Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 9 (2017) 25-34

Alzheimer's & Dementia

Neuroimaging

Performance of [^{18}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

Stephen Salloway[a,b,*], Jose E. Gamez[c], Upinder Singh[d], Carl H. Sadowsky[e], Teresa Villena[f], Marwan N. Sabbagh[g], Thomas G. Beach[h], Ranjan Duara[i], Adam S. Fleisher[h,§], Kirk A. Frey[j], Zuzana Walker[k], Arvinder Hunjan[l], Yavir M. Escovar[m], Marc E. Agronin[n,o], Joel Ross[p], Andrea Bozoki[q], Mary Akinola[r], Jiong Shi[g], Rik Vandenberghe[s], Milos D. Ikonomovic[t], Paul F. Sherwin[u], Gill Farrar[v], Adrian P. L. Smith[v], Christopher J. Buckley[v], Dietmar Rudolf Thal[s], Michelle Zanette[u], Craig Curtis[w]

[a]*Neurology and the Memory and Aging Program, Butler Hospital, Warren Alpert Medical School, Brown University, Providence, RI, USA*
[b]*Department of Neurology and Psychiatry, Warren Alpert Medical School, Brown University, Providence, RI, USA*
[c]*Galiz Research, LLC, Miami Springs, FL, USA*
[d]*Geriatric Solutions, LLC, Las Vegas, NV, USA*
[e]*Division of Neurology, Nova SE University, Fort Lauderdale, FL, USA*
[f]*Premiere Research Institute, West Palm Beach, FL, USA*
[g]*Division of Neurology, Barrow Neurological Institute, Phoenix, AZ, USA*
[h]*Banner Alzheimer's Institute, Phoenix, AZ, USA*
[i]*Mount Sinai Medical Center, Wien Center for Alzheimer's Disease and Memory Disorders, Miami Beach, FL, USA*
[j]*Department of Radiology (Nuclear Medicine), University of Michigan, Ann Arbor, MI, USA*
[k]*Division of Psychiatry, University College London and North Essex Partnership University NHS Foundation Trust, London, UK*
[l]*Hertfordshire Partnership University NHS Foundation Trust, Essex, UK*
[m]*VERITAS Research, Miami Lakes, FL, USA*
[n]*Mental Health and Clinical Research, Miami Jewish Health Systems, Miami, FL, USA*
[o]*University of Miami Miller School of Medicine, Miami, FL, USA*
[p]*Memory Enhancement Center, Eatontown, NJ, USA*
[q]*Department of Neurology, Cognitive and Geriatric Neurology Team, Michigan State University, East Lansing, MI, USA*
[r]*Kennington Health Centre, Oxford, UK*
[s]*Department of Neurology, University Hospitals Leuven, Leuven, Belgium*
[t]*Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA*
[u]*Life Sciences, GE Healthcare, Marlborough, MA, USA*
[v]*Life Sciences, GE Healthcare, Amersham, Buckinghamshire, UK*
[w]*Compass Research, Orlando, FL, USA*

[§]Current address: Eli Lilly and Company, Indianapolis, IN, USA.
*Corresponding author. Tel.: +44-401-455-6403; Fax: +44-401-455-6405.
E-mail address: SSalloway@Butler.org

http://dx.doi.org/10.1016/j.dadm.2017.06.001

Abstract

Introduction: Performance of the amyloid tracer [^{18}F]flutemetamol was evaluated against three pathology standard of truth (SoT) measures including neuritic plaques (CERAD "original" and "modified" and the amyloid component of the 2012 NIA-AA guidelines).

Methods: After [^{18}F]flutemetamol imaging, 106 end-of-life patients who died underwent postmortem brain examination for amyloid plaque load. Blinded positron emission tomography scan interpretations by five independent electronically trained readers were compared with pathology measures.

Results: By SoT, sensitivity and specificity of majority image interpretations were, respectively, 91.9% and 87.5% with "original CERAD," 90.8% and 90.0% with "modified CERAD," and 85.7% and 100% with the 2012 NIA-AA criteria.

Discussion: The high accuracy of either CERAD criteria suggests that [^{18}F]flutemetamol predominantly reflects neuritic amyloid plaque density. However, the use of CERAD criteria as the SoT can result in some false-positive results because of the presence of diffuse plaques, which are accounted for when the positron emission tomography read is compared with the 2012 NIA-AA criteria.

Keywords: Diagnostic; Sensitivity; Specificity; Alzheimer's disease; Amyloid PET; Autopsy; [^{18}F]Flutemetamol; Thal phasing

1. Introduction

The presence of brain amyloid-beta (Aβ) plaques is one of several pathologic hallmarks of Alzheimer's disease (AD) [1]. Historically, Aβ deposits could only be detected by histopathologic analysis of brain tissue obtained by brain biopsy or at autopsy, using Aβ antibodies [2] or histochemical dyes with high affinity for fibrillar structure of Aβ [3], and these methods remain the standard of truth (SoT) for clinical studies. The development of positron emission tomography (PET) imaging agents, which detect fibrillar Aβ in vivo [4–10], has improved *in-life* diagnostic accuracy. Three PET amyloid imaging agents (Neuraceq [florbetaben [11,12], Piramal], Amyvid [florbetapir [12,13], Eli Lilly], and Vizamyl [flutemetamol [14,15], GE Healthcare]) are approved in the USA, Europe, and Japan. In all these pivotal phase 3 studies image read performance of the tracer compared with measurement of neuritic amyloid (via Bielschowsky silver stain) was the SoT. [^{18}F]florbetapir reported 59 brains analyzed with a sensitivity and specificity of 92% and 100%, respectively [5]. A total of 74 autopsy cases were available for [^{18}F]florbetaben where Sabri et al. [16] reported the same metrics to be 98% and 89%.

[^{18}F]Flutemetamol was previously shown to selectively target amyloid in vivo with high specificity [7,17]. Curtis et al. [9] reported that blinded visual interpretations of [^{18}F]flutemetamol PET images in 68 end-of-life patients had high sensitivity and specificity for estimating neuritic plaque density.

This study expands the Curtis et al. study. Since that study's termination, 38 additional autopsy brains became available from that PET imaged cohort. New guidelines for the neuropathologic evaluation of AD were also released [18] and include Thal phasing [19,20] based on measurements of total (neuritic plus diffuse) Aβ plaque load. The Curtis et al. data contained two false-positive images suspected to be related to diffuse Aβ plaques. These data were reanalyzed including the additional 38 brains, and Thal phasing was used to help establish the cause of the apparent false-positive scans. To test the robustness of the SoT, variations in the neuritic amyloid plaque component of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria used in the Curtis et al. study were also explored. This included validation against the "original" CERAD (oCERAD) and "modified" CERAD (mCERAD) criteria as described in Table 1. In contrast to the methodology based on the CERAD criteria initially described by Mirra et al. [21], which use a single measurement in each assessed region, multiple samples per cortical region were obtained, with the aim of reducing regional heterogeneity, providing a better estimate of regional burden for comparison to PET imaging results. Finally, an electronic reader training program including image reading methodology and a testing module became available [22], obviating the in-person reader training used in the Curtis study [9].

2. Methods

2.1. Ethics

This study was conducted in accordance with the Declaration of Helsinki, the *Good Clinical Practice: Consolidated Guideline* approved by the International Conference on Harmonisation, and applicable national and local laws and regulations. The initial study imaged the 180-subject cohort and examined the first 68 brains [9] under institutional review board (IRB) and ethics committee approval, and all patients from whom samples would be analyzed in this study were deceased, the IRBs granted exemption from IRB review for this expanded autopsy collection. The investigators

Table 1
Summary of beta-amyloid standards of truth (SoT)

SoT priority in this study	Primary SoT	Secondary SoTs		
SoT name	Modified CERAD (mCERAD)	Original CERAD (oCERAD)	2012 NIA-AA for Thal phasing*	Thal phasing*
Description of SoT	Additional regions sampled compared with oCERAD; final score is the mean	CERAD 1991 regions†; final score is the mode‡	Additional regions sampled compared with oCERAD; final score is the mean	
Sampling (intraregional numeric† assessments, averaged regions)	30 × 8	30 × 4	19	19
Regions	MFL, MTG, STG, IPL, ACG, PCG, PRC, PVC	MFL, MTG, STG, IPL	MO, pons, MBr, CbCx, Dt, Th, STh, BG, Hi, EC, ACC, Amg, MFG, MTG, STG, IPL, OC (BA 17 and 18), WM	
Staining	BSS	BSS	4G8 IHC	
Case dichotomy: definition of "abnormal Aβ plaque count" (type of measure)	>1.5 (Quantitative)	Moderate and frequent (ordinal)‡	Thal phase ≥3 AND moderate and frequent oCERAD when Thal phase = 2 (quantitative)	N/A
Measure	Sensitivity/specificity	Sensitivity/specificity	Sensitivity/specificity	% of majority PET results in each Thal phase
Reference	Curtis et al. [9]	New SoT adapted from Mirra et al. [21]	Hyman et al. [18]; Thal et al. [19,20]	Thal et al. [19,20]
Rationale	Representative of the mean accumulation of neuritic plaques in the eight areas examined. EMA and FDA recommendation as primary SoT for product approval	Representative of the accumulation of neuritic plaques in the totality of the four areas examined, because of multiple assessments	Representative of the mean accumulation of total (neuritic and diffuse) amyloid across multiple regions	Representative of the degree of accumulation of total (neuritic and diffuse) amyloid across multiple regions

Abbreviations: ACC, anterior cingulate cortex; ACG, anterior cingulate gyrus; Amg, amygdala; BA, Broadmann area; BG, basal ganglia; BSS, Bielschowsky silver stain; CbCx, cerebellar cortex; Dt, dorsal terminal nucleus; EC, entorhinal cortex; EMA, European Medicines Agency; FDA, Food and Drug Administration; Hi, hippocampus; IHC, immunohistochemistry; IPL, inferior parietal lobe; MBr, midbrain; MFG, middle frontal gyrus; MFL, middle frontal lobe; MO, medial orbital cortex; MTG, middle temporal gyrus; OC, occipital cortex; PCG, posterior cingulate gyrus; PET, positron emission tomography; PRC, precommissural nucleus; PVC, paraventricular cortex; STG, superior temporal gyrus; STh, subthalamic nucleus; Th, thalamus; WM, white matter.

*Thal phasing is used in a dichotomized format in Hyman et al. [18] to provide a measure of image accuracy, as opposed to a measure of amyloid burden.

†CERAD 1991: earliest CERAD criteria described by Mirra et al. [21], where the level of neuritic plaque burden is evaluated by one global BSS stain assessment from the region of highest amyloid pathology. In contrast, in oCERAD for this work multiple samples (30 per region) were taken to reduce potential heterogeneity of pathology.

‡Mode is taken to allow a noncontinuous dichotomization of the amyloid load.

accepted the exemption and deemed additional informed consent unnecessary.

2.2. Participants

A total of 180 terminally ill subjects with a life expectancy of ≤1 year, aged more than 55 years, and deemed able to undergo study procedures were scanned between June 22, 2010 and November 23, 2011 as described in Curtis et al. [9]. Between July 2010 and February 2014, 108 subjects came to autopsy and 106 brains (68 of which previously reported by Curtis et al. [9]) were considered of sufficient tissue quality for histopathologic assessment. A single patient lacked image interpretation data, which were therefore imputed.

Table 2 displays subject baseline demographics and characteristics. Seventy-eight subjects (73.6%) had a history of dementia and 50% had a clinical diagnosis of dementia because of AD. [^{18}F]Flutemetamol PET imaging was performed at a mean of 7.5 months (ranging between 0 and 28 months) before death. Safety data concerning the administration of Flutemetamol (^{18}F) Injection were reported previously [9] (Clinicaltrials.gov: NCT02090855).

2.3. [^{18}F]Flutemetamol PET scanning

Patients underwent head computed tomography or magnetic resonance scanning either up to 6 months before the PET scan or during PET imaging. The administered activity of Flutemetamol (^{18}F) Injection (Vizamyl, GE Healthcare, Marlborough, MA, USA) was 185 to 370 MBq. PET acquisition was carried out for 10 to 20 min (depending on the injected activity) in 2-minute frames starting 90 minutes after injection. Images were corrected for attenuation. Axial slices (128 × 128) were generated by iterative reconstruction and smoothed with a Gaussian postreconstruction filter when required [9,23].

Table 2
Patient demographics and clinical characteristics

Variable	All subjects (N = 106)
Age* (y)	
Mean (SD)	80.8 (8.87)
Median	83.0
Min, max	59, 95
Gender, *n* (%)	
Male	48 (45.3)
Female	58 (54.7)
Race, *n* (%)	
American Indian or Alaska Native	0 (0.0)
Asian	0 (0.0)
Black	4 (3.8)
Native Hawaiian or other Pacific Islander	0 (0.0)
Caucasian	99 (93.4)
Other	3 (2.8)
Ethnicity, *n* (%)	
Hispanic or Latino	27 (25.5)
Not Hispanic or Latino	79 (74.5)
Time from PET scan to death† (mo)	
n	106
Mean (SD)	7.53 (6.948)
Median	5.45
Min, max	0.0, 28.2
Time from PET scan to death category, *n* (%)	
≤1 y	82 (77.4)
>1 to ≤2 y	19 (17.9)
>2 to ≤3 y	5 (4.7)
History of dementia‡	
Yes	87 (82.1)
No	19 (17.9)
Clinical diagnosis§, *n* (%)	
Alzheimer's disease	53 (50.0)
Other dementing disorder‖	25 (23.6)
Mild cognitive impairment	0 (0.0)
No history of cognitive impairment	27 (25.5)
Memory loss (unspecified)	1 (0.9)

Abbreviations: PET, positron emission tomography; SD, standard deviation.

*Age was calculated as (date of informed consent − date of birth)/365.25 rounded down to the nearest integer.

†Time from PET scan to death was calculated as (date of death − date of PET scan)/30.

‡History of dementia was determined based on medical history reported either at a trial site or by autopsy staff.

§Clinical diagnosis was determined based on reported medical history.

‖May also have been an unspecified form of dementia.

2.4. Blinded image evaluations

Blinded image evaluations (BIEs) of the [^{18}F]flutemetamol PET images were performed at a central review center by five nuclear medicine physicians or radiologists (referred to as "readers") trained in the interpretation of [^{18}F]flutemetamol scans using the validated electronic training program [22] and not involved in previous studies. Because the initial blind had been broken to analyze the data for the initial 68 brain autopsy analysis [9], images were rerandomized and reread in a blinded fashion. The study used a dichotomized read, requiring readers to classify images in a Sokoloff color scale anchored to pons, as either normal or abnormal but not as unevaluable. Readers were allowed to use anatomic brain images (primarily computed tomography and more rarely magnetic resonance imaging) to support interpretation at their discretion. The regions read were frontal lobe, lateral temporal lobe, parietal lobe, posterior cingulate/precuneus, and striatum. If any one of these regions was classified as positive (i.e., abnormal for the presence of amyloid), then the whole image was scored as positive. All readers were instructed to examine the same five regions for amyloid positivity and to report the read outcome both as a global measure and by the individual regions. Further [^{18}F]flutemetamol image read methodology details can be found in Buckley et al. [22].

2.5. Autopsy tissue histopathologic SoT measures

Brains were collected at 19 dementia clinics, memory centers, and hospice centers in the UK and USA. Postmortem routine histopathology was performed locally within 48 hours of collection on 5 μm thick paraffin sections after pretreatment with 88% (vol/vol) formic acid for 5 minutes and heat-mediated antigen retrieval [9]. Amyloid SoT measurements were carried out at a central pathology laboratory for several brain regions, two blocks for each brain region and three slides per block. Using a two-headed microscope, two blinded neuropathologists evaluated simultaneously five randomly chosen 2.5 mm^2 fields of view per slide until a consensus on the neuritic plaque count was reached. The average regional scores were determined by averaging slide scores, which in turn were determined by averaging the field of view scores. The SoTs used are described further in Table 1. The SoT used for the primary analysis was described as mCERAD and involved Bielschowsky silver staining of amyloid plaques in eight regions (the four original CERAD regions, plus anterior and posterior cingulate, precuneus, and primary visual cortex) [21,24]. mCERAD scores reflect the numerical value associated with neuritic plaque burden: scores none (0) and sparse (1) were interpreted as normal, whereas moderate (2) and frequent (3) scores were interpreted as abnormal. The cutoff score for abnormal plaque burden was set as 1.5. A more detailed description of the pathology methods and the neuropathologic analysis of the 106 individual autopsy cases are presented in Ikonomovic et al. [23]. In particular, cerebral amyloid angiopathy (CAA) was measured and categorized by Vonsattel grade, stage, and type.

As a secondary SoT, the oCERAD method [21] involved Bielschowsky silver staining assessment of neuritic plaques from four neocortical regions (mid-frontal, superior and middle temporal, and inferior parietal). The major difference in this current methodology compared with that described in Mirra et al. [21] was the expanded tissue sampling in each region to reduce focal heterogeneity (Table 1). In addition, age and dementia status, components used in the CERAD diagnosis of AD by Mirra et al., were not used, as calculations of sensitivity and specificity simply relied on a comparison of PET image read results with neuritic plaque scores.

Third, the 2012 National Institute of Ageing - Alzheimer's Association (NIA-AA) criteria with amyloid (Thal) phasing involving total amyloid staining (monoclonal antibody, clone

4G8; SIG-39220, Covance, USA, diluted 1:100) were used. The total amyloid staining pattern guided Thal phasing (a 0–5 scale according to the degree of regional spread of amyloid pathology, where a score ≥3 indicates an abnormal load, Fig. 1) [19]. A summary of each of these SoTs, collection methods, and rationale for use and measures are described in Table 1.

2.6. Data analysis

Each blinded visual image interpretation was compared with the SoT measures outlined in Table 1 and classified as a true-positive, false-positive, true-negative, or false-negative (FN), and the classification counts were used to calculate sensitivity and specificity. Intrareader reproducibility (IRR) was determined by including duplicate images of a random sample (11 of 108, ~10%) of the subjects' images, combined randomly and reread along with the other images. Inter-reader agreement (IRA) was determined as described in Curtis et al. [9]. A subset analysis on subjects who died within 1 year of PET imaging was also performed. In addition, results were analyzed by time from PET scan to death to determine whether there was an increase in the FN rate over time (e.g., if patients with true negative scans later developed abnormal amyloid levels). Finally, the proportion of scans associated with each Thal phase as described in the amyloid component of the 2012 NIA-AA criteria of Thal phasing was determined [18,20].

2.7. Statistical analysis

The efficacy population analyzed comprised all subjects who met the study entry criteria and had usable PET images. Sensitivity and specificity were determined as point estimates with exact two-sided 95% confidence intervals (CIs). IRA and IRR were reported as percentage agreement and Cohen's kappa coefficients, respectively. IRA was additionally reported as Fleiss' kappa. The agreement between the three SoT measures was also determined by Fleiss's kappa, whereas pairwise agreement was determined by Cohen's kappa. Sample size was determined by the number of autopsied subjects. Estimates of the distribution of SoT-positive and SoT-negative brains, derived from the initial assessment of the data reported in Curtis et al. [9], were used to formulate assumptions of "true" sensitivity and specificity to perform power calculations. The success threshold for the study was set to >75% for sensitivity and >60% for specificity with respect to the lower bound of the 95% exact binomial CI based on majority image interpretations. Missing image interpretations were imputed as abnormal (positive) if the SoT was normal (negative) and vice versa. The imputed results for missing image interpretations were then used to calculate the majority image interpretation. The impact of covariates (e.g., age, gender, race, dementia history, flutemetamol dose, neuritic plaque density, and time from PET to autopsy) on outcome measures was assessed using a logistic regression model. The PET-to-death time interval was also evaluated with subset analyses using exact two-sided 95% CIs.

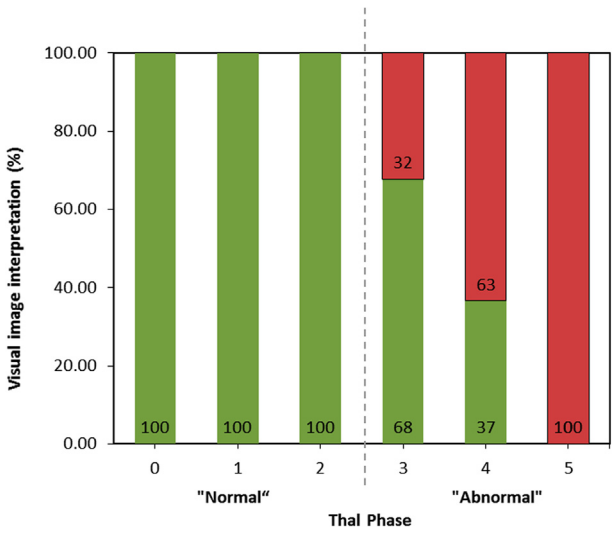

Fig. 1. Percentage of normal/abnormal [^{18}F]flutemetamol majority image interpretations by Thal phase. Green bar: negative scan interpretation; red bar: positive scan interpretation.

3. Results

In total, 106 evaluable subjects (105 with BIE performed and one imputed) were examined (Table 2). By majority image interpretation, 72 subjects' (67.9%) scans were amyloid-positive and 34 subjects' (32.1%) scans were amyloid-negative. Most image interpretations (87%) were performed without consulting anatomic images. The range of use of anatomic images of the five readers varied from 12% to 20%.

3.1. Neuropathologic findings

Using the Bielschowsky silver staining scoring in the primary SoT (mCERAD) regions, 76 subjects (71.7%) were Aβ-positive and the remaining 30 (28.3%) were Aβ-negative. The distribution of neuritic plaque density represented the full spectrum of amyloid neuritic plaque frequencies with cases of none (17%), sparse (20%), moderate (32%), and frequent (36%) plaque densities all being present in this cohort of 106 brains (Fig. 2 and [23]). In addition, the combined findings from the neuropathologic assessments are presented in Supplementary Table 1.

3.2. [^{18}F]Flutemetamol image interpretation: sensitivity and specificity

Sensitivity and specificity are shown in Table 3. On the basis of the primary SoT (mCERAD), the majority image interpretation had a sensitivity of 90.8% (FN rate of 9.2%) and specificity was 90.0% (false-positive rate of 10.0%, or 7% if the single imputed case was excluded) and the study met its primary objective. Using oCERAD gave similar results

(sensitivity 91.9%, specificity 87.5%). Using the 2012 NIA-AA criteria with Thal phasing as SoT, the specificity increased to 100% (thus eliminating false-positive reads), whereas the sensitivity slightly decreased to 85.7% (Table 3).

3.3. Relation between [^{18}F]flutemetamol image interpretation and Thal amyloid phasing

The percentages of normal/abnormal majority PET image interpretations based on Thal phases are shown in Fig. 1. The likelihood of an abnormal BIE increased with increased Thal phase. An abnormal PET scan predicted late Thal phase (4 or 5) with 90% probability (data not shown), whereas there was 0% probability of an early Thal phase (0, 1, or 2). Conversely, a normal scan was associated with a 65% probability of an early Thal phase (2 or less), but a 15% chance of a late Thal phase (4 or 5, data not shown).

3.4. Impact of PET-to-death time interval

Subset analyses probed if the time between PET imaging and death (up to 2.4 years) would increase the risk of FN results (i.e., PET-negative, pathology-positive). In patients who died within 1 year after the PET scan, by-reader sensitivity was 85.7% to 96.4% (majority value 89.3%) and by-reader specificity was 80.8% to 92.3% (majority value 88.5%). In patients who died 1 to 2.4 years after the PET scan, sensitivity and specificity were both >93.3%, indicating that time to death did not influence tracer performance over this time period. A logistic regression analysis confirmed these results, showing that PET-to-death time interval was not significantly associated with an FN image interpretation ($P = .1751$).

3.5. IRA, IRR, and SoT agreement

Pairwise IRA was >91.5%, with Cohen's kappa ranging from 0.80 to 0.92. Agreement across all five readers was 84.9%, with a Fleiss' kappa of 0.84. IRR was 100.0%, with a kappa of 1 for each reader. Agreement across the three SoTs was 89.6% with a Fleiss' kappa of 0.82. Cohen's kappa showed good agreement between the mCERAD and the NIA-AA 2012 criteria (0.80) and between the oCERAD and the NIA-AA 2012 criteria (0.75), whereas there was an excellent agreement between mCERAD and oCERAD.

4. Discussion

This study comprises a 106-subject autopsy cohort, which represents the largest cohort to date aiming to assess accuracy in amyloid PET image reading [5,16,25,26]. Blinded visual interpretation of [^{18}F]flutemetamol images exhibited high sensitivity and specificity (majority values of 90.8% and 90.0%, respectively) using the mCERAD SoT. Similar diagnostic performance was obtained using the secondary SoTs (oCERAD and 2012 NIA-AA criteria, as per Table 1). By reader (including the majority read), the three 95% CIs overlapped considerably for both sensitivity and

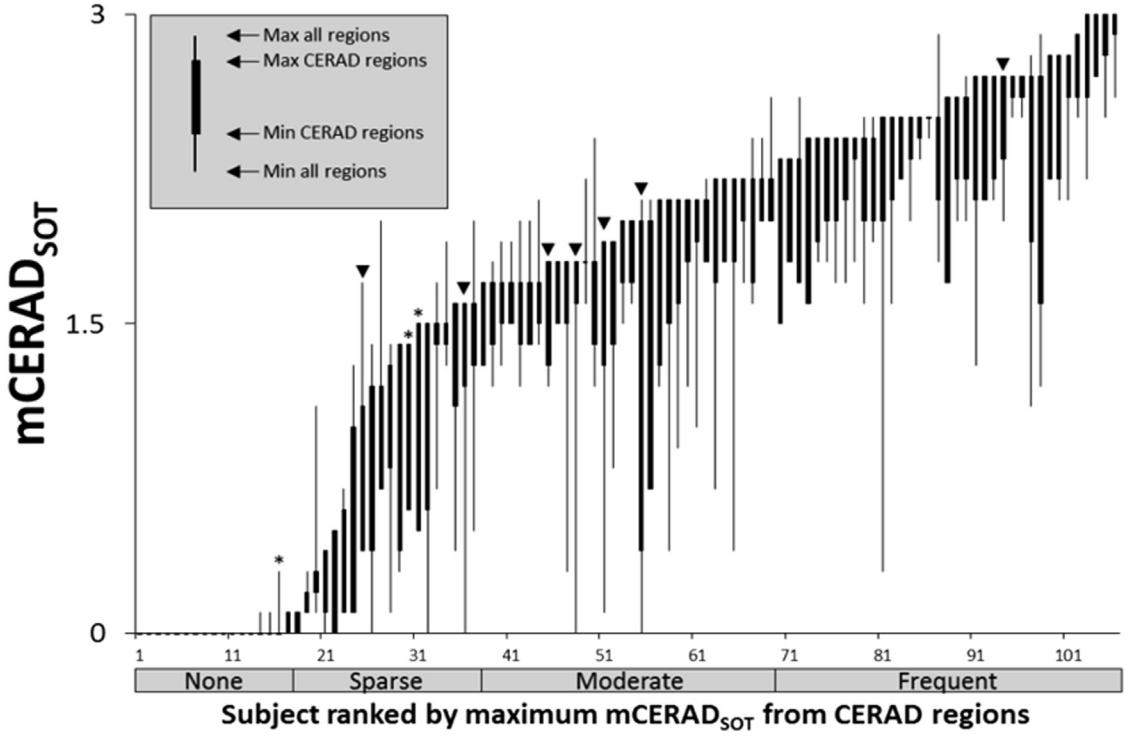

Fig. 2. Distribution of the selected population for the mean neuritic plaque score across the spectrum of amyloid pathology. ▼ = False-negative read. * = False-positive read. mCERAD$_{SOT}$ = neuritic plaque density CERAD scoring (0–3). Note: several cases had borderline pathology, leading to lower reader confidence. Reproduced with permission from Ikonomovic et al. [23].

Table 3
Sensitivity and specificity of blinded visual interpretation of flutemetamol images according to three SoTs

	Sensitivity,* % (95% CI)			Specificity,† % (95% CI)		
Reader	mCERAD	oCERAD	2012 NIA-AA Thal phasing	mCERAD	oCERAD	2012 NIA-AA Thal phasing
1	96.1 (88.9, 99.2)‡	95.9 (88.6, 99.2)	91.7 (83.6, 96.6)	83.3 (65.3, 94.4)§	78.1 (60.0, 90.7)	95.5 (77.2, 99.9)
2	89.5 (80.3, 95.3)‡	90.5 (81.5, 96.1)	84.5 (75.0, 91.5)	90.0 (73.5, 97.9)§	87.5 (71.0, 96.5)	100.0 (84.6, 100.0)
3	88.2 (78.7, 94.4)‡	87.8 (78.2, 94.3)	82.1 (72.3, 89.6)	93.3 (77.9, 99.2)§	87.5 (71.0, 96.5)	100.0 (84.6, 100.0)
4	88.2 (78.7, 94.4)‡	89.2 (79.8, 95.2)	84.5 (75.0, 91.5)	83.3 (65.3, 94.4)§	81.3 (63.6, 92.8)	95.5 (77.2, 99.9)
5	94.7 (87.1, 98.5)‡	95.9 (88.6, 99.2)	90.5 (82.1, 95.8)	86.7 (69.3, 96.2)§	84.4 (67.2, 94.7)	100.0 (84.6, 100.0)
Majority‖	90.8 (81.9, 96.2)‡	91.9 (83.2, 97.0)	85.7 (76.4, 92.4)	90.0 (73.5, 97.9)§	87.5 (71.0, 96.5)	100.0 (84.6, 100.0)

Abbreviations: CI, confidence interval; SoT, standard of truth.
*Primary analysis. Sensitivity = true positives/(true positives + false negatives).
†Specificity = true negatives/(true negatives + false positives).
‡Two-sided 95% CI with a lower bound >75%.
§Two-sided 95% CI with a lower bound >60%.
‖The majority interpretation is the interpretation made independently by more than half of the readers.

specificity, suggesting no significant differences between SoTs. These results compare favorably to those reported previously by Curtis et al. [9]. In this study, however, because of the larger number of subjects, the 95% CIs were narrower, reflecting increased precision around the measurements. A strong regional association between the tracer and mCERAD and oCERAD regions has also been reported [21], see Table 6 of Ikonomovic et al. [23] and a good correspondence between quantitative measures of fibrillar amyloid and [^{18}F]flutemetamol PET signal from normal pressure hydrocephalus patients who had cortical biopsy sampling either before or after PET imaging [27].

When dichotomized Thal phasing according to 2012 NIA-AA criteria [18] was used as the SoT, the specificity was raised to 100%. This SoT combines the presence of both neuritic and diffuse Aβ-immunoreactive plaques (i.e., total Aβ plaque burden), and these new results suggest that [^{18}F]flutemetamol detects both forms of Aβ plaques when present in sufficient amounts. These observations with the Thal-based SoT are thought to explain the three apparent false-positive cases observed when neuritic plaque density alone was used as the SoT. The approximate 90% accuracy of [^{18}F]flutemetamol performance when neuritic plaque density alone is the SoT suggests that in this cohort of 106 subjects most of the detected amyloid was present as neuritic plaques; the rise to 100% specificity indicates a small number of cases contributing to the visual PET signal when both neuritic and diffuse plaques are present and detected. These observations are expected to be similarly applicable to quantitative PET image read, as strong concordance with visual read results has been demonstrated previously [28].

This study (data in Fig. 1) also confirms the previous findings [19] that visual interpretation of [^{18}F]flutemetamol PET images strongly predicts Thal phase, which reflects the progressive deposition of brain amyloid from neocortex (phase 1), to allocortex (phase 2), to diencephalon and basal ganglia (phase 3), to midbrain and medulla oblongata (phase 4), and finally to cerebellum and pons (phase 5). [^{18}F]Flutemetamol is particularly effective in detecting the most advanced Thal phases (4 and 5) [19] where the deposition of amyloid was recorded in clinically proven AD cases in the original work by Thal et al. [20]. Similarly, in this work, a positive scan interpretation was best predicted by an increasing spread of amyloid burden, whereas a low burden restricted to neocortex and allocortex (Thal phases 0–2) was always associated with a negative scan. Conversely, a negative (normal) scan was associated with a 65% probability of an early Thal phase (0, 1, or 2), but only a 15% chance of a late Thal phase (4 or 5). These results confirm that the likelihood of a positive scan interpretation increases with the progressive deposition of amyloid. There were five cases, which were Thal phase 4 but were reviewed as normal by majority visual PET read (Fig. 1) and Fig. 6c of Ikonomovic et al. [29]. In these cases, atrophy played a significant part in diminishing the gray matter intensity in the PET images leading to a majority normal (FN) assessment. The majority reads for these cases were mostly not unanimous.

The results of the [^{18}F]flutemetamol-related Thal analysis are complementary to those of Murray et al. [30] who used the research tracer [^{11}C]PiB to compare autopsy findings with PET image results. Murray et al. report a boundary between Thal phases 1 and 2 equating to a [^{11}C]PiB threshold of >1.4 whereas [^{18}F]flutemetamol reports positivity coming in at Thal phase 3. These differences may be because of the use of quantitation for the [^{11}C]PiB measure, which is likely to be more sensitive than the visual inspection methods used currently for the routine use of fluorine-18 labeled tracers.

Additional insight into the diagnostic value of [^{18}F]flutemetamol can be gathered by its striatal uptake pattern [31]. [^{18}F]Flutemetamol is unique across the fluorinated PET amyloid tracers, in that it uses the striatum as one of its key read regions. The striatum contains amyloid primarily in the diffuse form [28] and is less susceptible to atrophy than the neocortex [32], hence its value in scan interpretation [22,31]. Increasing levels of striatal amyloid were found to be associated with increasing sensitivity of image interpretation, from 77% for moderate striatal plaque density to 82% for frequent plaques [31]. These values of sensitivity are lower than those observed with mCERAD and oCERAD as SoTs (Table 3) suggesting that compared

with neuritic plaques the [^{18}F]flutemetamol signal may be less intense in diffuse plaques. Further work is ongoing in this area to better understand the relationship between different forms of amyloid pathology and [^{18}F]flutemetamol imaging signal. For example, the use of plaque area and signal intensity to give an integrated density is being investigated by Ikonomovic et al. [29], who showed that using the fluorescent labeled probe cyano-flutemetamol neuritic plaques give a stronger signal than diffuse plaques of similar area. This finding is consistent with the lower sensitivity reported by Beach et al. [31] when striatal pathology alone is used as the SoT.

Regarding specificity, the observations in this article are consistent with the results of Beach et al. [31]; specificity was 100% (i.e., no false-positive image interpretations) using the 2012 NIA-AA criteria with Thal phasing, whether based on the striatum alone or on the global brain regions analyzed. Both articles contribute to understanding the nature of amyloid pathology in the brain and the limitations of using neuritic Aβ plaques as the SoT measurement for characterizing the performance of PET amyloid imaging agents.

Within the 106 autopsy data set were seven FN cases based on majority PET read. Three of these had borderline neuritic plaque scores, and three of the remaining four had some discordance between brain region plaque scoring and the PET read. Hence, most of the FN cases may be explained by the difficulty in assigning a positive or negative amyloid score for borderline cases, coupled with the cortical atrophy present in all FN cases. Six of the seven FN cases were not unanimous in terms of the reported read results (four cases of four negative/one positive; two cases of three negative/two positive, data not shown). There was only one case with extensive radial atrophy where all five readers unanimously interpreted the case as PET negative in the presence of positive cortical amyloid (data not shown). These observations point to the importance of using an anatomic image when atrophy is suspected [15,22].

In addition to neuritic and diffuse amyloid, the presence of CAA could contribute to the PET signal. In this autopsy cohort, there were no cases where CAA was present in the absence of fibrillar amyloid, although there was circumstantial evidence to suggest that when the cortical neuritic amyloid was borderline, the presence of CAA pathology may have contributed to a majority positive PET read. More evidence however is required, and studies to examine amyloid imaging in intracerebral lobar hemorrhage, for example, may give us clues as to the extent of CAA deposition and how that compares to a PET amyloid signal.

In this study, sensitivity and specificity were unaffected by time from PET scan to death, consistent with the research by Villemagne et al. [33], that amyloid deposition occurs over decades. IRA was higher than previously reported in Curtis et al. [9] (>91.5%, κ = 0.80–0.92 compared with ≥80%, κ = 0.44–0.97), although one reader in Curtis et al. [9] was clearly an outlier. The overall improvement in IRA may be the result of training readers with a validated electronic reader training program [22], which provides reliable, dependable instruction, and skill development for accurate PET image interpretation, even when a possibly substantial number of subjects manifested severe cortical atrophy (a potential confounding factor for image interpretation [34]). As for the SoT agreement, although overall a very good agreement between the three pathologic criteria was found, the slightly lower concordance between the mCERAD/oCERAD and NIA-AA 2012 would be expected because mCERAD and oCERAD are based on Bielschowsky silver stain identifying neuritic amyloid alone whereas NIA-AA 2012 is based on total fibrillar amyloid load.

The comparison between visual image reads and quantitation of [^{18}F]flutemetamol was the subject of a publication by Thurfjell et al. [28]. Here a threshold between negative and positive amyloid was derived based on a mean plus two standard deviations in a healthy group. When applied to a cohort of more than 170 subjects from a reader validation study [22], the concordance was >99% when using either the pons or whole cerebellum as a reference region. When quantitation was compared with the read results from cases where there was a histology SoT there was near identical sensitivity and specificity from both the visual read and quantitative comparisons [28].

This study had several limitations. Our primary results were the majority interpretation of five readers, a research setting that does not reflect clinical practice. The study's end-of-life population does not reflect the patient demographic anticipated to benefit most from amyloid PET imaging in the clinical setting (patients with mild cognitive impairment), but was a necessary trade-off for comparing PET image interpretations to autopsy findings. The high proportion of Caucasian subjects could be viewed as a limitation to this study; however, recent data indicate that [^{18}F]flutemetamol has similar performance in both Caucasian and Japanese subjects and that readers interpret negative and positive scans almost identically [35]. The Japanese study also used visual interpretation of images, which may have involved some degree of judgment for scans close to the pathology thresholds and those with severe atrophy; theoretically, these cases might benefit from quantitative measurement of [^{18}F]flutemetamol uptake [28] as described previously. Nevertheless, experience in image reading has shown that the inferior parietal area and the striatum are less susceptible to atrophy than other regions. Thus, from a visual read perspective, these areas provide a more robust region to observe when atrophy is suspected [20].

In conclusion, this reported study found that in a cohort of 106 autopsy cases, the visual interpretation of [^{18}F]flutemetamol PET images had high sensitivity and specificity in detecting brain amyloid in life, regardless of the SoT measure. It also provides evidence that [^{18}F]flutemetamol binds to both neuritic and diffuse plaques, and that the striatum, which is little affected by atrophy, is a robust region for assessing brain amyloid. The newly found specificity of [^{18}F]flutemetamol for both neuritic and diffuse Aβ plaques (which possibly precedes neuritic plaque formation)

constitutes an important advantage in the early detection of dementia [36].

Acknowledgments

We gratefully acknowledge Enrico R. Fantoni of GE Healthcare Amersham, UK, for providing advice and assistance in drafting, editing, and compiling review comments. We express our gratitude to the patients and their families, who selflessly agreed to participate in this study.

Declaration of Interests: GE Healthcare funded this study. Employees of GE Healthcare participated in the design, administration, monitoring, reporting, and interpretation of the study, and drafting and critical revision of the manuscript for important intellectual content. S.S. had full access to all the data and had final responsibility for the decision to submit the article for publication.

Author Contributions: S.S., R.V., M.D.I., P.F.S., G.F., A.P.L.S., C.J.B., and D.R.T. planned the study design. S.S., J.E.G., U.S., C.H.S., T.V., M.N.S., R.D., A.S.F., K.A.F., Z.W., A.H., Y.M.E., M.E.A., J.R., A.B., M.A., J.S., and C.C. recruited and scanned patients and collected tissues for autopsy. S.S., M.D.I., P.F.S., G.F., A.P.L.S., C.J.B., D.R.T., A.S.F., Z.W., K.A.F., and M.Z. performed the data review, interpretation, and preparation of the manuscript. T.GB., M.D.I., A.P.L.S., and D.R.T. provided amyloid pathology expertise, whereas C.J.B. provided image analysis expertise.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.dadm.2017.06.001.

RESEARCH IN CONTEXT

1. Systematic review: The authors reviewed the literature using traditional sources (e.g., PubMed), clinical study reports, and internal reports. The sensitivity and specificity of the visual interpretation of [^{18}F]flutemetamol images following in-person training were previously evaluated against an autopsy-based standard of truth, albeit in a smaller cohort.
2. Interpretation: The findings presented here led to a refined interpretation for apparently false-positive [^{18}F]flutemetamol image reads and to a more precise estimation of sensitivity and specificity.
3. Future directions: The additional benefit of quantitative image processing could add value to results obtained using visual assessment alone, particularly when pathology is borderline between sparse and moderate plaques.

References

[1] Cotran R, Kumar V, Robbins S. Robbins Pathologic Basis of Disease. 6th ed. Philadelphia: W.B. Saunders Company; 1999.
[2] Vallet PG, Guntern R, Hof PR, Golaz J, Delacourte A, Robakis NK, et al. A comparative study of histological and immunohistochemical methods for neurofibrillary tangles and senile plaques in Alzheimer's disease. Acta Neuropathol 1992;83:170–8.
[3] Cooper JH. An evaluation of current methods for the diagnostic histochemistry of amyloid. J Clin Pathol 1969;22:410–3.
[4] Clark CM, Schneider JA, Bedell BJ, Beach TG, Bilker WB, Mintun MA, et al. Use of florbetapir-PET for imaging beta-amyloid pathology. JAMA 2011;305:275–83.
[5] Clark CM, Pontecorvo MJ, Beach TG, Bedell BJ, Coleman RE, Doraiswamy PM, 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.
[6] Barthel H, Gertz HJ, Dresel S, Peters O, Bartenstein P, Buerger K, et al. Cerebral amyloid-β PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. Lancet Neurol 2011;10:424–35.
[7] Vandenberghe R, Van Laere K, Ivanoiu A, Salmon E, Bastin C, Triau E, et al. 18F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. Ann Neurol 2010;68:319–29.
[8] Rowe CC, Pejoska S, Mulligan RS, Jones G, Chan JG, Svensson S, et al. Head-to-head comparison of 11C-PiB and 18F-AZD4694 (NAV4694) for β-amyloid imaging in aging and dementia. J Nucl Med 2013;54:880–6.
[9] Curtis C, Gamez JE, Singh U, Sadowsky CH, Villena T, Sabbagh MN, et al. Phase 3 trial of flutemetamol labeled with radioactive fluorine 18 imaging and neuritic plaque density. JAMA Neurol 2015;72:287.
[10] Ikonomovic MD, Klunk WE, Abrahamson EE, Mathis CA, Price JC, Tsopelas ND, et al. Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. Brain 2008;131:1630–45.
[11] FDA. US Prescribing information for Neuraceq 2016 n.d. Available at: http://www.accessdata.fda.gov/drugsatfda_docs/label/2016/204677s012lbl.pdf. Accessed March 20, 2017.
[12] EMA. Neuraceq EU Summary of Product characteristics 2016 n.d. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002553/WC500162592.pdf. Accessed March 20, 2017.
[13] FDA. US Prescribing information for Amyvid 2013 n.d. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/202008s020lbl.pdf. Accessed March 20, 2017.
[14] FDA. US Prescribing information for Vizamyl 2017 n.d. Available at: http://www.accessdata.fda.gov/drugsatfda_docs/label/2017/203137s008lbl.pdf. Accessed March 16, 2017.
[15] EMA. Vizamyl EU Summary of Product characteristics 2017 n.d. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/002557/WC500172950.pdf. Accessed March 20, 2017.
[16] Sabri O, Seibyl J, Rowe C, Barthel H. Beta-amyloid imaging with florbetaben. Clin Transl Imaging 2015;3:13–26.
[17] Snellman A, Rokka J, López-Picón FR, Eskola O, Salmona M, Forloni G, et al. In vivo PET imaging of beta-amyloid deposition in mouse models of Alzheimer's disease with a high specific activity PET imaging agent [18F]flutemetamol. EJNMMI Res 2014;4:37.
[18] Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, et al. National Insitute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. Alzheimers Dement 2012;8:1–13.
[19] Thal DR, Beach TG, Zanette M, Heurling K, Chakrabarty A, Ismail A, et al. [18F]flutemetamol amyloid positron emission tomography in preclinical and symptomatic Alzheimer's disease: specific detection of advanced phases of amyloid-β pathology. Alzheimers Dement 2015;11:975–85.

[20] Thal DR, Rüb U, Orantes M, Braak H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. Neurology 2002;58:1791–800.
[21] Mirra S, Heyman A, McKeel D, Sumi S. The consortium to establish a registry for Alzheimer's disease (CERAD). Part II. standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 1991;46:142–5.
[22] Buckley CJ, Sherwin PF, Smith APL, Wolber J, Weick SM, Brooks DJ. Validation of an electronic image reader training programme for interpretation of [18F]flutemetamol β-amyloid PET brain images. Nucl Med Commun 2017;38:234–41.
[23] Ikonomovic MD, Buckley C, Heurling K, Sherwin PF, Jones PA, Zanette M, et al. Post-mortem histopathology underlying β-amyloid PET imaging following flutemetamol F 18 injection. Acta Neuropathol Commun 2016;4:130–54.
[24] Vemuri P, Whitwell JL, Kantarci K, Keith A, Parisi JE, Shiung MS, et al. Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. Neuroimage 2011;42:559–67.
[25] Seibyl J, Catafau AM, Barthel H, Ishii K, Rowe CC, Leverenz JB, et al. Impact of training method on the robustness of the visual assessment of 18F-florbetaben PET scans: results from a Phase 3 trial. J Nucl Med 2016;57:900–6.
[26] Lundqvist R, Lilja J, Thomas BA, Lötjönen J, Villemagne VL, Rowe CC, et al. Implementation and validation of an adaptive template registration method for 18F-flutemetamol imaging data. J Nucl Med 2013;54:1472–8.
[27] Leinonen V, Rinne JO, Virtanen KA, Eskola O, Rummukainen J, Huttunen J, et al. Positron emission tomography with [18F]flutemetamol and [11C]PiB for in vivo detection of cerebral cortical amyloid in normal pressure hydrocephalus patients. Eur J Neurol 2013; 20:1043–52.
[28] Thurfjell L, Lilja J, Lundqvist R, Buckley C, Smith A, Vandenberghe R, et al. Automated quantification of 18F-flutemetamol PET activity for categorizing scans as negative or positive for brain amyloid: concordance with visual image reads. J Nucl Med 2014;55:1623–8.
[29] Ikonomovic MD, Abrahamson EE, Buckley CJ, Mathis CA, Klunk WE, Farrar G. Contribution of neuritic and diffuse plaques to signal derived from CN-Flutemetamol: a preliminary study in AD autopsy brains. Human Amyloid Imaging 2017 Meeting Abstract 2017:150.
[30] Murray ME, Lowe VJ, Graff-Radford NR, Liesinger AM, Cannon A, Przybelski SA, et al. Clinicopathologic and ^{11}C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum. Brain 2015;138:1370–81.
[31] Beach TG, Thal DR, Zanette M, Smith A, Buckley C. Detection of striatal amyloid plaques with [18F]flutemetamol: validation with postmortem histopathology. J Alzheimers Dis 2016;52:863–73.
[32] Bertoux M, O'Callaghan C, Flanagan E, Hodges JR, Hornberger M. Fronto-striatal atrophy in behavioral variant frontotemporal dementia and Alzheimer's disease. Front Neurol 2015;6:147.
[33] Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid-beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol 2013;12:357–67.
[34] Mountz JM, Laymon CM, Cohen AD, Zhang Z, Price JC, Boudhar S, et al. Comparison of qualitative and quantitative imaging characteristics of [11C]PiB and [18F]flutemetamol in normal control and Alzheimer's subjects. Neuroimage Clin 2015;9:592–8.
[35] Miki T, Shimada H, Kim JS, Yamamoto Y, Sugino M, Kowa H, et al. Brain uptake and safety of Flutemetamol F 18 injection in Japanese subjects with probable Alzheimer's disease, subjects with amnestic mild cognitive impairment and healthy volunteers. Ann Nucl Med 2017;31:260–72.
[36] Price JL, McKeel DW, Buckles VD, Roe CM, Xiong C, Grundman M, et al. Neuropathology of nondemented aging: presumptive evidence for preclinical Alzheimer disease. Neurobiol Aging 2009;30:1026–36.