

# Assessing Amyloid Pathology in Cognitively Normal Subjects Using <sup>18</sup>F-Flutemetamol PET: Comparing Visual Reads and Quantitative Methods

Lyduine E. Collij\*<sup>1</sup>, Elles Konijnenberg\*<sup>2</sup>, Juhan Reimand<sup>1,3,4</sup>, Mara ten Kate<sup>2</sup>, Anouk den Braber<sup>2,5</sup>, Isadora Lopes Alves<sup>1</sup>, Marissa Zwan<sup>2</sup>, Maqsood Yaqub<sup>1</sup>, Daniëlle M.E. van Assema<sup>6</sup>, Alle Meije Wink<sup>1</sup>, Adriaan A. Lammertsma<sup>1</sup>, Philip Scheltens<sup>2</sup>, Pieter Jelle Visser<sup>2</sup>, Frederik Barkhof<sup>1,7</sup>, and Bart N.M. van Berckel<sup>1</sup>

<sup>1</sup>Department of Radiology and Nuclear Medicine, VU Medical Center, Amsterdam, The Netherlands; <sup>2</sup>Alzheimer Center and Department of Neurology, VU Medical Center, Amsterdam, The Netherlands; <sup>3</sup>Centre of Radiology, North Estonia Medical Centre, Tallinn, Estonia; <sup>4</sup>Department of Health Technologies, Tallinn University of Technology, Tallinn, Estonia; <sup>5</sup>Department of Biological Psychology, VU Amsterdam, The Netherlands; <sup>6</sup>Department of Radiology and Nuclear Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; and <sup>7</sup>Institute of Neurology and Healthcare Engineering, University College London, London, United Kingdom

Our objective was to determine the optimal approach for assessing amyloid disease in a cognitively normal elderly population. **Methods:** Dynamic <sup>18</sup>F-flutemetamol PET scans were acquired using a coffee-break protocol (a 0- to 30-min scan and a 90- to 110-min scan) on 190 cognitively normal elderly individuals (mean age, 70.4 y; 60% female). Parametric images were generated from SUV ratio (SUVr) and non-displaceable binding potential (BP<sub>ND</sub>) methods, with cerebellar gray matter as a reference region, and were visually assessed by 3 trained readers. Interreader agreement was calculated using  $\kappa$ -statistics, and semiquantitative values were obtained. Global cutoffs were calculated for both SUVr and BP<sub>ND</sub> using a receiver-operating-characteristic analysis and the Youden index. Visual assessment was related to semiquantitative classifications. **Results:** Interreader agreement in visual assessment was moderate for SUVr ( $\kappa = 0.57$ ) and good for BP<sub>ND</sub> images ( $\kappa = 0.77$ ). There was discordance between readers for 35 cases (18%) using SUVr and for 15 cases (8%) using BP<sub>ND</sub>, with 9 overlapping cases. For the total cohort, the mean ( $\pm$ SD) SUVr and BP<sub>ND</sub> were 1.33 ( $\pm$ 0.21) and 0.16 ( $\pm$ 0.12), respectively. Most of the 35 cases (91%) for which SUVr image assessment was discordant between readers were classified as negative based on semiquantitative measurements. **Conclusion:** The use of parametric BP<sub>ND</sub> images for visual assessment of <sup>18</sup>F-flutemetamol in a population with low amyloid burden improves interreader agreement. Implementing semiquantification in addition to visual assessment of SUVr images can reduce false-positive classification in this population.

**Key Words:** <sup>18</sup>F-flutemetamol PET; amyloid pathology; visual assessment; preclinical Alzheimer disease

**J Nucl Med 2019; 60:541–547**

DOI: 10.2967/jnumed.118.211532

Alzheimer disease (AD) is the most common cause of dementia, accounting for 60%–80% of cases above 65 y of age (1). Its pathologic hallmark is the accumulation of the amyloid- $\beta$  peptide, thought to start years before cognitive impairment (2). In fact, abnormal amyloid- $\beta$  levels are seen in 20%–40% of cognitively normal subjects between the ages of 60 and 90 y (3). These subjects are considered to be in the preclinical stage of AD (4,5), which provides a unique opportunity for secondary prevention studies and is gaining increasing research focus (6). To this end, reliable identification of amyloid pathology in vivo using PET is of the utmost importance in this population.

The identification of amyloid burden by means of visual interpretation of summed late images or of semiquantitative SUV ratio (SUVr) images is currently suggested to be sufficient. Previous studies have shown a high interreader agreement for the visual assessment of SUVr images and a high imaging–pathology correlation in clinical populations and end-of-life subjects (7–9). It has been shown, however, that SUVr overestimates amyloid burden compared with quantitative nondisplaceable binding potential (BP<sub>ND</sub>) (10). As such, quantitative BP<sub>ND</sub> images may be more reliable also for visual interpretation. In a memory clinic population, Zwan et al. showed that visual assessment of parametric BP<sub>ND</sub> <sup>11</sup>C-Pittsburgh compound B images resulted in a higher interreader agreement than the frequently used SUV and SUVr images (11). To date, it remains to be determined whether these findings translate to the increasingly available <sup>18</sup>F-labeled amyloid- $\beta$  targeting tracers, such as <sup>18</sup>F-flutemetamol, and, more importantly, to the challenging population of cognitively normal elderly participants who generally have a minimal amyloid load.

The purpose of this study was to compare 2 parametric imaging methods (SUVr vs. BP<sub>ND</sub>) to determine the optimal approach for assessment of early amyloid pathology. To this end, we investigated the agreement in visual assessment of SUVr and BP<sub>ND</sub> images between 3 readers and its relationship to (semi-)quantitative measures.

## MATERIALS AND METHODS

### Project

The data used in this study originate from the Innovative Medicines Initiative of the European Medical Information Framework for AD

Received Mar. 16, 2018; revision accepted Aug. 28, 2018.  
For correspondence or reprints contact: Lyduine E. Collij, Department of Radiology and Nuclear Medicine, VU Medical Center, De Boelelaan 1117, 1108 HV Amsterdam, The Netherlands.

E-mail: l.collij@vumc.nl

\*Contributed equally to this work.

Published online Oct. 12, 2018.

Immediate Open Access: Creative Commons Attribution 4.0 International License (CC BY) allows users to share and adapt with attribution, excluding materials credited to previous publications. License: <https://creativecommons.org/licenses/by/4.0/>. Details: <http://jnm.snmjournals.org/site/misc/permission.xhtml>.

COPYRIGHT © 2019 by the Society of Nuclear Medicine and Molecular Imaging.

(<http://www.emif.eu/>). The overall aim of this project is to discover and validate diagnostic markers, prognostic markers, and risk factors for AD in nondemented subjects.

### Subjects

In total, 199 subjects from the preclinical AD cohort were included at the Vrije Universiteit (VU) Medical Center. Inclusion criteria were an age of at least 60 y and normal cognition according to a delayed recall score that was more than  $-1.5$  SDs of the demographically adjusted normative data on the Consortium to Establish a Registry for Alzheimer Disease 10-word list (12), a Telephone Interview for Cognitive Status–modified score of 23 or higher (13), a 15-item Geriatric Depression Scale score of less than 11 (14), and a Clinical Dementia Rating Scale score of 0 (15). Exclusion criteria were any physical, neurologic, or psychiatric condition that interferes with normal cognition. PET acquisition failed in 3 subjects, and 6 BP<sub>ND</sub> images were lacking a visual assessment, resulting in 190 subjects who had a visual assessment for both SUV<sub>r</sub> and BP<sub>ND</sub> images. PET quantification failed in 5 subjects; thus, 185 subjects were used for the quantitative analysis. Written informed consent was obtained from all subjects, and the study was approved by the Medical Ethics Review Committee of the VU University Medical Center.

### PET

PET scans were obtained using an Ingenuity TF PET/MRI camera (Philips Healthcare). Thirty-minute scans were acquired immediately after a manual injection of <sup>18</sup>F-flutemetamol ( $191 \pm 20$  MBq) (16). After 60 min, during which the patient remained outside the scanner bed, a second scan of 20 min was acquired, starting 90 min after injection (17). Immediately before each part of the PET scan, a T1-weighted gradient echo pulse MRI scan was acquired for attenuation correction of the PET data. The first emission scan was reconstructed into 18 frames of increasing length ( $6 \times 5$ ,  $3 \times 10$ ,  $4 \times 60$ ,  $2 \times 150$ ,  $2 \times 300$ , and  $1 \times 600$  s) using the standard line-of-response–based row-action maximum-likelihood algorithm for the brain. The second scan was reconstructed with the same algorithm into 4 frames of 300 s each. First, Vinci Software, version 2.56 (Max Planck Institute for Neurologic Research), was used to combine the 2 PET scans into a single multi-frame image. Next, each individual's T1-weighted MR images were coregistered to the dynamic PET images using the generic multimodality setting of Vinci with a linear rigid-body schema and normalized mutual information as the similarity measure. Parametric BP<sub>ND</sub> images were generated from the entire image set using the receptor parametric mapping implementation in PPET (18–20). Generation of the SUV<sub>r</sub> images was based on the 90- to 110-min scan interval. Next, T1-based volumes of interest using the Hammers atlas implemented in PVElab software were projected onto the PET images to extract regional values (21). Cerebellar gray matter was used as reference tissue for both analyses (22). Finally, we computed global values based on the average of frontal (volume-weighted average of superior, middle, and inferior frontal gyri), parietal (volume-weighted average of posterior cingulate, superior parietal gyrus, postcentral gyrus, and inferolateral remainder of parietal lobe), and temporal (volume-weighted average of parahippocampal gyrus; hippocampus; medial temporal lobe; and superior, middle, and inferior temporal gyri) regions (23,24).

### MRI

Whole-brain scans were obtained using the 3-T Achieva scanner (Philips Healthcare) of the PET/MRI system described above equipped with an 8-channel head coil. Isotropic structural 3-dimensional T1-weighted images were acquired using a sagittal turbo field echo sequence with the following settings:  $1.00 \times 1.00 \times 1.00$  mm voxels, repetition time

of 7.9 ms, echo time of 4.5 ms, and flip angle of 8°. A 3-dimensional sagittal fat-saturated fluid-attenuated inversion recovery sequence was acquired using the following settings:  $1.12 \times 1.12 \times 1.12$  mm voxels, repetition time of 4,800 ms, echo time of 279 ms, and inversion time of 1,650 ms. The structural 3-dimensional T1 and 3-dimensional fluid-attenuated inversion recovery images were used for assessment of global cortical atrophy (25), average medial temporal atrophy (26), and Fazekas score for white matter hyperintensities (27,28).

### Visual Assessment of PET Images

Three trained readers, masked to clinical information, first assessed all SUV<sub>r</sub> images and subsequently all BP<sub>ND</sub> images, in a randomized order. Images deemed dubious by the reader were reassessed on a separate occasion. Images were scaled to 90% of the pons signal using rainbow color scaling, and transverse, sagittal, and coronal views were displayed using the software package Vinci, version 2.56. Images were rated as either positive (binding in one or more cortical brain regions or striatum unilaterally) or negative (predominantly white matter uptake) according to criteria defined by the manufacturer (GE Healthcare). PET images were assessed together with a T1-weighted MR scan to limit the influence of atrophy on the visual assessment.

The level of experience in visual assessment of <sup>18</sup>F-flutemetamol images differed among readers: a nuclear medicine physician with considerable experience, a nuclear medicine physician trainee with basic experience, and a radiologist in training with 6 mo of experience in nuclear medicine. All readers completed the <sup>18</sup>F-flutemetamol reader training provided by GE Healthcare.

### Statistical Analysis

Baseline demographics were assessed using simple descriptive statistical analyses.  $\kappa$ -statistics were used to assess interreader agreement among the 3 readers, intrareader agreement between the 2 methods, and agreement between visual and semiquantitative classifications. Agreement was considered poor if  $\kappa$  was less than 0.20, satisfactory if  $\kappa$  was 0.21–0.40, moderate if  $\kappa$  was 0.41–0.60, good if  $\kappa$  was 0.61–0.80, and excellent if  $\kappa$  was more than 0.80. Differences in MRI measurements between PET-negative and PET-positive cases were assessed using a Mann–Whitney *U* analysis. The correlation between semiquantitative SUV<sub>r</sub> and BP<sub>ND</sub> measurements was assessed using Spearman  $\rho$ . Cutoffs were calculated for both SUV<sub>r</sub> and BP<sub>ND</sub> using a receiver-operating-characteristic analysis and the Youden index. Possible overestimation of amyloid burden using semiquantitative SUV<sub>r</sub> was investigated by calculating the difference between SUV<sub>r</sub> – 1 and BP<sub>ND</sub> values. Differences in global overestimation between PET-negative and PET-positive cases were assessed using a Mann–Whitney *U* analysis. Regional differences in binding and overestimation were assessed using a Wilcoxon paired test. Amyloid status resulting from quantitative assessment was considered the true amyloid status for all analyses, in the absence of postmortem confirmation.

## RESULTS

Baseline demographics are provided in Table 1.

### Visual Reads

Interreader agreement in visual assessment was moderate for SUV<sub>r</sub> images ( $\kappa = 0.57$ ) and good for BP<sub>ND</sub> images ( $\kappa = 0.77$ ). There was discordance between readers for 35 cases (18%) using SUV<sub>r</sub> and for 15 cases (8%) using BP<sub>ND</sub>, with 9 overlapping cases. Figure 1 shows examples of agreement and disagreement in visual interpretation of <sup>18</sup>F-flutemetamol images. On average, the rating was positive in 35 (18%) of the SUV<sub>r</sub> images and in 26 (14%) of the BP<sub>ND</sub> images. The reader with the least experience classified

**TABLE 1**  
Demographics, MRI Measurements, and PET Values

Parameter	Data
Total cohort (n)	190
Women (n)	113 (59.5%)
Age (y)	70.4 ± 7.56
MMSE score	29 ± 1.13
Education (y)	15.15 ± 4.42
Global cortical atrophy score (0–3)	0.79 ± 0.72
Medial temporal atrophy score (0–4)	0.65 ± 0.72
Fazekas score (0–3)*	1.18 ± 0.82
Quantitative cohort (n)	185
SUVr	
Mean ± SD	1.33 ± 0.21
Range	0.79–2.13
BP <sub>ND</sub>	
Mean ± SD	0.16 ± 0.12
Range	0.20–0.66
Concordant cohort (n)	149
PET-negative (n)	130
Global cortical atrophy score (0–3)	0.74 ± 0.67
Medial temporal atrophy score (0–4)	0.57 ± 0.64
Fazekas score (0–3)*	1.18 ± 0.83
SUVr	
Mean ± SD	1.25 ± 0.09
Range	1.08–1.63
BP <sub>ND</sub>	
Mean ± SD	0.12 ± 0.05
Range	0.02–0.27
PET-positive (n)	19
Global cortical atrophy score (0–3)	0.89 ± 0.81
Medial temporal atrophy score (0–4)	0.82 ± 0.75
Fazekas score (0–3)*	1.26 ± 0.87
SUVr	
Mean ± SD	1.83 ± 0.16 <sup>§</sup>
Range	1.54–2.13
BP <sub>ND</sub>	
Mean ± SD	0.43 ± 0.12 <sup>§</sup>
Range	0.27–0.66

\*White matter hyperintensity (0 = 35; 1 = 101; 2 = 40; 3 = 14).  
<sup>§</sup>P < 0.01, compared with PET-negative group.

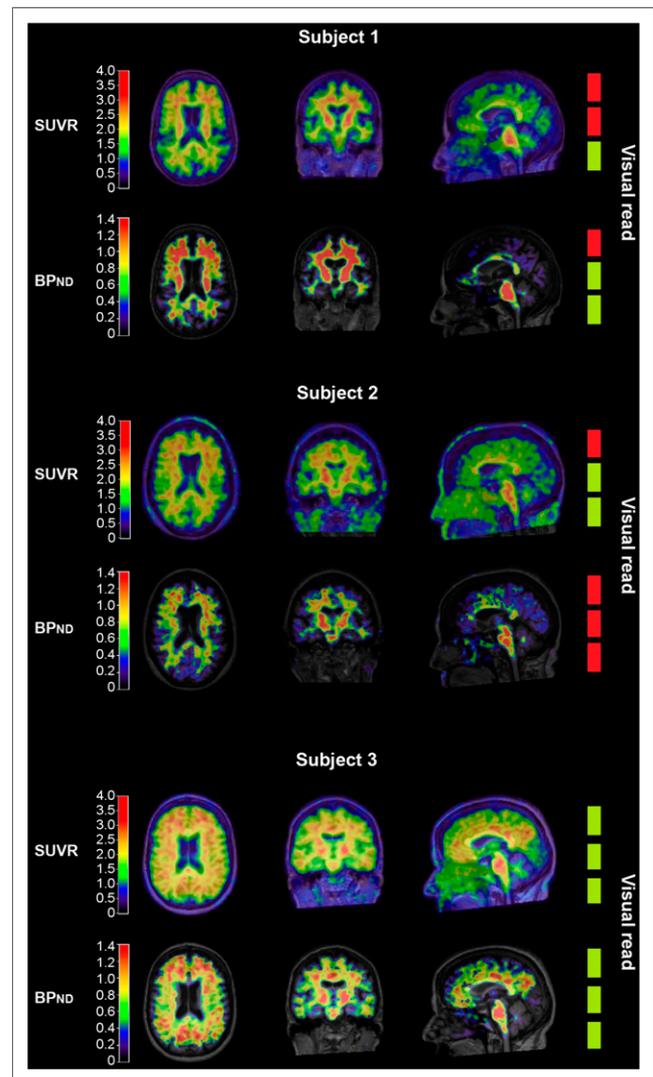
55 (29%) SUVr images as positive, compared with 21 (11%) and 29 (15%) by the intermediate and most experienced readers, respectively.

Intrareader agreement (i.e., within reader, between SUV<sub>r</sub> and BP<sub>ND</sub>) differed among readers, with moderate agreement ( $\kappa = 0.52$ ) between methods seen in the reader with least experience, excellent agreement ( $\kappa = 0.97$ ) in the reader with moderate experience, and good agreement in the reader with most experience ( $\kappa = 0.76$ ).

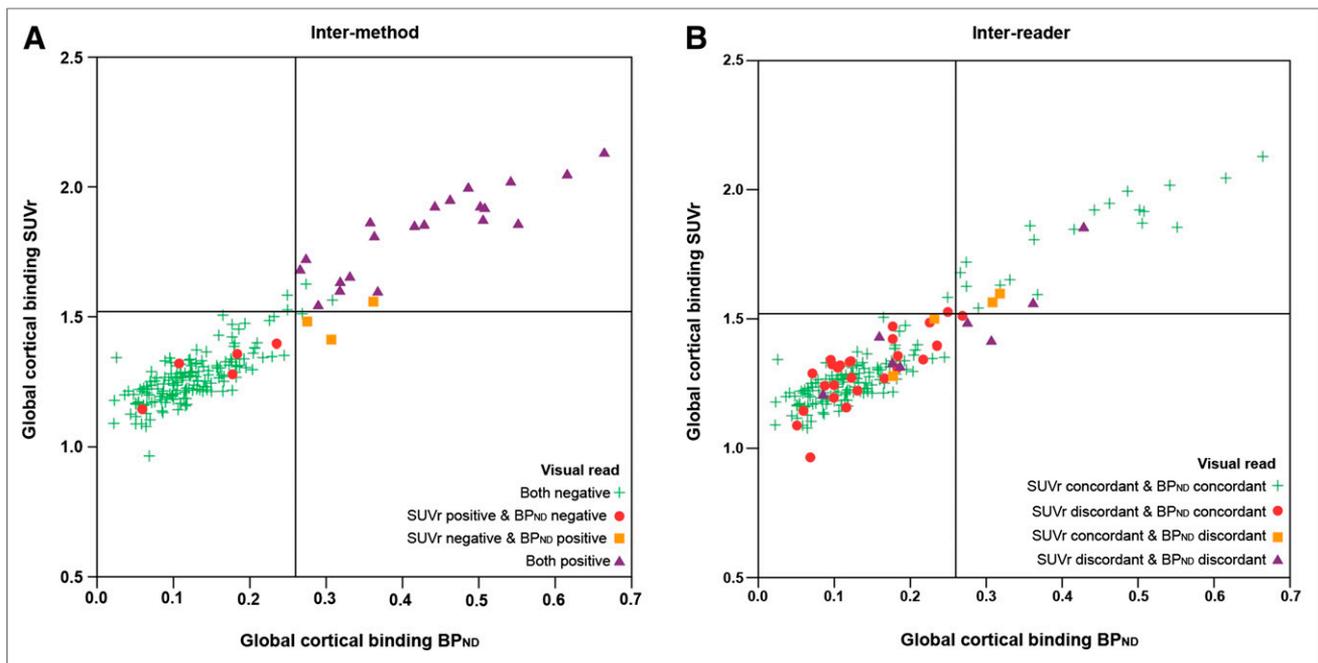
When applying majority rules (i.e., 2 of 3 readers agreed on a scan being either positive or negative), positivity was assigned to 27 (14%) cases based on SUV<sub>r</sub> and to 25 (13%) cases based on BP<sub>ND</sub>, with 22 overlapping cases. Thus, 8 cases showed inter-method discordance; that is, 5 cases were rated positive on SUV<sub>r</sub> but negative on BP<sub>ND</sub>, and 3 cases were rated positive on BP<sub>ND</sub> but negative on SUV<sub>r</sub>. The remaining 160 cases were classified as negative on both images (Fig. 2A).

#### Visual Reads Related to Quantitative Measures

For the total cohort, mean global SUV<sub>r</sub> and BP<sub>ND</sub> were  $1.33 \pm 0.21$  and  $0.16 \pm 0.12$ , respectively. There was good agreement between both measures (intraclass correlation coefficient, 0.89;  $P < 0.01$ ). Interreader-concordant positive cases had significantly higher SUV<sub>r</sub> and BP<sub>ND</sub> than concordant negative cases



**FIGURE 1.** Examples of SUV<sub>r</sub> and BP<sub>ND</sub> <sup>18</sup>F-flutemetamol images of 3 different patients. From left to right are shown axial, coronal, and sagittal views. The 3 boxes on right represent amyloid classification by 3 readers (red = negative; green = positive). Subject 1: Example of difficult case, represented by discordant visual reads on both SUV<sub>r</sub> and BP<sub>ND</sub> image. Subject 2: Example of possible overestimation of amyloid pathology when only SUV<sub>r</sub> image is assessed. Subject 3: Example of clearly positive case.



**FIGURE 2.** Scatterplot of quantitative measures compared with visual read. On x-axis is global cortical binding derived from BP<sub>ND</sub>. On y-axis is global cortical binding derived from SUVr. Reference lines denote cutoff (1.52 for SUVr and 0.26 for BP<sub>ND</sub>). Different colors demonstrate discordance or concordance between SUVr and BP<sub>ND</sub> visual read. (A) Visual read based on majority rules. For all intermethod discordant cases (red circles and orange squares), BP<sub>ND</sub> visual read was in accordance with quantitative value, whereas SUVr was not. (B) Most SUVr interreader discordant cases (red circles) are below cutoff for both SUVr and BP<sub>ND</sub>.

( $P < 0.01$ ) (Table 1). Based on the visual read–concordant cohort alone ( $n = 149$ ), the cutoff for positivity was 1.52 for SUVr (area under the curve, 0.98; sensitivity, 95%; specificity, 98%) and 0.26 for BP<sub>ND</sub> (area under the curve, 1.00; sensitivity, 100%; specificity, 98%) using a receiver-operating-characteristic analysis (Supplemental Fig. 1; supplemental materials are available at <http://jnm.snmjournals.org>). After applying both cutoffs to the dataset, the agreement between the SUVr majority visual read and semiquantitative negative–positive classification was good ( $\kappa = 0.78$ ), with 16 cases (9%) discordant between the 2 classification methods. The agreement analysis was also done with a literature-based cutoff (1.56) (8,29) resulting in a  $\kappa$  increase of 0.01. For BP<sub>ND</sub>, the agreement between the majority visual read and the quantitative negative–positive classification was excellent ( $\kappa = 0.93$ ), with 3 cases (2%) discordant between the 2 classification methods. Most of the 35 cases (91%) for which SUVr image assessment was discordant between readers were classified as negative using either cutoff (Fig. 2B). In addition, in the 8 cases with a discordant intermethod visual read, there was full agreement between visual and quantitative measurements when BP<sub>ND</sub> was used, which was not the case with SUVr (Fig. 2A).

#### SUVr $\neq$ BP<sub>ND</sub> Quantification

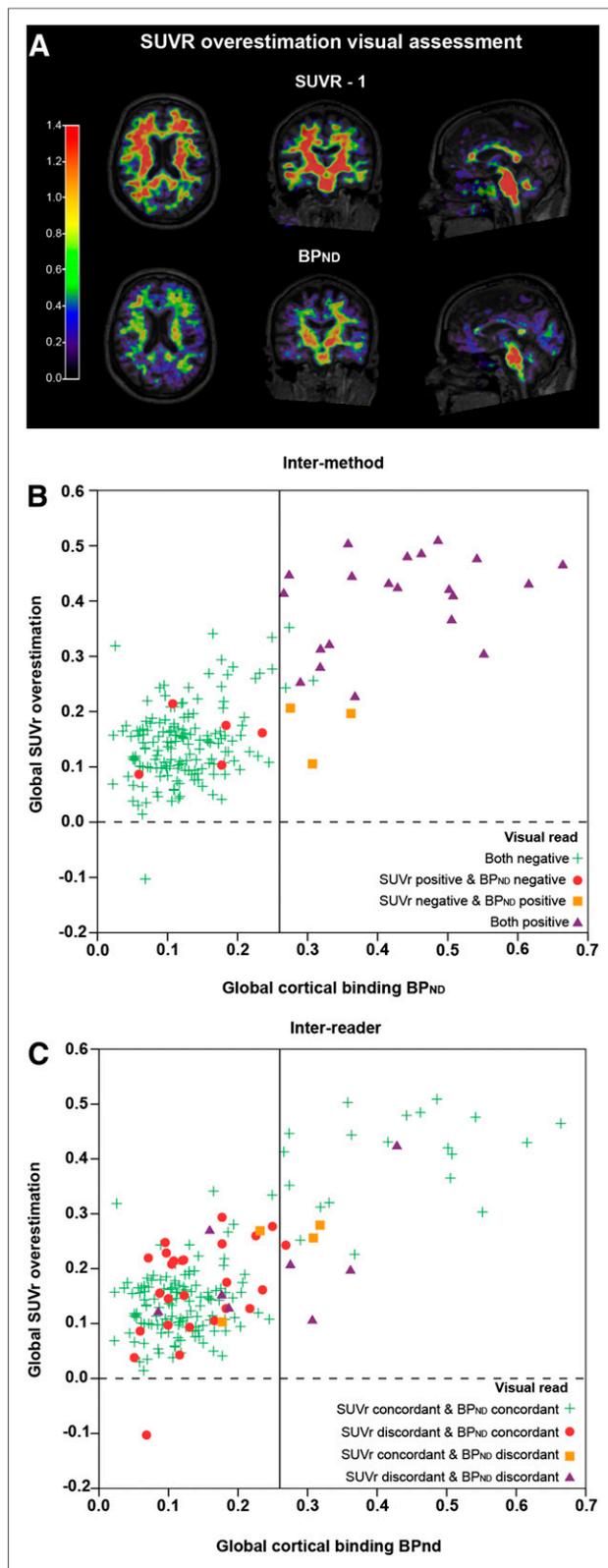
We investigated the relationship between the 2 quantitative measures with regard to the majority visual read to assess any violations of the equilibrium assumptions (i.e.,  $SUVr - 1 = BP_{ND}$ ) in this population. For all cases except one, global SUVr  $- 1$  values overestimated the corresponding global BP<sub>ND</sub> values. Participants with a positive read (mean overestimation [difference SUVr  $- 1$  and BP<sub>ND</sub>] =  $0.37 \pm 0.11$ ) had a significantly higher overestimation than participants with a negative read (mean overestimation =  $0.14 \pm 0.07$ ;  $P < 0.01$ ). This relationship was also

observed on a regional level, with the frontal lobe displaying the highest mean binding and the largest mean SUVr overestimation, compared with the parietal ( $P < 0.01$ ) and temporal ( $P < 0.01$ ) lobes. In turn, the parietal lobe did not show a significantly higher mean binding ( $P = 0.1$ ) but did show a significantly larger overestimation ( $P < 0.01$ ) than the temporal lobe (Supplemental Fig. 2; Table 1). The SUVr overestimation seems to have a limited influence on the visual read of the high-binding group (i.e., BP<sub>ND</sub>  $> 0.26$ ), considering no cases were visually assessed as positive on SUVr and negative on BP<sub>ND</sub> and only 2 SUVr images (7%) had a discordant read. For the low-binding group (i.e., BP<sub>ND</sub>  $\leq 0.26$ ), the SUVr overestimation might have influenced the visual read, considering that 26 cases (16%) had a SUVr-discordant visual read. However, no obvious pattern was discernible (Fig. 3).

#### DISCUSSION

In a cognitively normal elderly population with low amyloid burden, we show a considerable improvement in interreader agreement of <sup>18</sup>F-flutemetamol visual assessment when using BP<sub>ND</sub> rather than standard SUVr images. Misclassifications can be reduced using semiquantitative SUVr measures and avoided using fully quantitative BP<sub>ND</sub> measures.

Our results are in line with the <sup>11</sup>C-Pittsburgh compound B findings of Zwan et al., who found a comparable improvement in interreader agreement using BP<sub>ND</sub> images (11). This result suggests that the underlying reason for discrepant interreader agreements was tracer-independent and likely related to the distinctive metrics being used (SUVr and BP<sub>ND</sub>). SUVr is commonly used as a proxy for BP<sub>ND</sub>, under the assumption that a secular equilibrium is reached during scanning. However, these equilibrium conditions are rarely met in practice. As such, whereas parametric



**FIGURE 3.** Illustration of binding overestimation for semiquantitative PET acquisition. (A) From left to right are shown axial, coronal, and sagittal views. SUVR images with subtraction of 1 show clearly higher binding values than BP<sub>ND</sub> images, whereas in theory these images should be same. (B and C) Diagrams showing difference between SUVR - 1 and BP<sub>ND</sub> for each subject with regard to visual read. Overestimation of SUVR is higher with increasing cortical binding.

BP<sub>ND</sub> images reflect the density of available receptors (amyloid plaques), SUVR images are affected by a nondisplaceable (free and nonspecific) signal and may be affected by changes in regional flow and washout effects (28,30). As a result, SUVR can overestimate specific binding (10) and influence visual assessments (Fig. 3). Furthermore, our existing data show that this overestimation is not constant but instead increases with higher tracer binding (10,28).

The interreader agreement for the SUVR images and the concordance between semiquantitative and corresponding visual read classifications in our study are lower than previously reported (7–9). However, previous results were based on a clinical population of end-of-life subjects with a higher incidence of moderate to severe amyloid burden, which highlights the challenge of assessing amyloid pathology in a population with low amyloid burden. The challenge could be due to the nonspecific white matter uptake seen with <sup>18</sup>F-flutemetamol, which together with the overestimation resulting from static scanning may translate into a tendency to visually assign regions as positive (31). In our study, the frontal regions were most often perceived as difficult to assess, leading to the greatest doubt for final classification. Although the <sup>18</sup>F-flutemetamol reader training focuses on disentangling the white matter pattern from the cortical signal, assessment in this population seems additionally challenging, especially for less experienced readers. Indeed, the positive-assigning tendency was the strongest for the reader with the least experience, who also showed the lowest intrareader agreement between methods. This result stresses the need for experienced readers to make early assessments or for the reading guidelines to be updated, with the focus being on a cognitively normal elderly population. Of note, whereas the reference region used for visual assessment (i.e., pons) is different from that used for quantitative assessment (i.e., gray matter cerebellum), a separate agreement analysis using pons for quantification did not affect the agreement between classification methods.

Our results may have consequences for drug-intervention studies focused on early populations, since using the visual assessment of SUVR images as an inclusion criterion could result in false-positive inclusion due to the observed overestimation of cortical amyloid burden (32,33). Also, studies indicate that cerebral blood flow can change with age and disease progression (34,35). Therefore, using BP<sub>ND</sub> images in clinical trials could avoid false-positive classification in visual assessment (28) and ensure that measured changes are due to the treatment instead of a measurement error or blood flow confounders.

An important factor in considering dynamic PET acquisition is participant burden. In this cohort, 95% of participants indicated they had no objections to undergoing a second dynamic PET scan. The coffee-break protocol used in this study may have facilitated this response and suggests the feasibility of longitudinal dynamic acquisition in cognitively normal elderly persons.

In a clinical setting, however, amyloid burden will more likely be moderate to severe and dynamic acquisition more challenging. In addition, the utility of SUV or SUVR visual reads for the diagnosis of AD-type dementia in a clinical setting has been extensively shown (36). Thus, in this context, visual assessment of SUVR images may indeed be sufficient. Nevertheless, the present results illustrate that semiquantification using SUVR can help reduce false-positive classification, especially in a challenging population. Thus, the clinical preference for visual assessment could be revised in light of more available automatic

semiquantification methods, such as the one already provided for  $^{18}\text{F}$ -flutemetamol PET scans (8).

In this study, the standard manufacturer guidelines were used for reading both SUV<sub>r</sub> and BP<sub>ND</sub> images. Nonetheless, an interesting finding was the improvement in interreader agreement for BP<sub>ND</sub> images despite the lack of official guidelines and the limited experience of readers in assessing such images. However, it might still be of interest to formally assess whether the current guidelines are optimal for assessing BP<sub>ND</sub> images. In addition, optimizing visual assessment of SUV<sub>r</sub> images by updating the current guidelines and providing training specifically focused on early accumulation may also improve the certainty of classification, comparable to that observed using dynamically derived measures. Studies have suggested that, specifically, medial frontal, anterior/posterior/isthmus cingulate cortex, and precuneus are early-accumulating regions (37,38). These regions can be visually assessed using the sagittal view of the PET image. Thus, the importance of this plane may be of most interest for updating guidelines.

A limitation of this study is the lack of a gold standard, as no postmortem data were available, hampering the understanding of the findings in relation to underlying neuropathology. Furthermore, although the frequency of amyloid positivity in this cohort is comparable to previous reports (39), the low incidence may have induced reader bias with regard to searching for amyloid positivity. Lastly, both quantification and visual assessment of the PET images in this study were accompanied by structural MRI, which might not always be available.

## CONCLUSION

The use of parametric BP<sub>ND</sub> images for visual assessment of  $^{18}\text{F}$ -flutemetamol in a population with low amyloid burden improves interreader agreement. Implementing semiquantification in addition to visual assessment of SUV<sub>r</sub> images can reduce false-positive classification in this population.

## DISCLOSURE

This project received funding from the EU/EFPIA Innovative Medicines Initiative (IMI) Joint Undertaking (EMIF grant 115372) and the EU-EFPIA IMI-2 Joint Undertaking (grant 115952). This joint undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. Support was also received from the NIHR UCLH Biomedical Research Center, and in-kind sponsoring of the PET tracer was received from GE Healthcare. Philip Scheltens received grants from GE Healthcare, Piramal, and Merck, paid to his institution, and speaker's fees paid to the Alzheimer Center, VU University Medical Center, Lilly, GE Healthcare, and Roche. Pieter Jelle Visser received research support from Biogen and grants from EU/EFPIA IMI Joint Undertaking, EU Joint Programme—Neurodegenerative Disease Research (JPND), ZonMw, and Bristol-Myers Squibb; served as a member of the advisory board of Roche Diagnostics; and received nonfinancial support from GE Healthcare. Frederik Barkhof received payment and honoraria from Bayer-Schering Pharma, Sanofi-Aventis, Genzyme, Biogen-Idec, TEVA, Merck-Serono, Novartis, Roche, Jansen Research, IXICO Ltd., GeNeuro, and Apitope Ltd. for consulting; payment from the Serono Symposia Foundation, IXICO Ltd., and MedScape for educational presentations; and research support via grants from EU/EFPIA Innovative Medicines Initiative Joint Undertaking (AMYPAD consortium), EuroPOND (H2020), U.K. MS Society,

Dutch MS Society, PICTURE (IMDI-NWO), and ECTRIMS-MAGNIMS. No other potential conflict of interest relevant to this article was reported.

## REFERENCES

1. Alzheimer's Association. 2016 Alzheimer's disease facts and figures. *Alzheimers Dement.* 2016;12:459–509.
2. Jansen WJ, Ossenkoppele R, Knol DL, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. *JAMA.* 2015;313:1924–1938.
3. Jansen WJ, Ossenkoppele R, Knol DL, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. *JAMA.* 2015;313:1924–1938.
4. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 2007;6:734–746.
5. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer Dement.* 2011;7:280–292.
6. Sperling RA, Rentz DM, Johnson KA, et al. The A4 study: stopping AD before symptoms begin? *Sci Transl Med.* 2014;6:228fs13.
7. Buckley CJ, Sherwin PF, Smith AP, Wolber J, Weick SM, Brooks DJ. Validation of an electronic image reader training programme for interpretation of [ $^{18}\text{F}$ ]flutemetamol beta-amyloid PET brain images. *Nucl Med Commun.* 2017;38:234–241.
8. Thurfjell L, Lilja J, Lundqvist R, et al. Automated quantification of  $^{18}\text{F}$ -flutemetamol PET activity for categorizing scans as negative or positive for brain amyloid: concordance with visual image reads. *J Nucl Med.* 2014;55:1623–1628.
9. Salloway S, Gamez JE, Singh U, et al. Performance of [ $^{18}\text{F}$ ]flutemetamol amyloid imaging against the neuritic plaque component of CERAD and the current (2012) NIA-AA recommendations for the neuropathologic diagnosis of Alzheimer's disease. *Alzheimers Dement (Amst).* 2017;9:25–34.
10. van Berckel BN, Ossenkoppele R, Tolboom N, et al. Longitudinal amyloid imaging using  $^{11}\text{C}$ -PiB: methodologic considerations. *J Nucl Med.* 2013;54:1570–1576.
11. Zwan MD, Ossenkoppele R, Tolboom N, et al. Comparison of simplified parametric methods for visual interpretation of  $^{11}\text{C}$ -Pittsburgh compound-B PET images. *J Nucl Med.* 2014;55:1305–1307.
12. Morris JC, Heyman A, Mohs RC, et al. The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology.* 1989;39:1159–1165.
13. de Jager CA, Budge MM, Clarke R. Utility of TICS-M for the assessment of cognitive function in older adults. *Int J Geriatr Psychiatry.* 2003;18:318–324.
14. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res.* 1982–1983;17:37–49.
15. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology.* 1993;43:2412–2414.
16. Nelissen N, Van Laere K, Thurfjell L, et al. Phase 1 study of the Pittsburgh compound B derivative  $^{18}\text{F}$ -flutemetamol in healthy volunteers and patients with probable Alzheimer disease. *J Nucl Med.* 2009;50:1251–1259.
17. Heeman F, Yaqub M, Heurling K, et al. P20: optimized coffee-break protocol for quantitative [ $^{18}\text{F}$ ]flutemetamol studies. *Human Amyloid Imaging Conference.* 2018:83
18. Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *NeuroImage.* 1997;6:279–287.
19. Wu Y, Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. *J Cereb Blood Flow Metab.* 2002;22:1440–1452.
20. Boellaard R, Yaqub M, Lubberink M, Lammertsma A. PPET: a software tool for kinetic and parametric analyses of dynamic PET studies. *Neuroimage.* 2006;31:T62.
21. Svarer C, Madsen K, Hasselbalch SG, et al. MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. *Neuroimage.* 2005;24:969–979.
22. Hammers A, Allom R, Koeppe MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp.* 2003;19:224–247.
23. Tolboom N, Yaqub M, van der Flier WM, et al. Detection of Alzheimer pathology in vivo using both  $^{11}\text{C}$ -PiB and  $^{18}\text{F}$ -FDDNP PET. *J Nucl Med.* 2009;50:191–197.

24. Guo T, Dukart J, Brendel M, Rominger A, Grimmer T, Yakushev I. Rate of  $\beta$ -amyloid accumulation varies with baseline amyloid burden: implications for anti-amyloid drug trials. *Alzheimers Dement*. 2018;14:1387–1396.
25. Koedam EL, Lehmann M, van der Flier WM, et al. Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol*. 2011;21:2618–2625.
26. Scheltens P, Leys D, Barkhof F, et al. Atrophy of medial temporal lobes on MRI in probable Alzheimers-disease and normal aging: diagnostic-value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry*. 1992;55:967–972.
27. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5-T in Alzheimers dementia and normal aging. *Am J Neuroradiol*. 1987;8:421–426.
28. Lammertsma AA. Forward to the past: the case for quantitative PET imaging. *J Nucl Med*. 2017;58:1019–1024.
29. Vandenberghe R, Van Laere K, Ivanoiu A, et al.  $^{18}\text{F}$ -flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. *Ann Neurol*. 2010;68:319–329.
30. Carson RE, Channing MA, Blasberg RG, et al. Comparison of bolus and infusion methods for receptor quantitation: application to [ $^{18}\text{F}$ ]cyclofoxy and positron emission tomography. *J Cereb Blood Flow Metab*. 1993;13:24–42.
31. Mountz JM, Laymon CM, Cohen AD, et al. Comparison of qualitative and quantitative imaging characteristics of [ $^{11}\text{C}$ ]PiB and [ $^{18}\text{F}$ ]flutemetamol in normal control and Alzheimer's subjects. *Neuroimage Clin*. 2015;9:592–598.
32. Salloway S, Sperling R, Fox NC, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370:322–333.
33. Doody RS, Thomas RG, Farlow M, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med*. 2014;370:311–321.
34. Zhang N, Gordon ML, Goldberg TE. Cerebral blood flow measured by arterial spin labeling MRI at resting state in normal aging and Alzheimer's disease. *Neurosci Biobehav Rev*. 2017;72:168–175.
35. Sierra-Marcos A. Regional cerebral blood flow in mild cognitive impairment and Alzheimer's disease measured with arterial spin labeling magnetic resonance imaging. *Int J Alzheimers Dis*. 2017;2017:5479597.
36. Morris E, Chalkidou A, Hammers A, Peacock J, Summers J, Keevil S. Diagnostic accuracy of  $^{18}\text{F}$  amyloid PET tracers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging*. 2016;43:374–385.
37. Grothe MJ, Barthel H, Sepulcre J, et al. In vivo staging of regional amyloid deposition. *Neurology*. 2017;89:2031–2038.
38. Palmqvist S, Scholl M, Strandberg O, et al. Earliest accumulation of beta-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nat Commun*. 2017;8:1214.
39. Jansen W, Ossenkuppele R, Verhey FR, Visser PJ. The prevalence of amyloid pathology in healthy individuals and individuals with mild cognitive impairment: a meta-analysis of amyloid-PET studies [abstract]. *Alzheimer Dement*. 2013; 9(suppl):533.