with marked atrophy. Interpreter agreement for 8 of 10 cases was very high both for the in-person– and for the electronically trained interpreter groups. Only for 1 subject (case 75 in Table 6; Supplemental Table 1) did the visual assessments differ for the 2 methods.

Challenging Cases

The patient cohort included challenging cases, such as subjects with marked brain atrophy or scans with head motion. Figure 3 shows sample florbetaben PET scans from a patient with marked brain atrophy, with potential for false-positive assessment. A small percentage of scans demonstrated motion as illustrated in Figure 4.

DISCUSSION

Results from this study demonstrate a high sensitivity and specificity of florbetaben PET imaging for evaluation of β-amyloid plaques in end-of-life individuals. Importantly, the sensitivity and specificity were independent of the interpreter training method, the previous expertise of interpreters in β-amyloid PET scan assessment, and the histopathology method used (BSS with or without immunohistochemistry). However, specificity increased for BSS + immunohistochemistry compared with BSS alone, as 4 brains in the BSS + immunohistochemistry group were additionally categorized
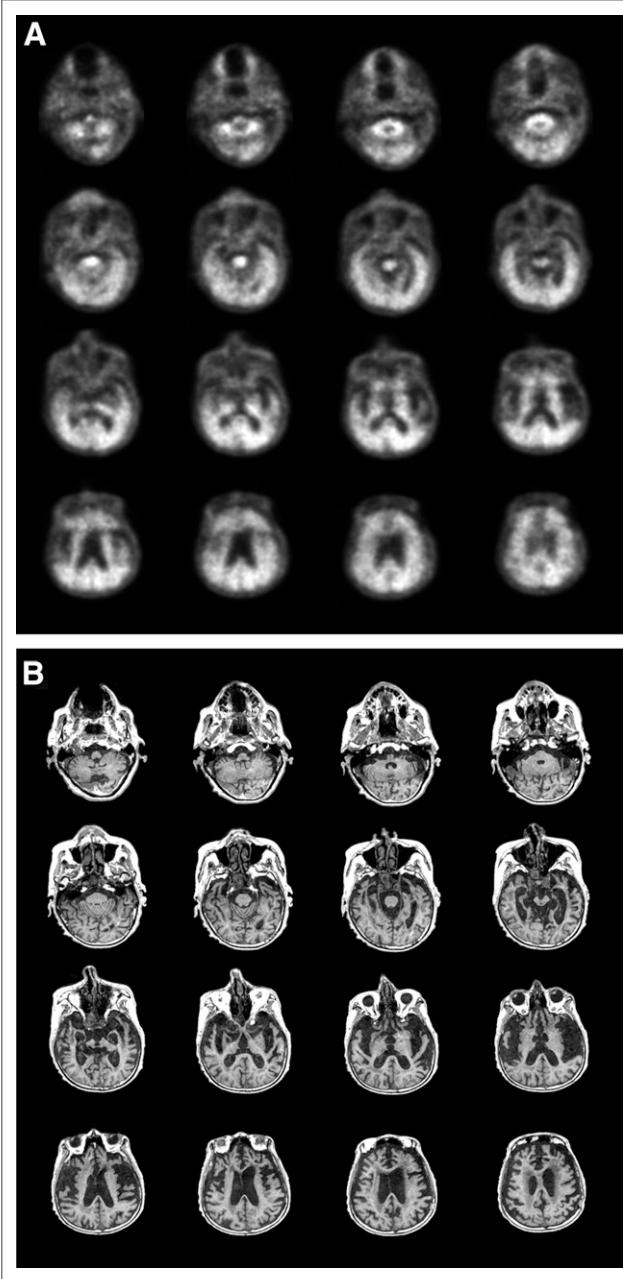
FIGURE 3. Florbetaben PET (A) and MRI (B) scans from patient 71 with marked brain atrophy, which was assessed by all interpreters as β-amyloid–positive. Patient was found to be β-amyloid–negative on postmortem histopathology and positive in quantitative assessment (SUVR, 1.53).

as β-amyloid–positive. Indeed, BSS + immunohistochemistry is recommended in current neuropathologic guidelines for the assessment of AD pathology (*10,11*). Intrainterpreter and interinterpreter agreement was very high for both training groups.

The individual differences in sensitivity and specificity among the 5 inexperienced interpreters are particularly informative. There was a wide range of specificity in particular, with 2 interpreters showing relatively lower performance than the other 3. Poststudy interviews with each individual interpreter suggested that those with poorer performance were less rigid in the application of the training rules to their visual assessments. Specifically, when assessing whether a particular region is positive, the algorithm requires that most (i.e., at

least 50%) of the axial slices comprising each region must be positive. In some instances, the poorer interpreters did not apply this rule systematically, leading to a region being classified as positive. Nonetheless, the sensitivity and specificity in the present study were consistent with earlier analyses from the same study (*3*) and also with previous reports with florbetaben PET (*6,12*) using clinical diagnosis as the SoT. Pathology examinations have shown, however, that clinical diagnosis of AD can be wrong (i.e., individuals diagnosed with AD dementia who do not show β-amyloid plaques on autopsy) in 10%–30% of cases (*13*). This can lead to false-negatives
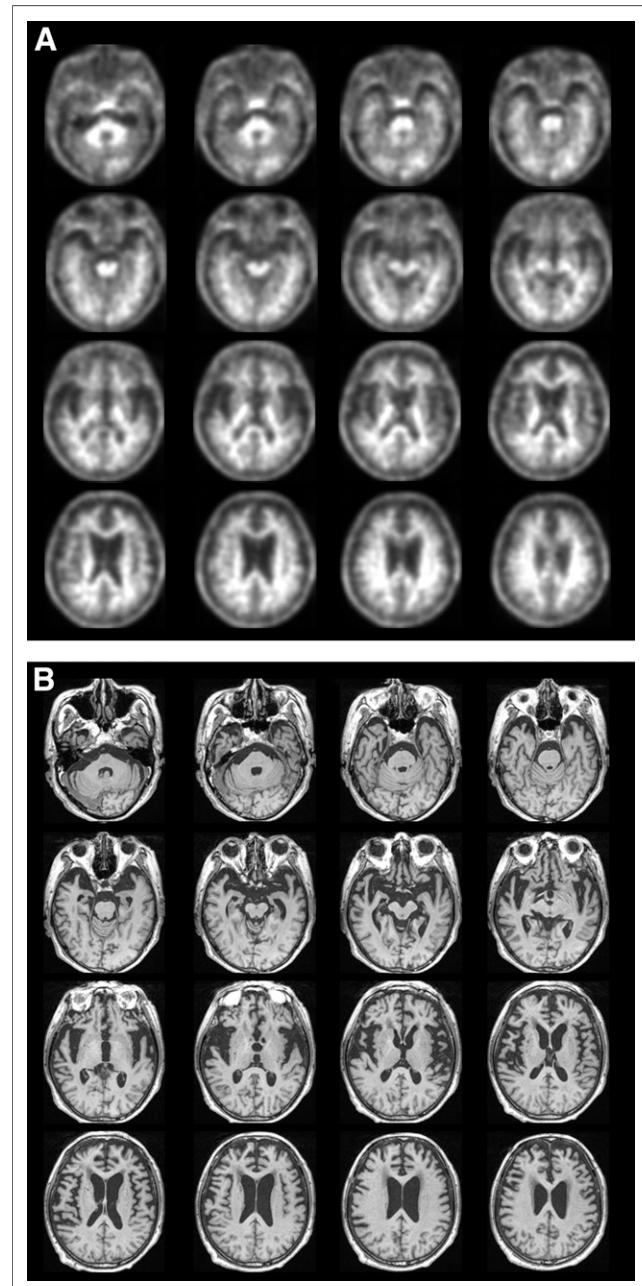
FIGURE 4. Florbetaben PET (A) and MRI (B) scans from patient 68 with motion artifact, potentially leading to false-positive visual assessment. All interpreters assessed scans as β-amyloid–negative. Patient was found to be β-amyloid–negative on postmortem histopathology and negative on quantitative assessment (SUVR, 1.10).

compromising the sensitivity estimation when using clinical diagnosis as SoT.

The relationship of visual interpretations to quantitative values was also consistent with prior reports (6,14), indicating high concordance between visual positivity and negativity and the composite SUVR using a quantitative cutoff for positive and negative scans. There were only a few cases (10 for e-trained and 9 for in-person–trained of 81 case evaluations) in which there was discordance between visual and quantitative assessments. The visual assessment method used did not allow comparison of the PET scan images with CT or MR. This may lead to difficulties in the interpretation of some cases with atrophy, and quantification could potentially help here. Most of the visual versus quantitative discordant cases showed marked atrophy. However, in this study the visual assessment of discordant cases matched with pathology results. Quantitative assessment of cases with marked atrophy were all discordant with pathology results. The interpreters were more adept at distinguishing gray matter from white matter uptake in scans with severe atrophic changes than the quantification method. Two exceptions to this are shown in Supplemental Figures 1 and 2. This suggests that atrophy may affect the quantification method used in this study more than the visual assessment. Further investigation is required to substantiate this and whether partial-volume error correction influences this result.

The end-of-life population used in this study is not the intended population for β-amyloid PET scanning. The clinically intended population will likely have fewer structural brain abnormalities than observed in this study cohort. Furthermore, the interpretation methodology designed for florbetaben in the present study was quite restrictive, with interpreters not permitted to use all the tools routinely available in nuclear medicine for PET assessment (3). In clinical practice, PET scans are interpreted using 3 spatial orientations, with structural images provided by CT or MRI to guide anatomic localization of findings, often in discontinuous color scales for fused images. Additional use of all available tools for image interpretation will likely increase the diagnostic performance. Nevertheless, excellent results were obtained, even allowing for the challenging nature of some scans and the stringent requirements of the applied interpretation methodology.

CONCLUSION

Overall, visual assessment of florbetaben images was robust, even in challenging scans from elderly end-of-life individuals. Sensitivity and specificity were high, as was interinterpreter agreement, independent of the interpreter expertise and training method used. The visual assessment strategy and respective training tools to analyze florbetaben PET scans are accurate and reliable in the detection of brain neuritic β-amyloid plaques as assessed using BSS, and cored plaques as assessed using immunohistochemistry, and showed a good correlation with quantitative assessments.

DISCLOSURE

REFERENCES

1. Piramal Imaging. NEURACEQ (florbetaben F 18 injection), for intravenous use [package insert]. Matran, Switzerland: Piramal Imaging; 2014.
2. Piramal Imaging Limited. *Neuraceq. Summary of Product Characteristics.* Cambridge, U.K.: Piramal Imaging Limited; 2014.
3. Sabri O, Sabbagh MN, Seibyl J, et al. Florbetaben PET imaging to detect amyloid plaques in Alzheimer disease: phase 3 study. *Alzheimers Dement.* 2015;11:964–974.
4. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med.* 2009;50(suppl 1):11S–20S.
5. Becker GA, Ichise M, Barthel H, et al. PET quantification of ^{18}F-florbetaben binding to beta-amyloid deposits in human brains. *J Nucl Med.* 2013;54:723–731.
6. Barthel H, Gertz HJ, Dresel S, et al. Cerebral amyloid-beta PET with florbetaben (^{18}F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. *Lancet Neurol.* 2011;10:424–435.
7. Barthel H, Luthardt J, Becker G, et al. Individualized quantification of brain beta-amyloid burden: results of a proof of mechanism phase 0 florbetaben PET trial in patients with Alzheimer's disease and healthy controls. *Eur J Nucl Med Mol Imaging.* 2011;38:1702–1714.
8. Mirra SS, Heyman A, McKeel D, 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;41:479–486.
9. Uchihara T. Silver diagnosis in neuropathology: principles, practice and revised interpretation. *Acta Neuropathol (Berl).* 2007;113:483–499.
10. Hyman BT, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement.* 2012;8:1–13.
11. Montine TJ, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol (Berl).* 2012;123:1–11.
12. Tiepolt S, Barthel H, Butzke D, et al. Influence of scan duration on the accuracy of beta-amyloid PET with florbetaben in patients with Alzheimer's disease and healthy volunteers. *Eur J Nucl Med Mol Imaging.* 2013;40:238–244.
13. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *J Neuropathol Exp Neurol.* 2012;71:266–273.
14. Jennings D, Seibyl J, Sabbagh M, et al. Age-dependence of brain β-amyloid load in Down syndrome: a [^{18}F]florbetaben PET study. *Neurology.* 2015;84:500–507.