

Use of Standardized Uptake Value Ratios Decreases Interreader Variability of [¹⁸F] Florbetapir PET Brain Scan Interpretation

A.P. Nayate, J.G. Dubroff, J.E. Schmitt, I. Nasrallah, R. Kishore, D. Mankoff, and D.A. Pryma;
for the Alzheimer's Disease Neuroimaging Initiative



ABSTRACT

BACKGROUND AND PURPOSE: Fluorine-18 florbetapir is a recently developed β -amyloid plaque positron-emission tomography imaging agent with high sensitivity, specificity, and accuracy in the detection of moderate-to-frequent cerebral cortical β -amyloid plaque. However, the FDA has expressed concerns about the consistency of interpretation of [¹⁸F] florbetapir PET brain scans. We hypothesized that incorporating automated cerebral-to-whole-cerebellar standardized uptake value ratios into [¹⁸F] florbetapir PET brain scan interpretation would reduce this interreader variability.

MATERIALS AND METHODS: This randomized, blinded-reader study used previously acquired [¹⁸F] florbetapir scans from 30 anonymized patients who were enrolled in the Alzheimer's Disease Neuroimaging Initiative 2. In 4 separate, blinded-reading sessions, 5 readers classified 30 cases as positive or negative for significant β -amyloid deposition either qualitatively alone or qualitatively with additional adjunct software that determined standardized uptake value ratios. A κ coefficient was used to calculate interreader agreement with and without the use of standardized uptake value ratios.

RESULTS: There was complete interreader agreement on 20/30 cases of [¹⁸F] florbetapir PET brain scans by using qualitative interpretation and on 27/30 scans interpreted with the adjunct use of standardized uptake value ratios. The κ coefficient for the studies read with standardized uptake value ratios (0.92) was significantly higher compared with the qualitatively read studies (0.69, $P = .006$).

CONCLUSIONS: Use of standardized uptake value ratios improves interreader agreement in the interpretation of [¹⁸F] florbetapir images.

ABBREVIATIONS: AD = Alzheimer disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; EAD = early AD; EMCI = early mild cognitive impairment; LMCI = late mild cognitive impairment; SUVr = standardized uptake value ratios

Estimated to affect approximately 5.4 million Americans and 30 million people worldwide, Alzheimer disease (AD) is the most common type of dementia.^{1,2} Cortical β -amyloid deposition is 1 pathologic hallmark² and is hypothesized to be pathogenic for

AD.^{2,3} The ability to noninvasively detect cortical β -amyloid deposition could considerably improve clinical diagnosis because >33% of patients with early signs of AD will be inaccurately identified.⁴ Fluorine-18 florbetapir, a recently developed β -amyloid plaque positron-emission tomography imaging agent, rapidly en-

Received May 16, 2014; accepted after revision January 12, 2015.

From the Division of Nuclear Medicine and Clinical Molecular Imaging, Department of Radiology, University of Pennsylvania, Philadelphia, Pennsylvania.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://www.adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

This study was supported in part by a grant from Siemens (Hoffman Estates, Illinois). Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica; Biogen Idec; Bristol-Myers Squibb; Eisai; Elan Pharmaceuticals; Eli Lilly and Company; F. Hoffmann-La Roche and its affiliated company Genentech; GE Healthcare; Innogenetics, N.V.; IXICO; Janssen Alzheimer Immunotherapy Research & Development; Johnson & Johnson Pharmaceutical Research &

Development; Medpace; Merck; Meso Scale Diagnostics; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer; Piramal Imaging; Servier; Synarc; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by National Institutes of Health grants P30 AG010129 and K01 AG030514.

Research findings previously presented at: Annual Meeting the Radiological Society of North America, December 1–6, 2013; Chicago, Illinois.

Please address correspondence to D.A. Pryma, MD, 3400 Spruce St, 110 Donner Building, Philadelphia, PA; e-mail: daniel.pryma@uphs.upenn.edu

Indicates open access to non-subscribers at www.ajnr.org

Indicates article with supplemental on-line table.

<http://dx.doi.org/10.3174/ajnr.A4281>

ters the brain and specifically binds to cortical fibrillar β -amyloid.^{5,6} Pathologic findings demonstrate the high sensitivity (87%), specificity (95%), and accuracy (90%) of [¹⁸F] florbetapir in the detection of moderate-to-frequent cortical β -amyloid plaque by using the Consortium to Establish a Registry for Alzheimer's Disease criteria.⁷

Sufficient concern about the consistency of [¹⁸F] florbetapir PET brain scan interpretation led the FDA to withhold approval until an interpretation training program was implemented to reduce interreader variability.⁸⁻¹⁰ Imaging and pathologic studies have previously demonstrated that cerebral cortical regions, including the frontal lobe, parietal lobe, temporal lobe, precuneus, and anterior and posterior cingulate gyrus, are regions in which β -amyloid deposition is commonly found in patients with AD.^{5,11} These findings motivated quantitative analysis of [¹⁸F] florbetapir PET brain images, comparing differential binding between these cortical regions to the whole cerebellum, a site not prone to amyloid deposition, expressed as cerebral-to-whole-cerebellar standardized uptake value ratios (SUVr).^{5,12} Subsequent pathologic analysis demonstrated high sensitivity (97%), specificity (100%), and accuracy (98%) between [¹⁸F] florbetapir PET SUVr and postmortem immunohistochemical measurements of β -amyloid.^{5,7,12} These studies showed that a mean [¹⁸F] florbetapir SUVr value of >1.17 was strongly associated with an intermediate-to-high likelihood of a neuropathologic diagnosis of AD.^{5,12} However, SUVr are onerous to calculate manually, and manual placement of ROIs is prone to variability. Computer assistance could provide an easier method to incorporate SUVr into the interpretation process,^{13,14} and semiautomated software has been developed to facilitate SUVr calculations. The current standardized methodology for amyloid PET brain interpretation does not use SUVr,¹⁴ which might be useful to improve reader agreement. We hypothesized that adding SUVr to qualitative image interpretation of [¹⁸F] florbetapir PET brain scans would reduce interreader variability.

MATERIALS AND METHODS

Participants

This randomized, blinded-reader study used previous [¹⁸F] florbetapir scans from 60 anonymized patients (37 men and 23 women) who were enrolled in the Alzheimer's Disease Neuroimaging Initiative (ADNI) 2 and had already provided written informed consent. Fluorine-18 florbetapir PET brain scans were obtained at multiple sites; however, all sites followed the same ADNI 2 protocol.¹⁵ Studies were randomly selected from the ADNI 2 population data base and anonymized. Patient age ranged from 55 to 94 years; each patient had an established clinical diagnosis of normal, early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), or early AD (EAD). We chose these groups (normal, EMCI, LMCI, and EAD) because their [¹⁸F] florbetapir PET studies were expected to be the most challenging to interpret and, therefore, could better test the potential benefit of using SUVr. In consultation with a biostatistician before the study, we determined that with 5 readers, approximately 30 cases would provide enough statistical power to test whether the addition of SUVr would provide a substantial improvement in reader consistency for our study. Due to the large ADNI data base

and the need to only include these groups, an uninvolved third party was used to select cases accordingly. The investigators played no role in choosing the cases, to maintain the blinded nature of the study. This study was approved by the local institutional review board.

Image Analysis and Reader Study

Five nuclear medicine board-certified or eligible physicians with no prior clinical experience in interpreting [¹⁸F] florbetapir PET brain scans (though some had experience with the studies in a research setting) and with varying years of clinical experience in nuclear medicine (3 readers had ≤ 3 years of experience, 1 reader had 8 years of experience, and 1 reader had 18 years of experience) underwent on-line [¹⁸F] florbetapir PET clinical training (<http://www.amyvidtraining.com>; Avid Radiopharmaceuticals, Philadelphia, Pennsylvania) before initiation of the study. The training included information about [¹⁸F] florbetapir, β -amyloid, and Alzheimer disease followed by demonstrations and self-assessment cases on study interpretation.

Fluorine-18 florbetapir PET brain scans of 60 participants were given 2 unique and random identifiers; each case was evaluated twice, once qualitatively and once with the inclusion of SUVr information. Each case was classified as either positive or negative for cortical β -amyloid deposition. All 60 cases were assigned to 4 reading sessions separated by at least 72 hours to avoid reader memory. All assignments were random except that no case was repeated during an individual session. For each case, the order of qualitative and SUVr-aided reads was also random.

We divided the 60 cases into 2 groups with the initial 30 cases to be used as a lead-in to give all readers a more similar experience in evaluating amyloid PET studies and using the SUVr software. The remaining 30 test cases were used for the test set. Readers never received feedback about their interpretations. Readers qualitatively interpreted the scans by using MIMfusion (MIM Software Inc, Cleveland Ohio) by determining the presence or absence of cortical β -amyloid deposition according to the clinical interpretation methodology. Axial, sagittal, and coronal images were presented, and the reader could manipulate image contrast to accentuate the gray-white interface as recommended by the training program.

SUVr cases were reviewed qualitatively on MIMfusion, and additionally SUVr were calculated automatically by using Scenium (Siemens, Erlangen, Germany). The SUVr software automatically registers [¹⁸F] florbetapir scans to the Montreal Neurological Institute atlas space and calculates the SUVr values without requiring any input from the user.¹⁶ However, the reader could review and manipulate the ROIs to fit the cerebral cortical gray matter and whole cerebellum, to assure proper anatomic registration. Readers recorded average cortical SUVr of these regions for each case along with the positive/negative interpretation. Because the SUVr values for all 5 readers were very similar, the manipulation of ROIs was at most minimal in all cases. Cortex-to-whole-cerebellum SUVr values were automatically calculated and presented by the software for 6 anatomically relevant cortical ROIs: frontal lobe, lateral parietal lobe, lateral temporal lobe, precuneus, and anterior and posterior cingulate gyrus, as well as the mean SUVr of these regions.¹² Prior studies demonstrated that with the aid of the average SUVr from these anatomic regions, sensitivity

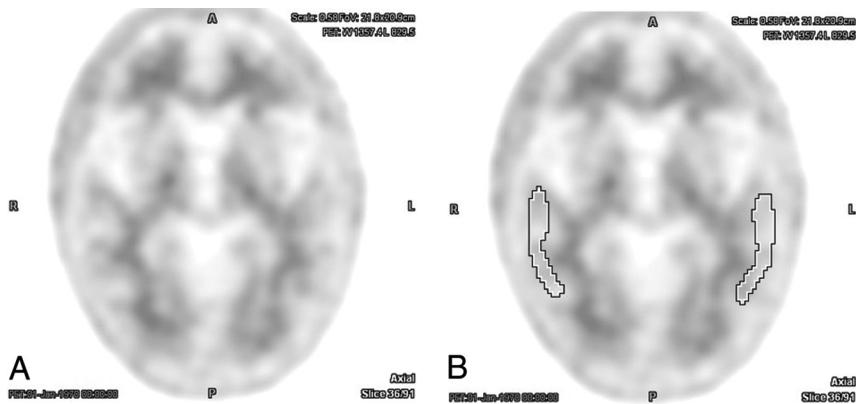


FIG 1. Example of an [¹⁸F] florbetapir PET brain scan with negative findings without (A) and with (B) ROIs centered over the bilateral temporal lobe cortex. There was perfect interreader agreement both with and without the aid of semiquantitative indices (average interreader SUVr = 0.97 ± 0.02).

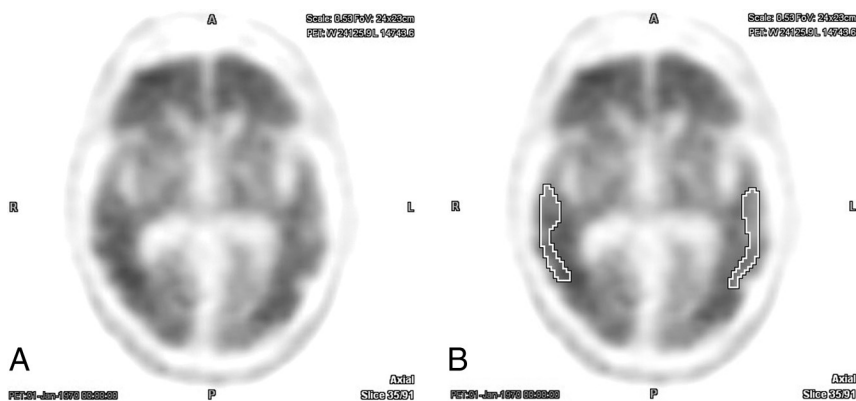


FIG 2. Example of an [¹⁸F] florbetapir PET brain scan with positive findings without (A) and with (B) ROIs centered over the bilateral temporal lobe cortex. There was perfect interreader agreement both with and without the aid of semiquantitative indices (average interreader SUVr = 1.78 ± 0.02).

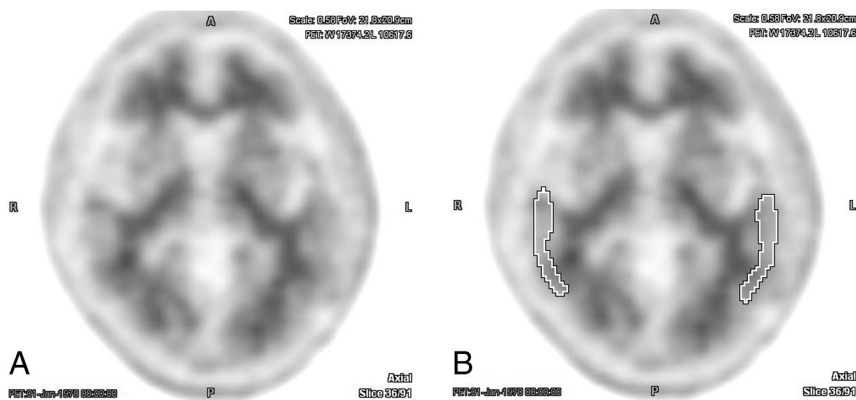


FIG 3. Example of a [¹⁸F] florbetapir PET brain scan without complete interreader agreement, displayed without (A) and with (B) an ROI over the temporal lobes. In this case, there was rater discordance both for qualitative interpretation and with the aid of semiquantitative indices (average interreader SUVr = 1.27 ± 0.05).

and specificity of reads increased by using both clinical diagnosis¹⁷ and pathology⁷ as comparisons. Although we chose to use the Scenium software to calculate SUVr, commercial programs from other vendors are available for calculating SUVr.^{5,12}

A recent study¹⁸ showed a high correlation between SUVr calculated by Scenium and other methods.^{5,7,19} Prior research suggested that a threshold for amyloid positivity was at SUVr ≥ 1.17.^{5,12} All readers were informed of this threshold. However, the SUVr value was available to the reader as an adjunct to assist in the primary qualitative interpretation. Therefore, the final interpretation relied on the reader's overall judgment, incorporating both the qualitative image data and SUVr. Sample cases are shown in Figs 1–3.

Statistical Analyses

The data were analyzed by using the R Statistical Computing Environment (<http://www.r-project.org/>).²⁰ To assess interrater reliability, we calculated the Fleiss multirater κ statistics for the 2 conditions separately (qualitative versus qualitative + SUVr) by using the “irr” package.²¹ Confidence intervals for each κ were calculated separately via bootstrap by using 1000 replicates. Statistical comparison was accomplished by the method for comparing correlated κ statistics described by Vanbelle and Albert.²² Briefly, for 2 correlated κ values, a difference score can be calculated such that

$$1) \quad \hat{\kappa}_D = \hat{\kappa}_{SUVr} - \hat{\kappa}_{qualitative}$$

The distribution of this statistic can be estimated via bootstrap by calculating the difference score for q subsamples with replacement. A new estimator is then calculated that under the null hypothesis ($\hat{\kappa}_{qualitative} = \hat{\kappa}_{SUVr}$) follows a t distribution with $q-1$ df :

$$2) \quad t_{\kappa_D} = \frac{\overline{\hat{\kappa}_D}}{SD(\hat{\kappa}_D)}$$

For the current study, a bootstrap analysis by using 1000 replicates was used for hypothesis testing. An α of .05 was set as the threshold for statistical significance.

For the interpretation of κ values, we used the magnitude guidelines published by Landis and Koch²³: Values of $\kappa < 0$ indicate no agreement; $\kappa = 0-0.20$, slight; $\kappa = 0.21-0.40$, fair; $\kappa = 0.41-0.60$, moderate; $\kappa = 0.61-0.80$,

excellent; and $\kappa = 0.81-1$, near-perfect agreement. Average interreader measurements of SUVr for the 4 different clinical diagnostic groups (normal, EMCI, LMCI, and EAD) were compared by using 1-way ANOVA with post hoc Tukey-

Kramer tests. We did not measure intrareader data per recommendations of the annual Clinical Trials Methodology Workshop of the Radiological Society of North America (https://www.rsna.org/Clinical_Trials_Methodology_Workshop.aspx).

RESULTS

One of the input images could not be successfully registered to the template for quantification due to nonstandard patient positioning. Most important, no disease process (eg, normal pressure hydrocephalus) or anatomic variant precluded this case being aligned to the template, and no similar issue arose with the other 59 cases. In this case, the readers could do only qualitative interpretation for both the qualitative and qualitative + SUVr reads. There was complete interreader agreement on 20 of 30 cases of [¹⁸F] PET brain scans by using qualitative-only interpretation and 27 of 30 scans interpreted with the adjunct use of the SUVr (On-line Table). Quantitative measures of interrater reliability confirmed excellent agreement between raters when using qualitative analysis alone ($\kappa = 0.69$; 95% CI, 0.50–0.82). The addition of SUVr data resulted in near-perfect agreement ($\kappa = 0.92$; 95% CI, 0.79–0.97). Interrater agreement was significantly increased with the addition of SUVr data ($t = 2.51, P = .006$) after adjusting for the correlated nature of the data.

Group differences in global SUVr were statistically significant ($P < .005$) with group means and SDs as follows: normal (1.05 ± 0.21), EMCI (1.27 ± 0.25), LMCI (1.32 ± 0.25), and EAD (1.65 ± 0.20). Post hoc Tukey-Kramer tests demonstrated significant mean differences between all pair-wise group combinations ($P < .0002$), with the exception of EMCI versus LMCI (not significant).

In the original 30 practice cases used as a lead-in, there was complete interreader agreement on 17 of 30 cases of [¹⁸F] florbetapir PET scans by using qualitative-only interpretation ($\kappa = 0.56$; 95% CI, 0.29–0.7) and 16 of 30 scans interpreted with the adjunct use of the SUVr ($\kappa = 0.55$; 95% CI, 0.36–0.73). These κ values were significantly different ($P < .05$) than the κ values achieved in our dataset of 30 cases. If all the data from the 60 cases were included for all 5 readers, then there was complete interreader agreement on 37/60 cases ($\kappa = 0.62$) only qualitatively evaluated and complete interreader agreement on 43/60 cases ($\kappa = 0.74$) when evaluated with the aid of SUVr. The κ values of all 60 cases evaluated with the aid of SUVr were higher than those of the same cases when only qualitatively evaluated; this approached but did not achieve statistical significance ($t = 1.34, P = .09$).

DISCUSSION

AD is the most common dementia to affect the elderly and traditionally has been diagnosed clinically.^{1,2} However, 10%–20% of patients clinically diagnosed with AD lack pathologic findings at postmortem examination.²⁴ Improved diagnosis could aid in medical and personal decision-making. Furthermore, on the basis of the suspected role of β -amyloid in the pathophysiology of AD, it has emerged as a potential drug target. In evaluating and potentially using such therapies, reliably establishing the presence of β -amyloid deposition would be of paramount importance. Recently, the Society of Nuclear Medicine and Molecular Imaging and the Alzheimer's Association proposed PET amyloid imaging

appropriate-use guidelines for patients who meet specific criteria.²⁵ For the test to be of greatest clinical utility, interreader variability needs to be minimized. We hypothesized that incorporating a method of quantification to standard image interpretation of [¹⁸F] florbetapir PET brain scans by the addition of SUVr would reduce interreader variability.

Our results show a significantly higher ($P = .006$) interreader agreement when [¹⁸F] florbetapir PET scans were evaluated with adjunctive SUVr data ($\kappa = 0.92$) compared with qualitative-only assessment of the same studies ($\kappa = 0.69$). This κ value of 0.69 is similar to the κ value in another study with 5 readers visually assessing 59 cases ($\kappa = 0.76$) and in a study with 2 readers visually assessing 46 cases ($\kappa = 0.71$).¹⁷ The values of our study are also similar to those seen with other β -amyloid imaging agents such as ¹¹C Pittsburgh compound-B.²⁶ The impact of the data of SUVr on κ values reported here is also higher than those seen in interobserver variability studies in lesion detection in other organ systems, for example, in the detection of pulmonary nodules with ($\kappa = 0.67$)²⁷ and without the use of computer-assisted detection software ($\kappa = 0.64$)²⁸ and the detection of breast lesions by using automatic breast scanners ($\kappa = 0.8$).²⁹

When [¹⁸F] florbetapir PET brain studies were read qualitatively, there was interreader disagreement in 9/30 (30%) cases; however, there was complete agreement between readers for 8 of these cases when independently evaluated with semiquantitative indices. With the adjunct use of the SUVr, there was interreader disagreement on 3/30 (10%) cases. One case was discrepant (interreader disagreement) on both qualitative and SUVr reads (the case in which the software failed to register the study to the atlas). In one of the remaining 2 discrepant cases, there was interreader disagreement on both the qualitative assessment and the assessment with the aid of SUVr; and on the other case, 1 reader interpreted the study differently than the others only on assessment with the aid of SUVr. The interreader average SUVr on these cases were 1.27 (range, 1.24–1.35) and 1.15 (range, 1.13–1.21) and were the closest to the threshold value of 1.17^{5,7,12} of the test cases (Fig 4). Therefore, we hypothesized that interreader disagreement on these 2 cases was probably a result of visually borderline scans because both of these subjects had diagnoses of early mild cognitive impairment, which often has intermediate values of SUVr.^{5,7,12}

Although we did not directly assess the extent of ROI manipulation by the readers, the manipulation by all readers was, at most, minimal as shown by the small SD between interreader average SUVr ($\leq 5\%$, On-line Table). These findings are concordant with a prior study that demonstrated minimal variance in the interrater reliability of manual and automated ROI delineation for Pittsburgh compound-B PET.³⁰ There was no certain relationship between the years of post-board certification and discrepancies, compared with most interpretations of individual cases (Table).

Our results also show a significant difference ($P < .0002$) in the values of SUVr between normal controls and patients with early mild and late mild cognitive impairment or EAD, with progressively higher SUVr values seen in patients with early mild and late mild cognitive impairment or EAD compared with normal patients, concordant with findings of prior studies.^{4,14,19} Com-

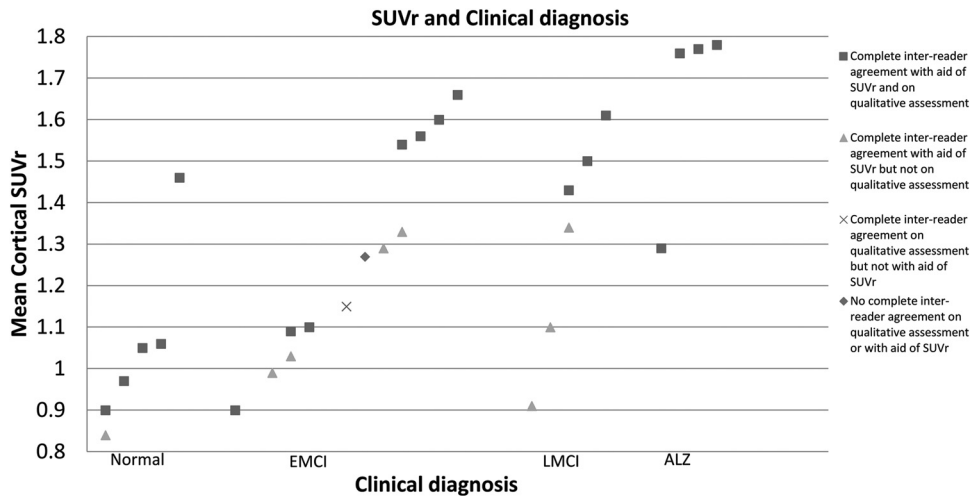


FIG 4. SUVr values for 29/30 test cases. No SUVr could be determined for one of the cases because input images could not be successfully registered to the template for quantification due to nonstandard patient positioning.

Comparison of years of experience in diagnostic imaging and discrepant reads (versus majority consensus) in 30 test cases

| Years of Experience (after Radiology or Nuclear Medicine Residency) | Discrepant Reads on Qualitative Assessment | Discrepant Reads with Aid of SUVr |
|---|--|-----------------------------------|
| 2 | 3 | 0 |
| 3 | 6 ^a | 2 ^a |
| 3 | 1 | 1 |
| 8 | 3 | 0 |
| 18 | 1 | 0 |

^a In 1 case, only qualitative interpretation for both the qualitative and qualitative + SUVr reads could be provided because input images could not be successfully registered to the template for quantification due to nonstandard patient positioning.

parison of SUVr between the EMCI and LMCI groups did not demonstrate a significant difference.

Our findings demonstrate that cases that had complete inter-reader agreement as positive for significant β -amyloid deposition with the aid of semiquantitative analysis but not complete agreement when qualitatively assessed had an interreader mean SUVr value of 1.32 ± 0.0 ($n = 3$). Cases that had complete agreement for positive scan findings by both methods had an interreader mean SUVr of 1.58 ± 0.15 ($n = 12$). These findings suggest that cases with values of SUVr closer to the cutoff value of 1.17 are often visually challenging and can therefore contribute to discrepant reads with qualitative assessment. However, a similar divergence in the mean SUVr was not seen in cases with discrepant interpretations by qualitative evaluation that were uniformly interpreted as negative with the aid of SUVr (interreader mean SUVr = 0.97 ± 0.1 ; $n = 5$) compared with cases that had complete interreader agreement for negative scan findings by both methods (interreader mean SUVr = 1.01 ± 0.09 ; $n = 7$). This finding is congruent with findings from prior studies^{5,12} and may be explained by the decreased [¹⁸F] florbetapir uptake seen in patients without significant β -amyloid cerebral cortical deposition and therefore creating a narrower range for SUVr values (Fig 4).

We designed our experiment to include 30 practice cases because all 5 readers had no prior clinical experience in interpreting [¹⁸F] florbetapir PET brain scans (which had just been approved

at the time of this study) and had varying amounts of research experience and varying years of experience in nuclear medicine. Therefore, we thought that the practice cases would help readers gain similar familiarity with the software and method. In the original 30 practice cases used as a lead-in, there was moderate inter-reader agreement between qualitative-only interpretation ($\kappa = 0.56$) and with the adjunct use of the SUVr ($\kappa = 0.55$). These κ values were significantly different ($P < .05$) than the κ values achieved in our dataset of 30 cases. This discrepancy is likely the result of inexperience and varying early proficiency in using the software. Improved agreement between readers was seen in our test set of the 30 cases for qualitative-only interpretation, emphasizing the importance of practice on physician performance. Improved agreement between readers on the test set compared with the training set when using SUVr signifies the need for familiarization with image analysis software and is congruent with findings seen in using other computer-aided diagnostic software such as in the detection of lesions on mammograms.³¹ Finally, the κ values of all 60 cases evaluated with the aid of SUVr were higher than those of the same cases when only qualitatively evaluated, showing a trend toward statistical significance ($P = .09$).

We could not determine the accuracy of interpretations in this study because no criterion standard pathologic data were available for the cohort. Clinical diagnosis is not the criterion standard for diagnosing Alzheimer disease³² and can be an unreliable diagnostic tool with 10%–20% of patients diagnosed with Alzheimer disease lacking pathology on postmortem histopathologic analysis²⁴; clinical diagnosis has intermediate sensitivity (84%) and low specificity (52.5%) in diagnosing Alzheimer disease.³³ Our findings show that 52% of mild cognitive impairment (EMCI + LMCI) cases were interpreted with complete interreader agreement as β -amyloid-positive with the aid of SUVr, while 37% of these mild cognitive impairment cases were interpreted with complete interreader agreement as β -amyloid-positive with qualitative-only assessment. Fourteen percent of cognitively normal cases were interpreted with complete interreader agreement as β -amyloid-positive, both with only qualitative assessment and with the aid of SUVr (On-line Table). These findings are congruent

ent with prior imaging studies¹³ and are within range of cortical β -amyloid deposition seen in postmortem case studies in cognitively normal, mildly cognitively impaired, and patients with Alzheimer disease.³³⁻³⁵ Our findings also demonstrate that the relationship between cognitive decline and the amount of cortical β -amyloid deposition is variable because we see studies with positive findings in all of our experimental groups (normal, EMCI, LMCI, and AD) and they are compatible with prior pathologic studies.³⁶ The degree of cortical β -amyloid deposition could have prognostic importance because recent studies have demonstrated that cognitively normal, mild cognitively impaired, and subjects with Alzheimer disease who have PET scans positive for amyloid demonstrate greater cognitive and global deterioration during an 18-month³⁷ and 3-year³⁸ follow-up than subjects with scans with negative findings.

We did not use the cutoff value of SUV_r of 1.17 solely as a method to quantitatively interpret the scans. The study that determined this value had a small sample size of 19 patients, and studies in larger community-based samples with a broader distribution of SUV_r would be needed to more definitively establish standard thresholds. Furthermore, the applicability of this value obtained from whole-brain cortical uptake to use in regional values is unknown.¹² In our study, the readers were aware of this empiric threshold and could choose to use it when evaluating the calculated regional and averaged regional SUV_r value from [¹⁸F] florbetapir PET images.

We did not measure the intraobserver variability because we wanted to emulate the reading sessions as a standard clinical practice in which studies are assigned a single interpretation by an individual physician reader. In this regard, our methodology was similar to that in other studies examining interreader performance in diagnostic imaging with and without an intervention such as computer-aided diagnostic software. Parallel methodology has been used in lung disease,³⁹⁻⁴³ breast imaging,^{29,44-46} and Alzheimer disease,^{26,47} without determining intraobserver variability.

The most important limitation of our study is the absence of the criterion standard of pathologic analysis to establish the presence or absence of cerebral cortical β -amyloid deposition in our cohort of subjects to determine the accuracy of physician interpretations of the [¹⁸F] florbetapir PET studies. As such, we could not determine whether the use of SUV_r improved diagnostic accuracy; our primary aim was to assess interreader variability. Prior studies have demonstrated a high correlation between PET SUV_r and immunohistochemical measurements of β -amyloid,^{5,7,12} suggesting that improving agreement will likely improve diagnostic accuracy. However, while high interreader agreement is desirable in diagnostic testing, it will be important to directly evaluate the effect on diagnostic accuracy in future studies, especially in patients with mild cognitive impairment because prior studies have primarily focused on determining the accuracy of amyloid PET in cognitively normal individuals or patients with probable Alzheimer disease.^{5,7} Second, although it was meant as an adjunct tool, we did not determine or prescribe the degree to which readers used the SUV_r values in determining their interpretations of scans. Third, due to differences in reader ROI manipulation, there was minimal variance in average SUV_r values on the test cases; therefore, this minimal variance is a potential weakness of a semi-

automated method. Fourth, 1 case from our 30 test cases did not successfully register to the template for quantification due to non-standard patient positioning. Therefore, registration errors and other technical failures of the software are an additional potential weakness of such a semiautomated method.

CONCLUSIONS

Our results support the use of SUV_r to improve interreader agreement in the interpretation of [¹⁸F] florbetapir images. Furthermore, using computer software to obtain values of SUV_r can be an appealing and efficient option for nuclear medicine physicians and radiologists in interpreting [¹⁸F] florbetapir PET brain scans and other brain imaging agents. The promising results from this initial study support future larger and prospective studies, including determining the performance of semiquantification strategies for [¹⁸F] florbetapir and other β -amyloid radiopharmaceuticals to establish ranges for negative and positive, compared against clinical and histopathologic reference standards.

ACKNOWLEDGMENTS

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative data base (adni.loni.ucla.edu), a \$60 million public-private partnership launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations. The primary goal of ADNI has been to test whether serial brain imaging studies, clinical and neuropsychological assessments, and other biologic markers can be combined to measure the progression of mild cognitive impairment and early AD to aid researchers and clinicians in developing new treatments and monitoring their effectiveness. ADNI, led by Principal Investigator Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco, is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from >50 sites across the United States and Canada. For up-to-date information and descriptions of PET imaging acquisition parameters, see www.adni-info.org.

Disclosures: Ameya P. Nayate—*RELATED: Grant:* This study was partially funded by Siemens.* Jacob G. Dubroff—*RELATED: Grant:* Siemens provided the software (syngo Scenium) to calculate the SUV_r and partial funding*; *Payment for Lectures (including service on Speakers Bureaus):* Webinar for Applied Radiology, *Comments:* Twenty-one months ago (August 2012), a \$1000 one-time honorarium was paid to me for giving a Webinar for Applied Radiology; my institution and department were/are aware of the relationship. James E. Schmitt—*UNRELATED: Grants/Grants Pending:* National Institutes of Health, Radiological Society of North America, *Comments:* T32 Training Grant from the National Institutes of Health (EB004311, to Mitch Schnall), Fellow Grant from the Radiological Society of North America (Imaging Genomic Analysis of the 22q11DS Syndrome). Rekha Kishore—*RELATED: Grant:* This study was partially funded by Siemens*; *UNRELATED: Stock/Stock Options:* Merck, *Comments:* I own a small amount of stock in Merck. David Mankoff—*RELATED: Grant:* Siemens (partial support)*; *UNRELATED: Consultancy:* GE Healthcare, Philips Healthcare, Siemens; *Grants/Grants Pending:* Siemens,* Philips Healthcare.* Daniel A. Pryma—*RELATED: Grant:* This work was partially funded by a grant from Siemens*; *Payment for Lectures (including service on Speakers Bureaus):* IBA Molecular. *Money paid to the institution.

REFERENCES

1. Alzheimer's Association. 2013 Alzheimer's disease facts and figures. *Alzheimers Dement* 2013;9:208–45

2. Holtzman DM, Morris JC, Goate AM. **Alzheimer's disease: the challenge of the second century.** *Sci Transl Med* 2011;3:77sr1
3. Hiltunen M, van Groen T, Jolkonen J. **Functional roles of amyloid-beta protein precursor and amyloid-beta peptides: evidence from experimental studies.** *J Alzheimers Dis* 2009;18:401–12
4. Löppönen M, Raiha I, Isoaho R, et al. **Diagnosing cognitive impairment and dementia in primary health care: a more active approach is needed.** *Age Ageing* 2003;32:606–12
5. Clark CM, Schneider JA, Bedell BJ, et al. **Use of florbetapir-PET for imaging beta-amyloid pathology.** *JAMA* 2011;305:275–83
6. Wong DF, Rosenberg PB, Zhou Y, et al. **In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (florbetapir [corrected] F 18).** *J Nucl Med* 2010;51:913–20
7. Clark CM, Pontecorvo MJ, Beach TG, et al. **Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study.** *Lancet Neurol* 2012;11:669–78
8. FDA advisory committee briefing document: Peripheral and Central Nervous Systems Drugs Advisory Committee—draft clinical review. December 20, 2010. <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/PeripheralandCentralNervousSystemDrugsAdvisoryCommittee/UCM240265.pdf>. Accessed September 12, 2013
9. FDA approves imaging drug Amyvid. <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm299678.htm>. Accessed September 14, 2013
10. **FDA approves 18F-florbetapir PET agent.** *J Nucl Med* 2012;53:15N
11. Braak H, Braak E. **Neuropathological staging of Alzheimer-related changes.** *Acta Neuropathol* 1991;82:239–59
12. Fleisher AS, Chen K, Liu X, et al. **Using positron emission tomography and florbetapir F18 to image cortical amyloid in patients with mild cognitive impairment or dementia due to Alzheimer disease.** *Arch Neurol* 2011;68:1404–11
13. Johnson KA, Sperling RA, Gidicsin CM, et al. **Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging.** *Alzheimers Dement* 2013;9:S72–83
14. Avid Radiopharmaceuticals. Amyvid Reader Training. 2012. <http://amyvidtraining.com/>. Accessed June 14, 2013
15. Petersen R, Weiner M. ADNI 2 Procedures Manual. 2008. <http://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf>. Accessed June 14, 2013
16. Ghosh P, Sorger P, Buelau A, et al. **White paper: quantitative software evaluation of amyloid brain PET imaging in dementia.** 2012. http://www.healthcare.siemens.com/siemens_hwem-hwem_sxxa_websites-context-root/wcm/idc/groups/public/@global/@imaging/@molecular/documents/download/mdaw/mj11/~edis/whitepaper15_amyloiddementia-00274875.pdf. Accessed October 15, 2013
17. Camus V, Payoux P, Barre L, et al. **Using PET with 18F-AV-45 (florbetapir) to quantify brain amyloid load in a clinical environment.** *Eur J Nucl Med Mol Imaging* 2012;39:621–31
18. Peyrat J, Joshi AD, Mintun MA, et al. **An automatic method for the quantification of uptake with florbetapir imaging.** *J Nucl Med* 2012; 53(suppl 1):210
19. Joshi AD, Pontecorvo MJ, Clark CM, et al. **Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects.** *J Nucl Med* 2012;53:378–84
20. R Development Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing; 2011
21. Gamer M, Lemon J, Fellows I, et al. The irr package, Ver. 0.83: various coefficients of interrater reliability and agreement. 2010. <http://cran.r-project.org/web/packages/irr/index.html>. Accessed April 1, 2013
22. Vanbelle S, Albert A. **A bootstrap method for comparing correlated kappa coefficients.** *Journal of Statistical Computing and Simulation* 2008;78:1009–15
23. Landis JR, Koch GG. **The measurement of observer agreement for categorical data.** *Biometrics* 1977;33:159–74
24. Ranginwala NA, Hynan LS, Weiner MF, et al. **Clinical criteria for the diagnosis of Alzheimer disease: still good after all these years.** *Am J Geriatr Psychiatry* 2008;16:384–88
25. Johnson KA, Minoshima S, Bohnen NI, et al. **Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association.** *J Nucl Med* 2013;54:476–90
26. Ng S, Villemagne VL, Berlangieri S, et al. **Visual assessment versus quantitative assessment of 11C-PIB PET and 18F-FDG PET for detection of Alzheimer's disease.** *J Nucl Med* 2007;48:547–52
27. Jeon KN, Goo JM, Lee CH, et al. **Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening computed tomography.** *Invest Radiol* 2012;47:457–61
28. Gierada DS, Pilgram TK, Ford M, et al. **Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening.** *Radiology* 2008;246:265–72
29. Wojcinski S, Gyapong S, Farrok A, et al. **Diagnostic performance and inter-observer concordance in lesion detection with the automated breast volume scanner (ABVS).** *BMC Med Imaging* 2013; 13:36
30. Rosario BL, Weissfeld LA, Laymon CM, et al. **Inter-rater reliability of manual and automated region-of-interest delineation for PiB PET.** *Neuroimage* 2011;55:933–41
31. Luo P, Qian W, Romilly P. **CAD-aided mammogram training.** *Acad Radiol* 2005;12:1039–48
32. Beach TG, Monsell SE, Phillips LE, et al. **Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010.** *J Neuropathol Exp Neurol* 2012;71:266–73
33. Lim A, Tsuang D, Kukull W, et al. **Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series.** *J Am Geriatr Soc* 1999;47:564–69
34. Bennett DA, Schneider JA, Bienias JL, et al. **Mild cognitive impairment is related to Alzheimer disease pathology and cerebral infarctions.** *Neurology* 2005;64:834–41
35. Gearing M, Schneider JA, Rebeck GW, et al. **Alzheimer's disease with and without coexisting Parkinson's disease changes: apolipoprotein E genotype and neuropathologic correlates.** *Neurology* 1995;45: 1985–90
36. Morishima-Kawashima M, Oshima N, Ogata H, et al. **Effect of apolipoprotein E allele epsilon4 on the initial phase of amyloid beta-protein accumulation in the human brain.** *Am J Pathol* 2000; 157:2093–99
37. Doraiswamy PM, Sperling RA, Coleman RE, et al. **Amyloid-beta assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study.** *Neurology* 2012;79:1636–44
38. Doraiswamy PM, Sperling RA, Johnson K, et al. **Florbetapir F 18 amyloid PET and 36-month cognitive decline: a prospective multicenter study.** *Mol Psychiatry* 2014;19:1044–51
39. Kitazono MT, Lau CT, Parada AN, et al. **Differentiation of pleural effusions from parenchymal opacities: accuracy of bedside chest radiography.** *AJR Am J Roentgenol* 2010;194:407–12
40. Muiyoyeta M, Maduskar P, Moyo M, et al. **The sensitivity and specificity of using a computer aided diagnosis program for automatically scoring chest X-rays of presumptive TB patients compared with Xpert MTB/RIF in Lusaka Zambia.** *PLoS One* 2014;9:e93757
41. Owrangi AM, Entwistle B, Lu A, et al. **Semi-automated scoring of pulmonary emphysema from X-ray CT: trainee reproducibility and accuracy.** *Eur J Radiol* 2013;82:e734–41
42. Pinto LM, Dheda K, Theron G, et al. **Development of a simple reliable radiographic scoring system to aid the diagnosis of pulmonary tuberculosis.** *PLoS One* 2013;8:e54235
43. Xie X, Zhao Y, Snijder RA, et al. **Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detec-**

- tor CT: an anthropomorphic phantom study.** *Eur Radiol* 2013;23:139–47
44. Carton AK, Gavenonis SC, Currivan JA, et al. **Dual-energy contrast-enhanced digital breast tomosynthesis: a feasibility study.** *Br J Radiol* 2010;83:344–50
45. Tagliafico A, Tagliafico G, Tosto S, et al. **Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one.** *Breast* 2009;18:35–40
46. Takeda K, Kanao S, Okada T, et al. **Assessment of CAD-generated tumor volumes measured using MRI in breast cancers before and after neoadjuvant chemotherapy.** *Eur J Radiol* 2012;81:2627–31
47. 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;38:1702–14