The early perfusion image is useful to support the visual interpretation of brain amyloid-PET with [^{18}F]flutemetamol in borderline cases

Franziska L. Mathies[1]*(ORCID ID 0000-0002-8041-9893) and Fiona Heeman[2,3,4,5]* (ORCID:0000-0001-7769-8329), Pieter Jelle Visser[6,7,8,9], Anouk den Braber[6,8], Maqsood Yaqub[4,5] (0000-0003-2122-740X), Wiesje van der Flier[6], Susanne Klutmann[1], Gill Farrar[10], Michael Schöll[2,3,11,12](0000-0001-7800-1781), Elsmarieke van de Giessen[4], Lyduine E. Collij[4,5] † and Ralph Buchert[1†] (ORCID ID 0000-0002-0945-0724)

[1]Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[2]Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Sweden

[3]Department of Psychiatry and Neurochemistry, Institute of Physiology and Neuroscience, University of Gothenburg, Gothenburg, Sweden

[4]Amsterdam UMC, location Vrije Universiteit Amsterdam, Department of Radiology and Nuclear Medicine, Boelelaan 1117, Amsterdam, The Netherlands

[5]Amsterdam Neuroscience, Brain Imaging, Amsterdam, The Netherlands

[6]Amsterdam UMC, location Vrije Universiteit Amsterdam, Department of Neurology, Alzheimer Center Amsterdam, Boelelaan 1117, Amsterdam, The Netherlands

[7]Alzheimer Center Limburg, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands

[8]Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

[9]Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska Institutet, Stockholm, Sweden

[10]GE Healthcare, Pharmaceutical Diagnostics, Amersham, HP7 9LL, UK

[11]Department of Clinical Physiology, Sahlgrenska University Hospital, Gothenburg, Sweden

[12]Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom

[13]Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden

* F. L. Mathies and F. Heeman contributed equally as first authors

† L.E. Collij and R. Buchert contributed equally as last authors

Corresponding author: Ralph Buchert, Martinistr. 52, 20246 Hamburg, Germany, Email: r.buchert@uke.de, Phone: +49 40 741054347, Fax: +49 40 741040265, ORCID ID 0000-0002-0945-0724

Abstract

Purpose: Visual interpretation of brain amyloid (Aß) PET can be difficult in patients with borderline Aß burden. Co-registration with individual MRI is recommended in these cases, which however is not always available in clinical practice. This study therefore evaluated co-registration with the early perfusion image acquired immediately after tracer injection to support the visual interpretation of the late Aß-image in brain PET with [^{18}F]flutemetamol (FMM).

Methods: Fifty dual-time-window FMM-PET scans of cognitively normal subjects with Centiloid (CL) scale between 0 and 60 points were included retrospectively (70.1±6.9 years, 56% female, Mini-Mental State Examination score 28.9±1.3, 42% APOE ε4 carrier). Regional amyloid load was scored in each of the 10 regions-of-interest (ROI) recommended for FMM reading using three different settings: Aß-image only, Aß-image co-registered with MRI and Aß-image co-registered with the perfusion image. This was performed by 3 independent raters using a 6-point Likert scale. The impact of setting, within- and between-readers variability, ROI and amyloid-status (Aß-negative or -positive: CL ≤ or > 21) was tested by repeated measures analysis of variance of the Likert score.

Results: The CL scale ranged between 2 and 52 (interquartile range 7-19). Support of visual scoring by the perfusion image resulted in the best discrimination between Aß-positive and Aß-negative cases, mainly by improved certainty of excluding Aß plaques in Aß-negative cases ($p = 0.030$). It also resulted in significantly higher between-raters agreement. The setting effect was most pronounced in the frontal lobe and in the posterior cingulate cortex/precuneus area ($p = 0.005$).

Conclusions: The early perfusion image is a suitable alternative to T1w-MRI to support the visual interpretation of the late Aß-image in FMM-PET.

Key words: Amyloid PET; Flutemetamol; Regional visual read; Perfusion; Centiloid

Abbreviations

Aß: amyloid-ß

AIC: Akaike information criterion

ANOVA: analysis of variance

AUC: Area under the ROC curve

CI: confidence interval

CL: Centiloid

CL-Aß-positive/negative: CL > / ≤ 21 points

FDG: [^{18}F]fluorodeoxyglucose

FMM: [^{18}F]flutemetamol

MNI: Montreal Neurological Institute

MRI: Magnetic Resonance Imaging

PC: precuneus

PCC: posterior cingulate cortex / precuneus area

PET: positron emission tomography

ROC: receiver operating characteristic curve

ROI: Region-of-interest

SPM: statistical parametric mapping

SUVR: standardized uptake value ratio

T1w: T1-weighted

vis-Aß-positive/negative: visually Aß-positive/negative (automatically derived from the regional Likert 6-scale scores)

VSS: visual sum score

Introduction

Positron emission tomography (PET) with amyloid-ß (Aß) tracers allows *in vivo* detection (or exclusion) of fibrillar Aß deposition in the cerebral cortex, one of the earliest pathological hallmarks of Alzheimer's disease (AD) [1]. In the clinical routine, binary interpretation of Aß-PET scans as Aß-positive or -negative is based on the visual assessment of the PET images according to tracer-specific guidelines defined by the tracers' manufacturers. This is rather straightforward in the majority of the cases, but approximately 10 % of the cases are equivocal, i.e., they are not clearly Aß-positive nor clearly Aß-negative [2-4]. This may be either due to borderline amyloid burden or due to marked brain atrophy that makes discrimination between specific cortical tracer binding and unspecific tracer uptake in white matter challenging. Co-registration of the Aß-PET image with the individual T1-weighted (T1w) MRI is recommended in these cases to assist the visual assessment [5, 6]. However, T1w-MRI with sufficiently high isotropic resolution is not always available in clinical routine, for example in patients with MRI-incompatible implants or claustrophobia.

Previous studies have shown that stable surrogates of relative cerebral blood flow (i.e., relative tracer delivery, R_1, or standardized uptake value ratio, SUVR) can be obtained from the early distribution phase of Aß-PET tracers [7-9]. Furthermore, high correlations have been reported between early uptake images of the different Aß-PET tracers and cerebral glucose metabolism measured by PET with [^{18}F]fluorodeoxyglucose (FDG) [10-14], as expected, given the tight coupling between blood flow and glucose metabolism in the brain [15]. These findings suggest that Aß-PET is suitable for dual-biomarker imaging, by providing measures of both amyloid burden and neuronal dysfunction/degeneration.

An additional putative benefit of the early (perfusion) image in Aß-PET is that it might be used to support the discrimination between specific cortical tracer binding and unspecific white matter uptake in the late Aß image [16, 17]. Preliminary data even suggest a small improvement in reader confidence in the interpretation of Aß-PET with [^{18}F]florbetaben when using the co-

registered early perfusion image for this task rather than co-registered T1w-MRI [18]. At sites that regularly acquire an early perfusion image as proxy of FDG-PET, it would be readily available to support the interpretation of the late Aß PET image without additional cost.

Against this background, the aim of the current study was to assess the value of the early perfusion image for the visual interpretation of the late Aß image in PET with [^{18}F]flutemetamol (FMM). More specifically, the study evaluated visual scoring of the late Aß image alone with respect to amyloid pathology compared to the Aß image co-registered with either the early perfusion image or a high-resolution T1w-MRI. The patient sample used for this purpose was enriched for cases with borderline findings regarding Aß load, since support of the visual PET reading is hypothesized to be most needed and/or beneficial in borderline cases.

Materials and Methods

Subjects

Cognitively unimpaired subjects with dual-time-window FMM PET were included retrospectively from the Amsterdam twin-pair sub-study of the EMIF-AD PreclinAD cohort (http://www.emif.eu/emif-ad-2/) [19, 20]. After randomly selecting one subject from each twin-pair and excluding subjects with Centiloid (CL) score outside the range $0 < CL < 60$ (in order to enrich the sample for borderline cases), 50 subjects were selected randomly from the remaining dataset.

MR Imaging

MR imaging was performed with a 3T Philips Achieva scanner using a 32-channel head coil at the Department of Radiology & Nuclear Medicine of the Amsterdam UMC, location VUmc, Amsterdam, as described previously [19]. Only the 3D T1w scan was used in the current study (sagittal turbo field echo sequence, 1.00 mm x 1.00 mm x 1.00 mm voxels, repetition time = 7.9 ms, echo time = 4.5 ms, flip angle = 8 degrees).

PET Imaging

FMM PET had been performed with a Philips Ingenuity Time-of-Flight PET-MRI system at the Amsterdam UMC, location VUmc, as described previously [19]. In brief, dual-time-window scans were acquired from 0-30 min (6 × 5 s, 3 × 10 s, 4 × 60 s, 2 × 150 s, 2 × 300 s, 1 × 600 s) and from 90-110 min (4 x 5 min) after intravenous injection of 185 (± 10%) MBq FMM [21]. Attenuation correction of the PET emissions scans was based on a dedicated MR sequence acquired immediately before each PET scan. Global amyloid burden was characterized according to the CL scale computed with a validated standard CL pipeline [22, 23].

In preparation of visual PET assessment, frames 10-15 of the early dynamic scan, corresponding to the time interval 1-8 min post injection, were integrated with rigid-body inter-frame motion correction to obtain a static perfusion image [10]. The first minute after injection was not included in order to avoid contamination by the high intravascular activity concentration immediately after intravenous tracer injection. The four frames of the late dynamic scan (corresponding to the 90-110 min post injection time interval) were integrated with rigid-body inter-frame motion correction to obtain the Aß-image. The Aß-image was co-registered to the perfusion image using the 'coregister' tool of the statistical parametric mapping software package (version: SPM12) with the 'estimate only' option. Finally, the perfusion and the co-registered Aß image were co-registered to the subject's 3D-T1w-MRI using the SPM12 'coregister' tool with the individual MRI as reference and the perfusion image as source.

For intensity scaling of the PET images, the pons mask provided by the Centiloid project [22] was elastically transformed from the anatomical space of the Montreal Neurological Institute (MNI) to the individual patient space using the 'Normalize' tool in SPM12. The perfusion image was scaled on a voxel-by-voxel basis to the mean voxel intensity in the pons to create an SUVR image. The Aß image was scaled to 90% of the mean intensity of the 1 ml hottest voxels in the pons. The rationale for this was to simplify compliance with the recommendation by the FMM prescribing information [24] stating that *"images should be viewed with the minimum image*

intensity set to zero and the maximum set such that the signal level in the easily identifiable pons is at 90% of maximum", namely by setting the lower and the upper threshold of the colour table for display of the scaled Aß image to 0.0 and 1.0, respectively.

Visual scoring

Visual scoring was performed independently by three raters blinded to all clinical information. Each rater had successfully completed the electronic training program provided by GE Healthcare [25] (https://www.readvizamyl.com/). Two scoring sessions of the 50 cases were performed, each consisting of three subsessions with different randomization of the cases, one subsession for each of the three settings tested for visual scoring: Aß-PET image alone (setting A), Aß-PET image together with the co-registered individual T1w-MRI (setting B), and Aß image together with the co-registered perfusion image (setting C). As a warming-up exercise, each subsession started with two additional training cases (not included in the 50 test cases), one clearly Aß-negative and one clearly Aß-positive case. Raters had to complete all reads from a given subsession within one week. The next subsession was started after a wash-out period of at least one week to a maximum of four weeks.

Transverse, sagittal, and coronal views were displayed using the 'View' tool of the PMOD software package (version 3.8.04) or using the Vinci software (version 5.06). In subsessions including co-registered T1w-MRI or perfusion images, the two modalities (Aß image and MRI or perfusion image) were viewed side-by-side with corresponding crosshairs, and as fusion image (Figure 1). "Spectrum" or "rainbow" were the standard colour scales for the display of the Aß images. However, the raters were allowed to use also other colour scales (e.g., "Sokoloff") or to switch between different colour scales while reading a case. The lower and the upper threshold of the colour scale for the (intensity-scaled) Aß image were set to 0.0 and 1.0, respectively. The raters were allowed to vary the thresholds while reading a case. The raters were also allowed to use 2-dimensional colour scaling (such that the lower and the upper threshold were set to the minimum and maximum of the corresponding 2-dimensional slice, separately for each 2-

dimensional slice). The T1w-MRI was displayed with the grey colour scale, the perfusion image was displayed with grey or inverse grey colour scale. In settings B and C, the raters were asked to use the corresponding crosshairs to support anatomical localization of the FMM signal in white or grey matter in the corresponding 'anatomical' image (MRI or perfusion).

Tracer uptake was evaluated in the following regions-of-interest (ROI) separately for left and right hemisphere: lateral temporal cortex, frontal cortex including anterior cingulate cortex, striatum, posterior cingulate cortex (PCC) / precuneus (PC) area and temporo-parietal cortex including insula. A region was considered Aß-negative if the tracer signal was distinctly lower than adjacent white matter and similar in intensity to the grey matter of the cerebellum. A region was considered Aß-positive if the signal was similar to or of higher intensity as adjacent white matter and greater than the grey matter of the cerebellum. The raters were asked to score each of the 10 ROIs separately using the following Likert 6-point scale: -3 = most likely Aß-negative, -2 = probably Aß-negative, -1 = possibly Aß-negative, 1 = possibly Aß-positive, 2 = probably Aß-positive to 3 = most likely Aß-positive.

In order to assess the visual scoring on a global binary level, the regional Likert 6-point scores were dichotomized and then 'globalized'. More precisely, an Aß-PET scan was considered 'visually Aß-positive' (vis-Aß-positive) if at least one of the 10 regional Likert 6-point scores was positive, otherwise (i.e., all regional Likert 6-point scores were negative) it was considered 'visually Aß-negative' (vis-Aß-negative). This was performed separately for each setting, each rater, and each scoring session.

Finally, visual scoring of an Aß-PET scan was considered 'certain' if the overall binary visual interpretation (automatically derived from the regional 6-point scores) agreed across all 6 reads (3 raters * 2 sessions) as either vis-Aß-positive or vis-Aß-negative. This was performed separately for each setting.

Statistical analyses

Subjects were categorized as Aß-positive or Aß-negative using 21 points on the CL scale as cut-off, which provides the highest accuracy for CL-based detection of moderate to frequent plaque density [26, 27]. More precisely, Aß images with CL > 21 were considered 'CL-Aß-positive', and Aß images with a CL ≤ 21 were considered 'CL-Aß-negative'.

Factors influencing visual assessment

The impact of setting (N = 3), ROI (N = 10), rater (N = 3), and scoring session (N = 2) on the Likert 6-point score was tested using repeated measures analysis of variance (ANOVA) with setting, ROI, rater, and session as within-subject factors and CL-Aß-status as between-subjects factor. This was considered the 'full model'.

For post-hoc testing of pairwise differences between two settings, the repeated measures ANOVA of the regional Likert 6-point score was performed with the same within- and between-subjects factors as in the full model, except that the setting factor was restricted to the two considered settings.

In order to eliminate putative training effects, the repeated measures ANOVA of the regional Likert 6-point score was restricted to the scores from the second scoring session, while keeping all other factors the same as in the full model.

Semi-quantitative visual characterization of the global amyloid load

In order to evaluate the potential of visual scoring for semi-quantitative characterization of the global amyloid load, the regional Likert 6-point score was summed over all 10 ROIs, separately for each rater, each session, and each setting. The results were then averaged over all raters and both scoring sessions, separately for each setting. This resulted in three visual sum scores (VSS) ranging from -30 to + 30, one for each setting. Low/high values of the VSS are expected to indicate low/high global amyloid load.

The relationship between the VSS and the CL score was tested with the non-parametric Spearman test.

The ability of the VSS to discriminate between CL-Aß-positive and CL-Aß-negative scans was assessed by receiver operating characteristic (ROC) analyses. The difference in the area under the ROC curve between the three VSS was tested pairwise using DeLong's test for correlated ROC curves [28].

Since visual inspection of scatter plots suggested a linear relationship between VSS and the CL scale only above a given break-point (and constant VSS below the break-point), the following broken-stick model was tested to describe the relationship between VSS and CL, separately for each setting:

(1) $VSS = VSS_0$, if $CL \leq$ break-point, and

$VSS = VSS_1 + slope * CL$, if $CL >$ break-point,

with the following boundary condition in order to avoid a discontinuous step at the break-point:

(2) $VSS_1 = VSS_0 + slope * break\text{-}point$.

This broken-stick model has three free parameters (VSS_0, break-point, and slope) that were optimized by minimization of the sum of the squared differences between the data points and the model. For comparison, straight lines were fitted to the scatter plots. The quality of the fits was compared between the broken-stick model and the straight line using the Akaike information criterion (AIC) with correction for finite sample size.

Within- and between-raters variability

Fleiss kappa was used to characterize within- and between-raters variability of the regional Likert 6-point score and of the global binary visual interpretation (derived from the Likert score). Statistical significance of the difference between two kappa estimates was tested by checking their 83.4% confidence intervals (CI) for overlap, with non-overlapping 83.4%-CI indicating

statistical significance with 5% type 1 error probability [29]. The 83.4%-CI was computed as: kappa ± 1.385* standard error of kappa [30]. The interpretation of Fleiss kappa values regarding the strength of within- and between-raters agreement was according to Landis and Koch [31].

The proportion of 'certain' cases was compared between two settings using cross tables and the chi-squared test, separately for CL-Aß-negative and CL-Aß-positive cases.

Statistical analyses were performed using SPSS version 27 (SPSS Inc., Chicago, Illinois). All *p*-values reported are two-sided. Statistical significance was defined as $p \leq 0.05$. The broken-stick model fit was performed with MATLAB (The Mathworks). In all repeated measures ANOVA, sphericity was tested for with Mauchly's test. Since sphericity could not be assumed in the vast majority of cases, Greenhouse-Geisser correction was used for all within-subject tests.

Results

The mean age of the 50 included individuals was 70.1±6.9 years (range 61-86 years), 56% were female, and the mean Mini-Mental State Examination score was 28.9±1.3 (range 25-30). Fourty-two percent of the subjects carried at least one APOE ε4 allele.

The global amyloid burden ranged from 2 to 52 CL (mean 15±11 CL, median 12 CL, interquartile range 7-19 CL). Applying the CL cut-off of 21 resulted in 42 (84%) CL-Aß-negative and 8 (16%) CL-Aß-positive cases.

Factors influencing visual assessment

The results of the different repeated measures ANOVA of the regional Likert 6-point score are summarized in Table 1 and Figure 2. In the full model, all factors and their interactions had a statistically significant impact on the regional Likert 6-point score, except for the setting*session*CL-Aß-status and session*CL-Aß-status interaction factors. The setting effect reached trend level significance (p = 0.078). Results from the ANOVA restricted to the scores

from the second scoring session were highly comparable to the results from the full model, except for the lack of significance of the rater*CL-Aß-status interaction factor (Table 1).

The pairwise post-hoc analyses demonstrated a significant setting effect between setting C and setting A, and a significant setting*CL-Aß-status interaction effect between setting C and both, setting A and B, with better separation between CL-Aß-positive and CL-Aß-negative cases for setting C (Figure 2a).

All ANOVA consistently showed a highly significant rater effect (Figure 2d,e) and a highly significant ROI effect on the Likert 6-point score (Figure 2b,c).

Semi-quantitative visual characterization of the global amyloid load

The VSS correlated significantly with the CL score for each of the three settings for visual scoring: Spearman correlation coefficient = 0.48, 0.54 and 0.64 for setting A, B and C, respectively (all $p < 0.0005$).

The area under the ROC curve of the VSS for differentiating between CL-Aß-positive and CL-Aß-negative cases was 0.87, 0.91, and 0.93 for setting A, B and C, respectively (Figure 3). None of the pairwise differences were statistically significant ($p \geq 0.298$).

The broken-stick fits of the relationship between VSS and CL scale are shown in Figure 4. The broken-stick provided a better fit than a straight line for all three settings (AIC = 195 versus 197, 161 versus 174, and 174 versus 182 in setting A, B, and C, respectively). The break-point of the broken-stick was at 13.2, 14.9, and 13.2 CL for setting A, B, and C, respectively. The slope of the broken-stick beyond the break-point was 0.81, 1.01, and 1.06 VSS points per CL point for setting A, B, and C.

Within- and between-raters variability

The results on Fleiss within- and between-raters kappa of the regional Likert 6-point score across all ROIs and of the binary global visual interpretation are shown in Figure 5. The only significant setting effects were higher between-raters kappa of the Likert 6-point score for setting C (kappa = 0.193, 83.4%-CI [0.171, 0.215]) compared to setting A (0.125, [0.103, 0.147]) and compared to setting B (0.102 [0.080, 0.124]).

Between-raters agreement on individual categories was best for the Likert score +3 (most likely Aß-positive) followed by the Likert score -3 (most likely Aß-negative) in all settings (Likert score = +3: kappa = 0.368, 0.238 and 0.531 for settings A, B and C; Likert score = -3: kappa = 0.203, 0.200 and 0.322 for settings A, B and C).

The Fleiss kappa values for between- and within-rater variability of the Likert 6-point score for different ROIs are shown in Supplementary Figure 1.

Amongst the CL-Aß-negative/positive FMM-PET, the proportion of visually 'certain' cases was 36/38%, 19/63%, and 36/63% for setting A, B, and C, respectively (Figure 6). The lower proportion of visually certain cases amongst the CL-Aß-negative FMM-PET with setting B reached trend level statistical significance compared to settings A and C (both $p = 0.087$). All other pairwise differences were not significant ($p \geq 0.317$).

Discussion

The current study tested the utility of the early perfusion image to support the visual assessment of the late Aß-image in FMM-PET (setting C) compared to visual inspection of the Aß-image alone (setting A) or accompanied by a co-registered individual high-resolution T1w-MRI (setting B). The clinical relevance of this research question arises out of the fact that MRI is not always available in clinical routine. The results support the use of the early FMM-PET perfusion image as suitable alternative to the T1w-MRI for visual interpretation of the late Aß-image.

The primary finding of the analyses was a setting effect on the regional Likert 6-score that was statistically significant when comparing setting C with setting A and reached trend level significance in the full model including all settings (Table 1). The setting effect was driven by highest certainty with setting C both in excluding Aß plaques in CL-Aß-negative cases (lowest Likert scores among all settings) as well as in detecting Aß plaques in CL-Aß-positive cases (highest Likert scores). The difference in (estimated marginal means of) the regional Likert 6-point scores between CL-Aß-positive and CL-Aß-negative cases increased from 1.66 in setting A to 1.83 in setting B to 2.10 in setting C (Figure 2a), indicating that setting C provided the best power for discrimination of CL-Aß-negative and CL-Aß-positive cases. This finding was corroborated by the tests of the VSS to discriminate between CL-Aß-positive and CL-Aß-negative cases, as the ROC analyses demonstrated the highest area under the curve for setting C, although the difference to the two other settings did not reach statistical significance.

The improvement by setting C was most pronounced in CL-Aß-negative cases: the certainty for the exclusion of regional Aß plaques in CL-Aß-negative cases was 0.39 points higher for setting C compared with setting A and 0.16 points compared with setting B. This level of improvement might be considered clinically relevant considering that the dynamic range of negative Likert score covers only 2.0 points (from -3 to -1).

Further support for setting C was provided by the within- and between-raters analysis. Fleiss between-raters kappa was significantly higher in setting C compared to both other settings (Figure 5). In line with this, the proportion of visually ‘certain’ cases, that is, cases with consistent binary global visual interpretation across all reads, was largest in setting C (Figure 6). The higher proportion of certain cases amongst the CL-Aß-negative cases in setting C compared to setting B reached trend level significance ($p = 0.087$).

The setting effect significantly differed between ROIs, and the ROI-dependence of the setting effect significantly depended on CL-Aß-status. The improvement in the certainty to exclude Aß plaques in CL-Aß-negative cases from setting A to B to C was most pronounced in the frontal

and temporal lobe and in the striatum. In CL-Aß-positive scans, the certainty for detecting Aß plaques improved from setting A to B to C for the frontal lobe and the PCC/PC. Interestingly, the PCC/PC assessment shifted from a negative mean Likert score (suggesting no local Aß burden) for settings A and B to a positive Likert score (suggesting local Aß burden) for setting C. For the other ROIs (i.e., temporal, parietal, striatum) there was no change or even a small decrease of the Likert score starting from negative values in setting A to more strongly negative values in settings B and C. A possible explanation for this is that these ROIs were actually Aß-negative in some (or most) of the globally CL-Aß-positive FMM-PET. This is not surprising, as most CL-Aß-positive scans were close to the cut-off of 21 CL, which is indicative of relatively low and likely focal Aß-burden, most often visually detectable first in the frontal lobe [27]. Overall, the added value of setting C was largest in the frontal lobe and in the PCC/PC, where it improved the certainty of both, exclusion of Aß plaques in CL-Aß-negative cases and detection of Aß plaques in CL-Aß-positive cases. This is practically relevant, given that the frontal lobe is most often perceived as difficult to assess, leading to the greatest doubt for final classification [32].

Further evidence for an added value of setting C was provided by the analyses regarding the utility of the VSS for the semi-quantitative assessment of the global Aß load. Spearman analyses revealed the highest correlation between the VSS and the CL for setting C. When the relationship of the VSS with the CL scale was modelled by a broken-stick, the break point ranged between 13-15 CL in all settings (Figure 4), in excellent agreement with the 12 points CL cut-off to exclude Aß-pathology reported previously [33]. Below the breakpoint, the VSS was more or less constant and, therefore, not useful to characterize global Aß load. From the breakpoint onwards, the VSS seems to be a useful proxy of the CL scale, probably most useful when using setting C which resulted in the steepest linear relation between VSS and CL after the break point.

The utility of the early perfusion image to support the visual interpretation of the Aß-image in FMM-PET has been shown before by Belohlavek and colleagues who segmented grey matter areas on the early perfusion image and overlaid the resulting grey matter mask onto the Aß-image.

This procedure improved between-raters agreement of visual scoring compared to conventional scoring of the Aß-image alone, but it was not compared to support by MRI [16, 17].

A secondary finding of the current study was the generally quite high within- and between-raters variability of the visual scoring, despite the fact that all 3 raters had successfully completed the FMM reader training. The ANOVA (full model) revealed that almost 50% of the total variance of the Likert 6-point score was explained by between-sessions (= within-raters) variability (20.4%) and between-raters variability (25.6%) (Table 1). This finding was reiterated by the Fleiss kappa analyses. Even for the *global binary* visual interpretation, within-raters agreement was only moderate to substantial and between-raters agreement was only fair to moderate (Figure 5). In part, this can be explained by the rather fine-grained 6-point Likert score implemented in the current study, as Fleiss kappa does not take into account the magnitude of differences, that is, minor discrepancies in the Likert 6-point score (e.g., -3 versus -2) were weighted the same as major discrepancies (e.g., -3 versus +3). Furthermore, the patient sample of the current study was enriched with borderline cases (CL interquartile range 7-19). Together these two factors explain the larger within- and between-raters variability as reported in previous studies [16, 23, 32, 34]. It is important to note that the proportion of borderline cases in typical patient samples referred to amyloid-PET in clinical routine is only about 10% [2-4]. Thus, the high within- and between-raters variability in the current patient sample, that was strongly enriched with borderline cases, does not directly translate to typical clinical patient cohorts. Nevertheless, it might be useful to extend amyloid-PET reader trainings to give more detailed recommendations regarding borderline cases and to include more such cases in the test dataset to be scored for successful completion of the training.

Finally, we hypothesize that the findings of the current study apply also to other F-18 labelled amyloid tracers with similar brain kinetics early after intravenous injection (to allow perfusion imaging) and similar grey-to-white matter contrast in the amyloid image, such as [^{18}F]florbetaben and [^{18}F]florbetapir.

Limitations of the current study include the following. First, the CL scale was used as reference for Aß-pathology, as post-mortem histopathological data were not available. Second, the use of 21 points as cut-off on the CL scale to dichotomize Aß load led to a rather unbalanced sample with respect of Aß-status. This might have limited the statistical power of the study for the detection of further actual effects. Finally, the readers were not asked to perform a global interpretation of the FMM-PET that was binary from the start, which is the standard procedure in clinical practice.

In conclusion, the perfusion image is a suitable alternative to high-resolution T1w-MRI to support the visual interpretation of the late Aß image in FMM-PET.

Statements and Declarations

Competing Interests: GF is employee of GE Healthcare. This did not influence the content of this manuscript, neither directly nor indirectly. The nonemployee authors had full control of the data and information that might present a conflict of interest for the employee authors. LEC received research support from GE Healthcare and Spinger Healthcare (funded by Eli Lilly). Both contributions have been paid to the institution. There is no actual or potential conflict of interest for the other authors.

Author Contributions: All authors contributed to study conception, design and data collection. The analyses were performed by Franziska Mathies, Fiona Heeman, Lyduine Collij and Ralph Buchert. The first draft of the manuscript was written by Franziska Mathies, Fiona Heeman and Ralph Buchert. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval / Consent to participate: The EMIF-AD PreclinAD study was approved by the ethics committee of the Amsterdam UMC, location VUmc, and all research was performed

in accordance with the principles of the 1964 Helsinki declaration and its later amendments. Written informed consent was obtained from all participants.

References

1. Chapleau M, Iaccarino L, Soleimani-Meigooni D, Rabinovici GD. The Role of Amyloid PET in Imaging Neurodegenerative Disorders: A Review. J Nucl Med. 2022;63:13S-9S. doi:10.2967/jnumed.121.263195.

2. Hosokawa C, Ishii K, Hyodo T, Sakaguchi K, Usami K, Shimamoto K, et al. Investigation of (11)C-PiB equivocal PET findings. Ann Nucl Med. 2015;29:164-9. doi:10.1007/s12149-014-0924-8.

3. Oh M, Seo M, Oh SY, Kim H, Choi BW, Oh JS, et al. Clinical significance of visually equivocal amyloid PET findings from the Alzheimer's Disease Neuroimaging Initiative cohort. Neuroreport. 2018;29:553-8. doi:10.1097/WNR.0000000000000986.

4. Payoux P, Delrieu J, Gallini A, Adel D, Salabert AS, Hitzel A, et al. Cognitive and functional patterns of nondemented subjects with equivocal visual amyloid PET findings. Eur J Nucl Med Mol Imaging. 2015;42:1459-68. doi:10.1007/s00259-015-3067-9.

5. Healthcare G. Vizamyl. Summary of product characteristics. European Medicines Agency. 2014 (assessed July 20, 2022).

6. Minoshima S, Drzezga AE, Barthel H, Bohnen N, Djekidel M, Lewis DH, et al. SNMMI Procedure Standard/EANM Practice Guideline for Amyloid PET Imaging of the Brain 1.0. J Nucl Med. 2016;57:1316-22. doi:10.2967/jnumed.116.174615.

7. Chen Y, Rosario B, Laymon C, Bi WZ, Matan C, Lopresti B, et al. Evaluation of PiB relative delivery value (R-1) as a proxy of relative cerebral blood flow. J Cerebr Blood F Met. 2012;32:S170-S1.

8. Heeman F, Yaqub M, Lopes Alves I, Heurling K, Bullich S, Gispert JD, et al. Simulating the effect of cerebral blood flow changes on regional quantification of [(18)F]flutemetamol and [(18)F]florbetaben studies. J Cereb Blood Flow Metab. 2021;41:579-89. doi:10.1177/0271678X20918029.

9. Heeman F, Visser D, Yaqub M, Verfaillie S, Timmers T, Pijnenburg YA, et al. Precision estimates of relative and absolute cerebral blood flow in Alzheimer's disease and cognitively normal individuals. J Cereb Blood Flow Metab. 2023;43:369-78. doi:10.1177/0271678X221135270.

10. Boccalini C, Peretti DE, Ribaldi F, Scheffler M, Stampacchia S, Tomczyk S, et al. Early-phase (18)F-Florbetapir and (18)F-Flutemetamol images as proxies of brain metabolism in a memory clinic setting. J Nucl Med. 2022;64:266-73. doi:10.2967/jnumed.122.264256.

11. Daerr S, Brendel M, Zach C, Mille E, Schilling D, Zacherl MJ, et al. Evaluation of early-phase [(18)F]-florbetaben PET acquisition in clinical routine cases. Neuroimage Clin. 2017;14:77-86. doi:10.1016/j.nicl.2016.10.005.

12. Hsiao IT, Huang CC, Hsieh CJ, Hsu WC, Wey SP, Yen TC, et al. Correlation of early-phase 18F-florbetapir (AV-45/Amyvid) PET images to FDG images: preliminary studies. Eur J Nucl Med Mol Imaging. 2012;39:613-20. doi:10.1007/s00259-011-2051-2.

13. Rostomian AH, Madison C, Rabinovici GD, Jagust WJ. Early 11C-PIB frames and 18F-FDG PET measures are comparable: a study validated in a cohort of AD and FTLD patients. J Nucl Med. 2011;52:173-9. doi:10.2967/jnumed.110.082057.

14. Oliveira FPM, Moreira AP, de Mendonca A, Verdelho A, Xavier C, Barroca D, et al. Can 11C-PiB-PET Relative Delivery R1 or 11C-PiB-PET Perfusion Replace 18F-FDG-PET in the Assessment of Brain Neurodegeneration? J Alzheimers Dis. 2018;65:89-97. doi:10.3233/JAD-180274.

15. Paulson OB, Hasselbalch SG, Rostrup E, Knudsen GM, Pelligrino D. Cerebral blood flow response to functional activation. J Cerebr Blood F Met. 2010;30:2-14. doi:10.1038/jcbfm.2009.188.

16. Belohlavek O, Jaruskova M, Skopalova M, Szarazova G, Simonova K. Improved beta-amyloid PET reproducibility using two-phase acquisition and grey matter delineation. Eur J Nucl Med Mol Imaging. 2019;46:297-303. doi:10.1007/s00259-018-4140-y.

17. Belohlavek O, Jaruskova M. An easy way to increase confidence in beta-amyloid-PET evaluation. Nuclear Medicine Review. 2017;20:107-9.

18. Mathies F, Lange C, Apostolova I, Frings L, Klutmann S, Meyer PT, et al. Visual interpretation of amyloid-ß PET might be supported by co-registration to the early uptake image. Nuklearmedizin-NuclearMedicine. 2020;59.02:P154.

19. Konijnenberg E, Carter SF, Ten Kate M, den Braber A, Tomassen J, Amadi C, et al. The EMIF-AD PreclinAD study: study design and baseline cohort overview. Alzheimers Res Ther. 2018;10:75. doi:10.1186/s13195-018-0406-7.

20. Coomans EM, Tomassen J, Ossenkoppele R, Tijms BM, Lorenzini L, Ten Kate M, et al. Genetically identical twin-pair difference models support the amyloid cascade hypothesis. Brain. 2023. doi:10.1093/brain/awad077.

21. Heeman F, Yaqub M, Alves IL, Heurling K, Berkhof J, Gispert JD, et al. Optimized dual-time-window protocols for quantitative [F-18]flutemetamol and [F-18]florbetaben PET studies. Ejnmmi Res. 2019;9. doi:ARTN 3210.1186/s13550-019-0499-4.

22. Klunk WE, Koeppe RA, Price JC, Benzinger TL, Devous MD, Sr., Jagust WJ, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement. 2015;11:1-15 e1-4. doi:10.1016/j.jalz.2014.07.003.

23. Collij LE, Salvado G, Shekari M, Lopes Alves I, Reimand J, Wink AM, et al. Visual assessment of [(18)F]flutemetamol PET images can detect early amyloid pathology and grade its extent. Eur J Nucl Med Mol Imaging. 2021;48:2169-82. doi:10.1007/s00259-020-05174-2.

24. Healthcare G. Vizamyl - Flutemetamol F18 Injection: Highlights of prescribing information. 2017.

25. Buckley CJ, Sherwin PF, Smith AP, Wolber J, Weick SM, Brooks DJ. Validation of an electronic image reader training programme for interpretation of [18F]flutemetamol beta-amyloid PET brain images. Nucl Med Commun. 2017;38:234-41. doi:10.1097/MNM.0000000000000633.

26. Amadoru S, Dore V, McLean CA, Hinton F, Shepherd CE, Halliday GM, et al. Comparison of amyloid PET measured in Centiloid units with neuropathological findings in Alzheimer's disease. Alzheimers Res Ther. 2020;12:22. doi:10.1186/s13195-020-00587-5.

27. Pemberton HG, Collij LE, Heeman F, Bollack A, Shekari M, Salvado G, et al. Quantification of amyloid PET for future clinical use: a state-of-the-art review. Eur J Nucl Med Mol Imaging. 2022;49:3508-28. doi:10.1007/s00259-022-05784-y.

28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837-45.

29. Knol MJ, Pestman WR, Grobbee DE. The (mis)use of overlap of confidence intervals to assess effect modification. Eur J Epidemiol. 2011;26:253-4. doi:10.1007/s10654-011-9563-8.

30. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? BMC Med Res Methodol. 2016;16:93. doi:10.1186/s12874-016-0200-9.

31. Landis JR, Koch GG. Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33:159-74. doi:Doi 10.2307/2529310.

32. Collij LE, Konijnenberg E, Reimand J, Kate MT, Braber AD, Lopes Alves I, et al. Assessing Amyloid Pathology in Cognitively Normal Subjects Using (18)F-Flutemetamol PET: Comparing Visual Reads and Quantitative Methods. J Nucl Med. 2019;60:541-7. doi:10.2967/jnumed.118.211532.

33. La Joie R, Ayakta N, Seeley WW, Borys E, Boxer AL, DeCarli C, et al. Multisite study of the relationships between antemortem [(11)C]PIB-PET Centiloid values and postmortem measures of Alzheimer's disease neuropathology. Alzheimers Dement. 2019;15:205-16. doi:10.1016/j.jalz.2018.09.001.

34. Walker Z, Inglis F, Sadowsky C, Safirstein B, Farrar G, Buckley C, et al. Reproducibility of [18f]flutemetamol pet amyloid image interpretation. Journal of the Neurological Sciences. 2013;333:e352.

.

Tables

Table 1 Within-subjects effects according to repeated measures ANOVA of the regional Likert 6-point score including all levels of all factors (full model, 2nd column), with setting C removed (for post-hoc comparison of settings A and B, 3rd column), with setting B removed (setting A versus C, 4th column), with setting A removed (setting B versus C, 5th column), and restricted to the scores from the second scoring session (to eliminate putative learning effects, 6th column). The table specifies the partial eta-squared characterizing the proportion of the variance of the Likert 6-point score explained by the considered factor after accounting for the variance explained by all other factors. All p-values are corrected for non-sphericity using the Greenhouse-Geisser method. Significant effects are indicated by bold type. ("p = 0.000" = p < 0.0005, n.a. = not applicable)

	Partial eta-squared (p-value)				
Factor	Full model	Setting B versus A	Setting C versus A	Setting C versus B	2nd scoring session only
Setting	0.054 (0.078)	0.055 (0.100)	**0.078 (0.050)**	0.003 (0.718)	0.010 (0.582)
Setting*CL-Aß-status	**0.075 (0.030)**	0.018 (0.350)	**0.122 (0.013)**	**0.093 (0.031)**	**0.066 (0.048)**
Setting*ROI	**0.052 (0.005)**	0.012 (0.720)	0.319 (0.054)	**0.090 (0.000)**	**0.069 (0.000)**
Setting*ROI*CL-Aß-status	**0.051 (0.006)**	0.043 (0.054)	**0.419 (0.005)**	0.022 (0.375)	**0.044 (0.018)**
Setting*rater	**0.083 (0.007)**	**0.127 (0.003)**	0.047 (0.323)	**0.109 (0.005)**	**0.069 (0.015)**
Setting*rater*CL-Aß-status	**0.097 (0.002)**	**0.131 (0.002)**	**0.203 (0.005)**	0.006 (0.719)	**0.058 (0.033)**
Setting*session	**0.092 (0.013)**	0.035 (0.192)	**0.130 (0.010)**	**0.083 (0.042)**	n.a.
Setting*session*CL-Aß-status	0.002 (0.898)	0.001 (0.869)	0.003 (0.722)	0.002 (0.772)	n.a.
ROI	**0.248 (0.000)**	**0.234 (0.000)**	**0.828 (0.000)**	**0.231 (0.000)**	**0.240 (0.000)**
ROI*CL-Aß-status	**0.121 (0.000)**	**0.092 (0.002)**	**0.685 (0.000)**	**0.143 (0.000)**	**0.114 (0.000)**
Rater	**0.256 (0.000)**	**0.239 (0.000)**	**0.258 (0.001)**	**0.273 (0.000)**	**0.145 (0.002)**
Rater*CL-Aß-status	**0.073 (0.031)**	0.037 (0.168)	0.057 (0.250)	**0.135 (0.002)**	0.038 (0.164)
Session	**0.204 (0.001)**	**0.351 (0.000)**	**0.127 (0.011)**	0.041 (0.160)	n.a.
Session*CL-Aß-status	0.011 (0.468)	0.018 (0.356)	0.006 (0.579)	0.003 (0.693)	n.a.

Figures

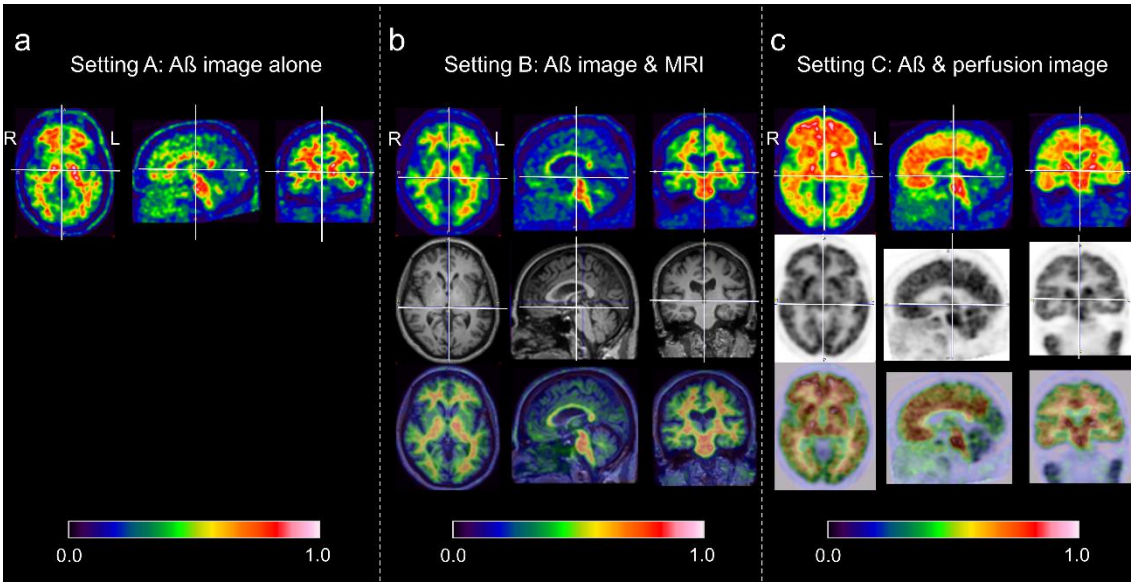


Figure 1 Example displays used for visual scoring of the Aß image alone (setting A, a), of the Aß image together with the co-registered T1w-MRI (setting B, b), and of the Aß image together with the perfusion image (setting C, c). In settings B and C, the raters were asked to use the corresponding cross-hairs to support anatomical localization of the FMM signal in white or gray matter in the corresponding 'anatomical' image (MRI or perfusion). The example shown with setting A is one of the 50 test cases (male, 65 years, ApoE E4 positive, Minimental State Examination Score 29, Centiloid scale 18). The global visual binary interpretation was 'visually Aß-positive' in 3 of the 6 scoring sessions (3 raters x 2 sessions), it was 'visually Aß-negative' in the remaining 3 scoring sessions. The examples shown with settings B and C are clearly Aß-positive (Centiloid = 95) and clearly Aß-negative (Centiloid = -11) training cases not included in the test dataset.

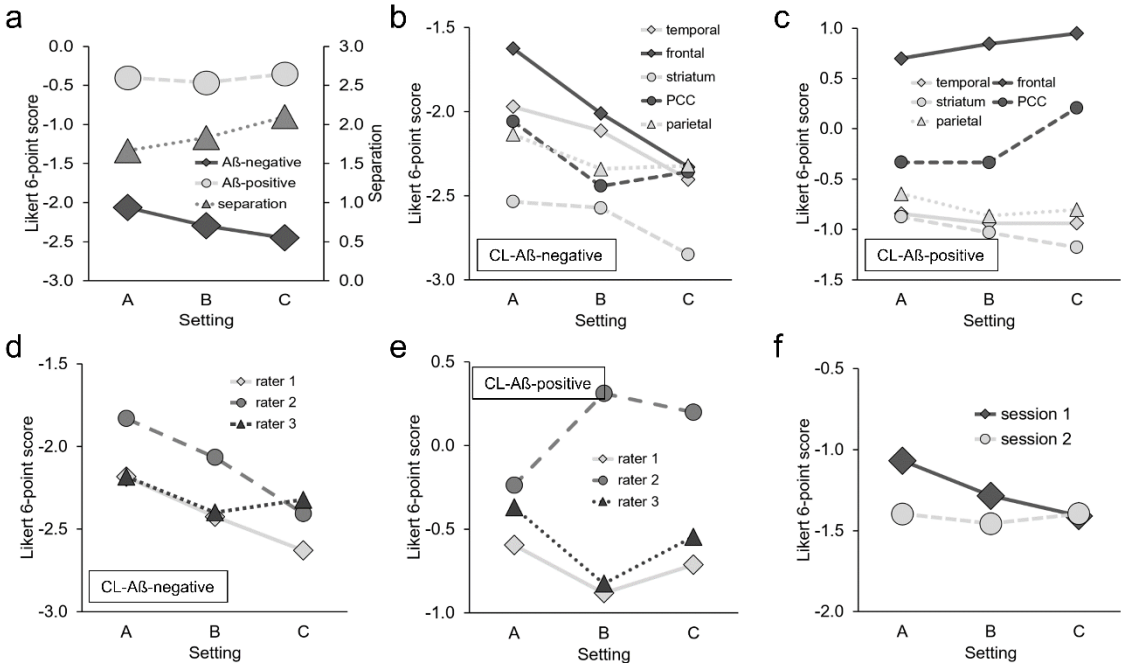


Figure 2 Marginal means of the Likert 6-point score for the different visual scoring settings estimated by repeated measures ANOVA (full model). The marginal means are presented according to CL-Aß-status (a), ROI (average of left and right hemisphere), separately in CL-Aß-negative (b) and CL-Aß-positive (c) cases, rater, separately in CL-Aß-negative (d) and CL-Aß-positive (e) cases, and according to scoring session (f). In part a, 'separation' indicates the difference of the marginal mean of the Likert 6-point score between CL-Aß-positive and CL-Aß-negative cases according to the secondary (right) vertical axis. (setting A = Aß-PET only, B = Aß-PET & T1w-MRI, C = Aß-PET & perfusion-PET)

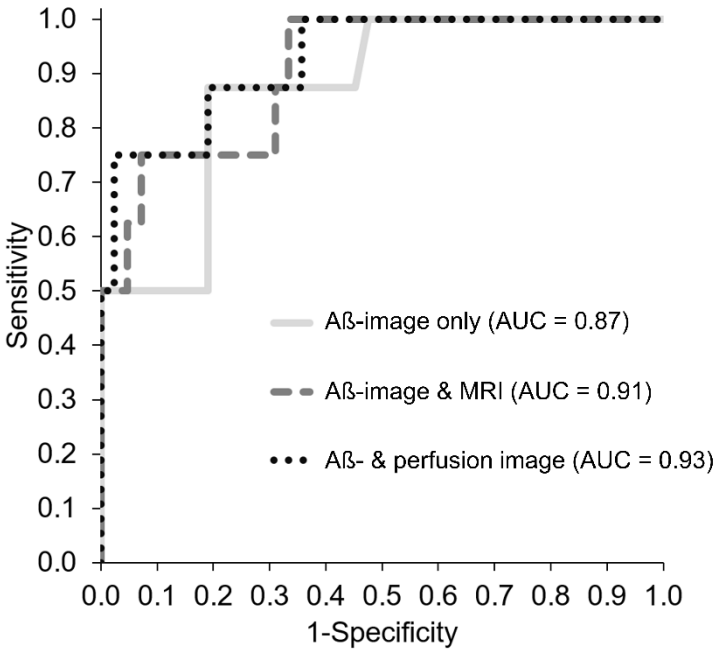


Figure 3 ROC curves for the differentiation between CL-Aß-positive and CL-Aß-negative FMM-PET by the visual sum scores for each of the three tested settings.

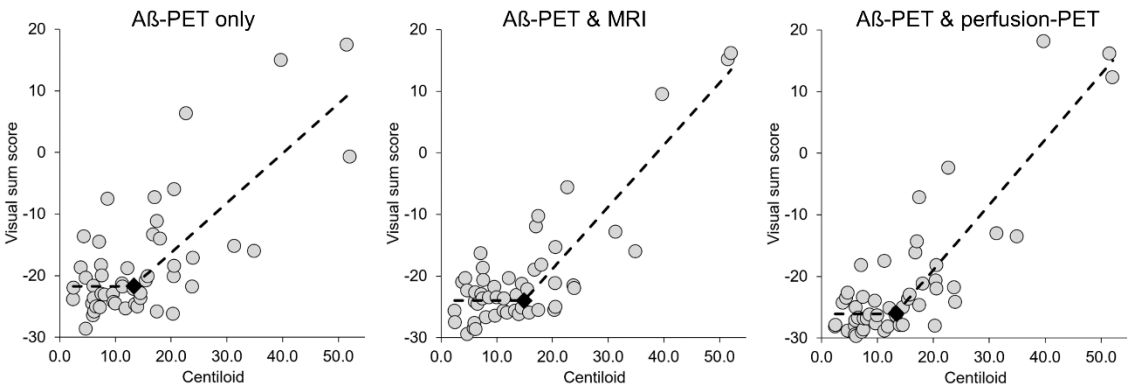


Figure 4 Scatter plots of the visual scum scores versus the Centiloid scale in the three settings for visual scoring. The dashed lines indicate the broken-stick model fit. The black rhombus indicates the break-point of the broken-stick.

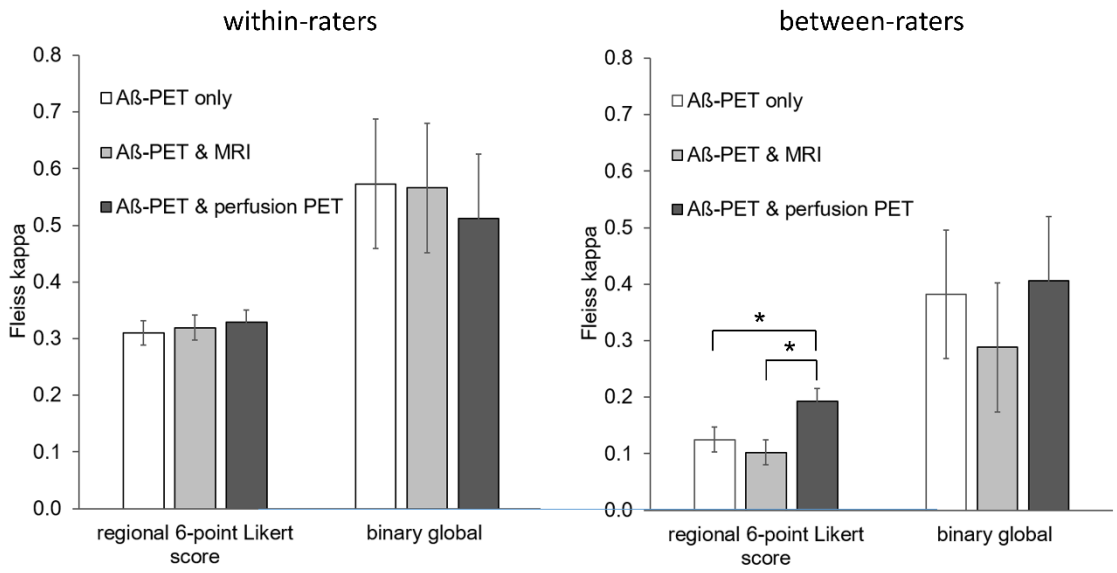


Figure 5 Left: Fleiss kappa of within-raters agreement regarding the regional visual Likert 6-point score (across all ROIs and across all readers) and the global binary visual score (across all readers). **Right:** Fleiss kappa of between-raters agreement for the second scoring session regarding the regional visual Likert 6-point score (across all ROIs) and the global binary visual score. The error bars represent the 83.4%-confidence interval. (*: $p < 0.05$)

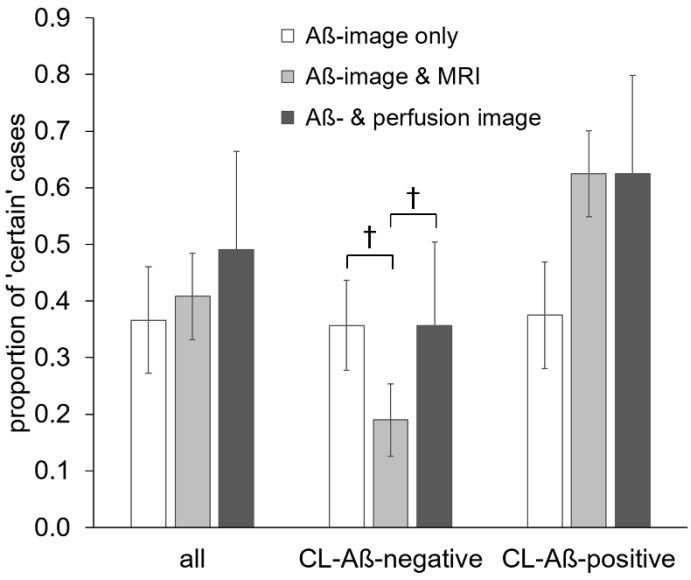
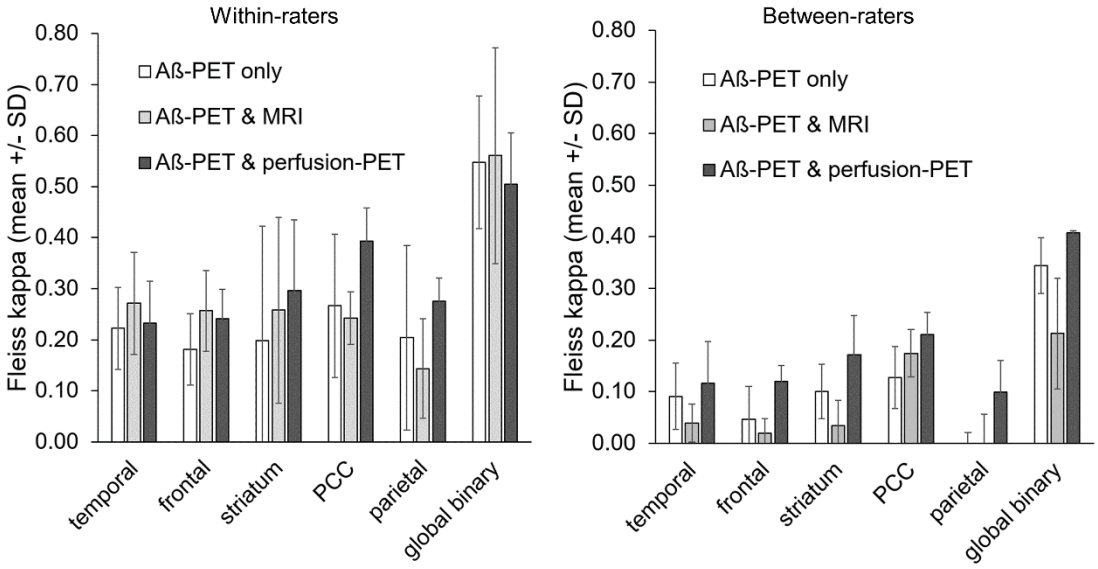


Figure 6 Proportion of cases with 'certain' global binary visual score (i.e., all 6 reads either 'visually Aß-positive' or 'globaly Aß-negative'). (†: $p < 0.10$)

Supplementary Material



Supplementary Figure 1 Fleiss kappa of within-rater (left) and between-raters (right) variability of the regional visual Likert 6-point score for the different ROIs and of the global binary visual interpretation. Within-raters kappa values were averaged over the 3 raters and over left and right hemisphere (ROIs only). Between-raters kappa values were averaged over both scoring sessions and both hemispheres (ROIs only).