AJNR

Concordance between Centiloid quantification and visual interpretation of amyloid PET scans across the Alzheimer's disease continuum

Mohammad Khalafi, Seyed Hani Hojjati, Xiuyuan H. Wang, Liangdong Zhou, Anna S. Nordvig, Yi Li, Tracy A. Butler, Qolamreza R. Razlighi, Emily B. Tanzi, Silky Pahlajani, Lidia Glodzik, Nancy S. Foldi, Mony J. de Leon and Gloria C. Chiang

This represents the accepted version of the manuscript and also includes the supplemental material; it differs from the final version of the article.

Concordance between Centiloid quantification and visual interpretation of amyloid PET scans across the Alzheimer's disease continuum

Mohammad Khalafi, Seyed Hani Hojjati, Xiuyuan H. Wang, Liangdong Zhou, Anna S. Nordvig, Yi Li, Tracy A. Butler, Qolamreza R. Razlighi, Emily B. Tanzi, Silky Pahlajani, Lidia Glodzik, Nancy S. Foldi, Mony J. de Leon, Gloria C. Chiang

ABSTRACT

BACKGROUND AND PURPOSE: Accurate identification of cerebral beta-amyloid (Aß) accumulation is crucial for diagnosing Alzheimer's disease (AD) and determining eligibility for anti-Aß therapies. The Centiloid (CL) scale has emerged as a standardized method to harmonize Aß positron emission tomography (PET) quantification across different tracers and sites. We aimed to evaluate the concordance between CL quantification and visual interpretation in a cohort of cognitively impaired (CI) and unimpaired (CU) participants who underwent Aß PET.

MATERIALS AND METHODS: Two hundred twenty-one participants (mean age 69 ± 12.3 years) were prospectively enrolled in AD studies and underwent 247 Aß PET scans, including 157 with [^{11}C]Pittsburgh Compound B (PiB) and 90 with [^{18}F]Florbetaben (FBB). Standardized uptake value ratios (SUVRs) were converted to the CL scale following Global Alzheimer's Association Interactive Network (GAAIN) guidelines. Percent agreement and kappa statistics were used to evaluate the concordance between CL thresholds and visual interpretation in determining Aß positivity.

RESULTS: The highest concordance rate for the whole cohort was 93% using a CL cutoff of 18 (kappa coefficient 0.84). Using FBB, the concordance rate was highest using a CL cutoff of 24 (97%), whereas the concordance rate for PiB peaked at 94% at a CL cutoff of 18. Concordance was higher in negative than positive Aß PET cases, 98% versus 90%. Concordance was slightly higher in CI participants, compared to CU (96% versus 93%). Disagreement commonly occurred when focal areas of Aß positivity were identified on visual interpretation but did not meet the threshold globally by CL quantification.

CONCLUSIONS: Global CL quantification of Aß PET scans is highly concordant with visual interpretation. Combining both methods may provide a more complete assessment of the extent of Aß deposition in the brain.

ABBREVIATIONS: AD = Alzheimer's disease; Aß = beta-amyloid; PET = positron emission tomography; PiB = Pittsburgh Compound B; FBB = Florbetaben; CL = Centiloid.

SUMMARY SECTION

PREVIOUS LITERATURE: Amyloid PET imaging plays a crucial role in accurate diagnosis of Alzheimer's disease and selection for anti-amyloid therapy. Both visual interpretation and quantification methods have been reported in the literature. Visual interpretation is widely used clinically due to accessibility and efficiency, but it can be subjective and prone to inter-reader variability. The Centiloid scale was developed to provide standardized and reproducible assessments to compare amyloid burden across sites and aid in tracking disease progression. The integration of both approaches may improve diagnostic accuracy and ensure better clinical decision-making.

KEY FINDINGS: We found high concordance between visual interpretation and Centiloid quantification, particularly for diffusely positive cases, which were more commonly seen in cognitively impaired individuals. Discordant cases were primarily observed with focal amyloid deposition that were detected visually but below the global Centiloid threshold for positivity.

KNOWLEDGE ADVANCEMENT: This study highlights the importance of integrating visual and quantitative assessment of amyloid PET scans. Even with Centiloid quantification, visual reads remain important, particularly in identifying individuals with focal amyloid deposition at earlier disease stages, who may still benefit from anti-amyloid therapy.

INTRODUCTION

A key pathologic hallmark of Alzheimer's disease (AD) is the presence of beta-amyloid (Aβ) plaques in the brain, which begin to accumulate years before the onset of clinical symptoms. These extracellular plaques are believed to lead to neurofibrillary tangle formation, neuronal dysfunction, and synaptic loss, ultimately leading to the cognitive decline observed in AD [1, 2]. Aβ positron emission tomography (PET) imaging has revolutionized our ability to visualize Aβ plaques in vivo [3-7] in both cognitively impaired (CI) and cognitively unimpaired (CU) individuals [3, 4]. Importantly, Aβ PET has enhanced our ability to identify preclinical AD, stratify for clinical trials, diagnose patients early, and select patients for anti-Aβ therapies [1, 2].

In clinical practice, Aβ PET studies are interpreted visually by radiologists who assess for Aβ tracer uptake in the cerebral cortex[1, 2, 8]. Guidelines for interpreting the three FDA-approved Aβ PET radiotracers are similar in assessing for signal, extending to the cortical margin, that is as intense or greater than adjacent white matter radioactivity, although the number of regions to assess, size requirement of positive areas, and image display color scale differ slightly [9-11]. However, visual interpretation can be prone to inter-reader variability, which has been reported to result in disagreement among readers in 6-22% of cases, due to differences in experience and training [8, 12, 13]. This variability highlights the need for a quantitative method alongside visual interpretation to assess the Aβ burden [14,15-17].

Prior studies have compared visual interpretation of Aβ PET scans with quantitative assessment using standardized uptake value ratios (SUVR), although these can be affected by technical parameters, including scanner variation, differences in image acquisition parameters, and selection of reference regions [18-20]. More recently, the Centiloid (CL) project was led by Klunk et al [17] to develop a standardized, quantitative measure of Aβ deposition, which would allow for direct comparison of Aβ burden across imaging centers and different tracers, using a scale of 0 to 100 [15-17]. Using this standardized CL metric mitigates variability and enhances comparability of results across sites and Aβ tracers, which can be particularly valuable in multi-center and multi-tracer research studies and clinical trials [15, 16].

As quantitative tools become adopted into clinical practice, it is crucial to understand instances in which visual interpretation and quantification may be discordant, since this could impact treatment decisions in this era of anti-Aβ therapy. Therefore, in our cohort we had two major aims: first, to assess the concordance between visual interpretation and CL quantification of Aβ PET scans, and second, to explore reasons for discordance, including whether this occurred more commonly in positive or negative cases, or CI versus CU individuals.

MATERIALS AND METHODS

Study design

Two hundred twenty-three individuals who had undergone Aβ PET in prospectively enrolled NIH-funded studies at the Weill Cornell Medicine (WCM) Brain Health Imaging Institute between 2019 and 2024 were included in this study. Two were excluded due to withdrawal of consent, declining further participation, and evidence of traumatic brain injury. The Standards for Reporting of Diagnostic Accuracy Studies (STARD) checklist was followed, and the flowchart is shown in the **Online Supplemental File**. Our final cohort consisted of 221 individuals (154 CU, 67 CI). Two hundred and forty-seven Aβ PET studies including 221 cross-sectional and 26 longitudinal scans were obtained, 90 using the Aβ PET tracer, FBB, and 157 using PiB. All projects complied with ethical standards for human research and were approved by the WCM Institutional Review Board (IRB). All participants provided written informed consent before participation. Each participant underwent screening, detailed interviews, and clinical and neuropsychological assessments. Cognitive impairment was determined based on the Montreal Cognitive Assessment, Clinical Dementia Rating, and clinical consensus conferences based on National Alzheimer's Coordinating Center guidelines.

MR imaging acquisition

All participants underwent brain MRI scans on 3T scanners (Siemens and GE) equipped with 64-channel head/neck receiver coils. High-resolution structural images were obtained using a standardized 3D T1-weighted sequence. For Siemens scanners, parameters included a repetition time (TR) of 2400 ms, echo time (TE) of 2.96 ms, flip angle of 9°, field of view (FOV) of 256 × 256 mm, matrix size of 512 × 512, isotropic voxel size of 0.5 × 0.5 × 0.5 mm³, and 416 sagittal slices. On GE scanners, imaging was performed with a TR of 8.23 ms, TE of 3.2 ms, flip angle of 12°, FOV of 256 × 256 mm, matrix size of 512 × 512, isotropic voxel size of 0.5 × 0.5 × 0.5 mm³, and 312 sagittal slices.

Aβ PET imaging acquisition

Aβ PET imaging was performed on a Siemens Biograph mCT (64-slice) PET/CT scanner. The PiB radiotracer was synthesized by the WCM core radiochemistry facility. PiB PET data were acquired in list mode 50 to 70 minutes after a 10-second intravenous bolus injection of approximately 555 MBq (15 mCi) of PiB. The images were reconstructed into a 400 × 400 × 109 matrix, with voxel dimensions of 1.018 × 1.018 × 2.027 mm, in 5-minute frames. For FBB PET imaging, an intravenous injection of approximately 300 MBq (8.1 mCi) of FBB was administered. Images were acquired from 90 to 110 minutes after injection and reconstructed into four 4×5-minute frames. The images were reconstructed into a 400 × 400 × 55 matrix, and the voxel size was 2.036 × 2.036 × 3 mm.

Quantification using the Centiloid scale

The Global Alzheimer's Association Interactive Network (GAAIN) has played a crucial role in standardizing the quantification of cerebral Aβ deposition on PET across tracers and sites using the CL scale [17]. This standardization approach utilizes a 100-point scale, in which 0 represents the average Aβ burden in Aβ-negative controls, and 100 represents the average Aβ burden in mild-to-moderate AD patients. However, the full range of the scale is not limited. The GAAIN database provides calibrated CL thresholds for Aβ PET tracers,

including PiB and FBB, allowing for direct comparison and conversion among different tracers across studies and centers [17].

Following the GAAIN guidelines[17], we first calculated the global SUVR for Aβ PET by assessing tracer uptake in key cortical gray matter regions, including the frontal cortex, temporal cortex, parietal cortex, cingulate cortex, and precuneus [21]. Specifically, we used the standard global cortical target (CTX) mask for both tracers and the whole cerebellum mask as the reference region, both provided on the GAAIN website to ensure standardized quantification [17]. For the PIB tracer, we replicated all the steps described as the Level 1 analysis protocol in the Klunk et al. paper [17]. For the FBB tracer, first, dynamic frames were realigned using rigid-body registration and averaged to obtain a composite volume to correct motion artifacts. Next, the frames were co-registered to T1-weighted MR images in FreeSurfer space using the normalized mutual information method in FSL [22]. The resulting images were then transformed from each participant's FreeSurfer space to MNI152 space using Advanced Normalization Tools (ANTs) [23]. Finally, we calculated the standardized uptake value ratio (SUVR) by calculating the standardized uptake value (SUV) using the standard CTX mask and normalized to the whole cerebellum mask [17]. All results underwent quality control using data from the GAAIN website, following the guidelines suggested in Klunk et al. [17] and the quality metrics met the expected standards—e.g., slope ranging from 0.98 to 1.02, intercept between -2 and 2 CL, and an R^2 value exceeding 0.98. We derived the equations to convert PIB and FBB SUVR values to CL units directly:

$$CL = 100 * (CTX\_SUVR - 1.014513) / 1.092145$$

Moreover, the level 2 calibration equation for the conversion of FBB SUVR to PiB SUVR is:

$$PiB\ SUVR = (CTX\_SUVR\_FBB - 0.5147) / 0.5936$$

Visual interpretation of Aβ PET studies

The Aβ PET images were read by a board-certified radiologist with subspecialty certification in neuroradiology and more than 14 years of experience reading brain Aβ PET studies (G.C.), who had completed required reader training for FBB PET and was blinded to all clinical information and quantitative metrics. The standard clinical method for PET scan interpretation was applied as described [10]. Briefly, images were viewed in the axial plane using gray scale. The images were reviewed systematically, starting at the cerebellum and scrolling upward through the lateral temporal and frontal lobes, the posterior cingulate cortex/precuneus, and the parietal lobes. If the majority of slices in at least one of these regions (i.e. lateral temporal, frontal, posterior cingulate cortex/precuneus, parietal) showed cortical gray matter signal intensity equal to or greater than adjacent white matter signal intensity, the scan was considered positive.

The first set of Aβ PET scans obtained at our Institute between 2019 and 2022 were independently double-read in batch by another radiologist (Y.L.), who is trained in nuclear medicine and also has more than 14 years of experience interpreting brain Aβ PET studies. This was performed for internal quality assurance to confirm high reader accuracy in our research studies. Subsequent Aβ PET studies, performed between 2022 and 2024, included in this analysis were visually interpreted prospectively by the first reader. CL quantification was performed in batch on all available Aβ PET scans performed at our Institute at the time of drafting this manuscript, with the objective of comparing this new standardized quantification metric with our visual reads.

Centiloid thresholds and statistical analysis

All statistical analyses were conducted using R software (version 4.4.1, R Core Team). Several CL cutoffs were considered for these analyses based on the literature. A prior study reported that a CL cutoff between 12 and 20 best predicted future Aβ accumulation for enrollment in secondary prevention trials [24]. Another study reported that a CL cutoff of 12 best predicted clinical progression on the Clinical Dementia Rating scale [25], whereas another showed that a CL cutoff of 24 was most concordant with pathologic findings of Aβ [26]. As a result, the following CL thresholds were tested in our analyses: 12, 16, 18, 20, and 24, as a sensitivity analysis.

To assess concordance between CL quantification thresholds and visual interpretation of Aβ PET studies, we used Cohen's kappa and chi-squared tests. Cohen's kappa coefficient was calculated to assess the level of agreement between the CL threshold-based Aβ PET classification and visual interpretation. A kappa value of 0.6 or higher indicates substantial agreement. The psych package in R was used to compute the kappa coefficient [27]. Also, a chi-squared test was performed to determine whether there was a statistically significant association between the classification based on the CL threshold and visual interpretation. This test was executed using the 'chisq.test' function in R. Additionally, to assess whether patient characteristics were associated with agreement between visual PET interpretation and CL quantification, we conducted a logistic regression analysis. Agreement was used as the dependent variable, with age, sex, and clinical status included as independent predictors. All p-values reported are two-tailed, and values below 0.05 were considered statistically significant [28]. We also performed subgroup analyses to assess whether agreement was higher in 1) positive versus negative Aβ PET cases, and in 2) CI or CU cases. Finally, we used receiver operating characteristic (ROC) curves, using visual reads as the reference standard, to comprehensively assess the range of CL cutoffs for assessing positivity.
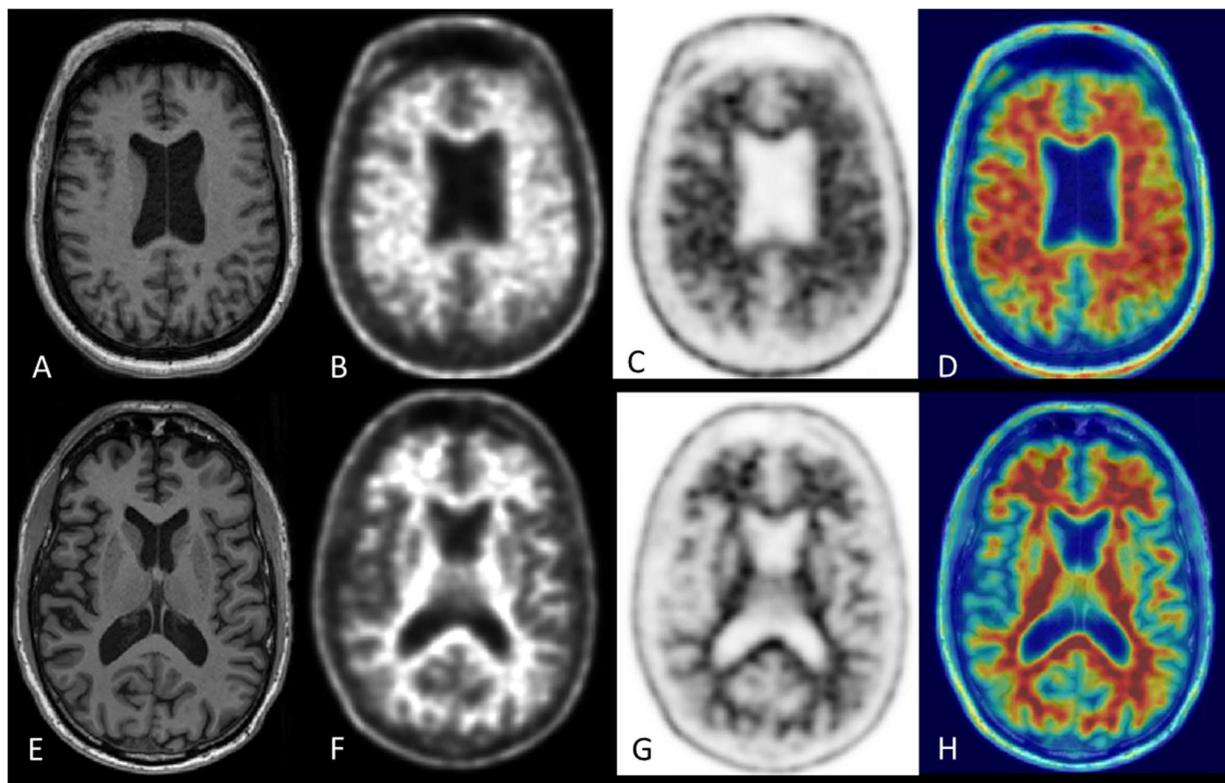
FIG 1. Concordant cases of [18F]Florbetaben PET, interpreted as positive (A-D) and negative (E-H) Aß PET scans, both visually and using a Centiloid threshold of 18. Axial 3D T1 MPRAGE images through the ventricles (A, E) and [18F]Florbetaben PET scans, shown in grayscale (B, F), inverse grayscale (C, G), and rainbow fused with the coregistered T1 (D, H).
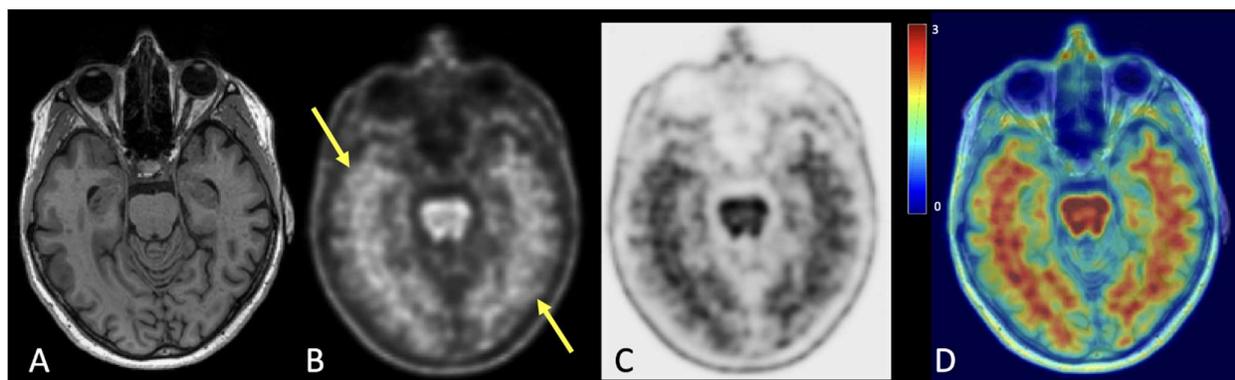

FIG 2. Discordant case of [18F]Florbetaben PET, interpreted as a positive Aß PET scan visually, but negative using the Centiloid quantification pipeline. Axial 3D T1 MPRAGE image through the temporal lobes (A) and [18F]Florbetaben PET, shown in grayscale (B), inverse grayscale (C), and rainbow fused with the coregistered T1 (D). This study was visually interpreted as a positive Aß PET scan due to confluent areas of cortical tracer uptake in the temporal lobes (yellow arrows). However, the study was quantified as 11 Centiloid units, below most thresholds for Aß positivity.

RESULTS

Baseline demographic information is shown in **Table 1**. There were no differences in sex distribution or years of education between the participants who underwent PiB PET versus FBB PET. Of note, the PiB PET group was older (p=0.003) and had more cognitively impaired participants (p<0.001). There was also no difference in proportion of ApoE ε4 carriers (p=0.23). The distribution of CL units, stratified by visual read are shown in the **Online Supplemental File**.

Centiloid quantification and visual interpretation of Aß PET studies are highly concordant

Examples of Aβ PET scans that had concordance between the visual reads and the quantitative assessment by CL threshold of 18 are shown in **Figure 1**. For the whole cohort, the highest concordance between visual interpretation and CL quantification was observed at a CL threshold of 18, yielding 93% agreement (κ = 0.84) **(Table 2)**. Similar concordance rates were found at CL thresholds of 20 and 24 (κ = 0.83, 93% agreement for both thresholds), indicating strong concordance between quantitative and visual assessments across these thresholds. When analyzed by tracer, the highest concordance for FBB PET occurred at CL 24, with 97% agreement (κ = 0.89), while for PiB PET, the strongest concordance was observed at CL 18, reaching 94% agreement (κ = 0.87). These findings suggest that optimal CL

thresholds may vary slightly between tracers yet remain highly consistent overall. Chi-squared test results also showed highly statistically significant associations between CL classification and visual interpretation (**Online Supplemental File**).

Across all tracers, concordance was higher for negative scans compared to positive ones **(Table 2)**. In the whole cohort, scans classified as negative by visual read showed 98% agreement with CL-based classification at thresholds of 20 or 24, with PiB PET demonstrating an even higher concordance of 99% at these levels. For positive scans, agreement rates were slightly lower but remained high, with 90% concordance across the whole cohort at CL thresholds of 12 or 16. This pattern was particularly evident for FBB PET, where 94% of visually positive scans aligned with CL-based classification at CL 24. These results underscore the reliability of CL quantification in capturing Aβ burden while highlighting slight variations in optimal thresholds across different tracers. The concordance range was wider in FBB PET scans compared to PiB PET scans, showing that FBB PET is slightly less sensitive at lower CL thresholds.

For quality assurance, a subset of our Aβ PET scans, comprising 47 cases, had been independently reviewed by a second radiologist, yielding a concordance rate of 98% between the two radiologists, with only a single discordant read, illustrated in the **Online Supplemental File**.

The ROC curve for the whole cohort is included in the **Online Supplemental File**, with additional ROC curves stratifying the analyses by CU versus CI. Notably, the optimal threshold using ROC analysis for a positive Aβ PET scan was also 18 CL.

Characterization of cases with discordance between visual reads and Centiloid quantification

Using a CL threshold of 18, there were 17 cases of discordance between visual reads and CL classification, and of those, seven were cases that used the FBB tracer and 10 that used PiB (**Online Supplemental File**). Of the seven discordant FBB PET cases, there was one case in which the visual read was positive due to small focal areas of cortical FBB uptake in the temporal lobes, but the CL value was only 11 (**Figure 2**). This 68-year-old participant had a clinical diagnosis of mild cognitive impairment, with a CDR of 0.5, and was an APOE ε4 heterozygote. The other 6 discordant FBB PET cases were visually read as negative, but the CL units were above 18 (**Online Supplemental File**). All six were cognitively normal with a CDR of 0. They ranged in age from 53 to 72 years. One was an APOE ε4 heterozygote and one was an APOE ε2 heterozygote; the other four had an APOE ε3/3 genotype.

Ten discordant cases used PiB PET; eight were visually interpreted as positive but did not meet the CL threshold of 18. Focal regions of positivity were noted in the temporal lobes in 4 cases (**Online Supplemental File**), the parietal lobe in two cases, and both the temporal and parietal lobes in 1 case. In one case, there was a diffuse Aβ deposition, and the CL value of 17 was just below the set CL threshold of 18 for classification. These eight cases ranged in age from 56 to 85 years. Four were APOE ε4 heterozygotes, one of whom had clinical dementia, and four had an APOE ε3/3 genotype. Two of the discordant cases were visually read as negative but reached the CL threshold. One had atypical Aβ deposition in the left basal ganglia (**Online Supplemental File**). Both were cognitively normal. One had an APOE ε3/3 genotype, while the other was an APOE ε4 carrier.

Concordance between centiloid quantification and visual interpretation was slightly higher with cognitive impairment

The concordance rate was 96% in the CI cohort using a CL threshold of 18, compared to 93% in the CU cohort. Among CI participants, PiB demonstrated slightly higher concordance than FBB (96% versus 94%). For CU participants, FBB showed higher concordance than PiB (97% versus 94%), using a CL threshold of 24. Interestingly, the highest concordance rate for PiB in the CU participant group was at a low threshold of 16. (**Online Supplemental File**). In addition, males were less likely to have concordant scans (odds ratio = 0.24, p=0.009), possibly because they were less likely to be CI (p=0.07). Age was not associated with likelihood of concordant scans (p=0.80). No participants progressed to cognitive impairment during follow-up (**Online Supplemental File**).

DISCUSSION

Aβ deposition is a key pathological hallmark of AD, beginning years before clinical symptoms manifest, making early detection crucial, particularly now in this new era of monoclonal antibody therapy [1, 4]. Accurate interpretation of Aβ PET impacts both timely intervention and disease management, whether using visual or quantitative assessments of Aβ burden [4, 16]. Visual interpretation allows clinicians to determine the presence or absence of cerebral Aβ deposition based on regional and global patterns of tracer uptake, whereas quantitative measures like CLs offer a standardized approach to quantify Aβ load across different tracers and sites. [16, 29]. The CL scale has recently gained traction for its potential to improve consistency in Aβ PET interpretation and facilitate comparison across studies. [16, 29-31]. Since both visual and quantitative assessments of Aβ PET are being widely adopted, we investigated the concordance between CL-based quantification and visual interpretation of Aβ PET imaging in our cohort of CI and CU participants. A key aim was to better understand when and why these two assessment methods could be discordant, to enhance diagnostic accuracy and reliability in both clinical and research settings.

Our first major finding was that expert visual interpretation and CL quantification were highly concordant, reaching 93% agreement across the whole cohort using a CL threshold of 18, up to 97% using FBB (CL threshold of 24), and up to 94% using PiB (CL threshold of 18). Two recent papers that compared CL with visual reads, using [18F]Flutemetamol reported 83% [32] and 95% agreement [8], using CL thresholds of 22 and 12, respectively; while our findings show similarly high agreement, some variability in concordance rates may be due to differences in cognitive status of the cohort. Notably, almost 30% of one of these cohorts had dementia, which typically shows diffuse Aβ deposition on PET and thus higher concordance, compared to our cohort that was predominantly at an earlier stage of AD [8]. In addition, similar to Zeydan et al.,[32] we found that many of the discordant cases occurred when the visual read was positive using PiB, but the CL units were below the threshold of 18. It is known that [18F] amyloid tracers, such as FBB and [18F]Flutemetamol have greater off-target, background uptake in the white matter, resulting in decreased conspicuity of the border between cortical gray and white matter [33-36] and

decreased sensitivity.

On the other hand, PiB typically shows greater differentiation between white matter and gray matter compared to other Aβ tracers due to its lower nonspecific binding in white matter and its high affinity for fibrillar Aβ plaques in the gray matter of AD patients [33-35]. As a result, it is easier to detect small focal areas of cortical Aβ deposition on PiB PET scans, which would render a scan positive visually. Notably, five of the eight discordant PiB PET scans at the CL 18 threshold had focal cortical Aβ in the temporal lobes. On the other hand, the CL units are global, so these small areas of Aβ-positivity would be averaged with other Aβ-negative brain regions, leading to a global CL value that is below the threshold. In clinical practice, this could lead to misclassifying an individual as Aβ-negative and could preclude them from anti-Aβ therapy, despite having early, focal Aβ deposition.

Unlike PiB, many of the discordant FBB PET scans were visually interpreted as negative but met the CL threshold for positivity—these cases were observed among the CU individuals; this again may be due to greater off-target binding in the white matter for FBB and FBB's narrower dynamic range [37, 38], making it more difficult to visually identify early areas of Aβ deposition. In our experience, coregistration with structural MRI can be helpful to confirm uptake in gray versus white matter.

Concordance between visual and quantitative assessments was higher for negative Aβ PET scans, up to 98% across the whole cohort, suggesting the marked utility of Aβ PET to rule out AD, especially for concordantly negative cases. Concordance was also slightly higher in the CI cohort, probably because they were more likely to be diffusely, rather than focally, positive on Aβ PET. Clinically, visual interpretation of Aβ PET scans has traditionally been used to assess Aβ deposition because it is quick and requires no advanced software or image processing capabilities [39] [14]. Among experienced readers, such as those in our study, concordance is very high. However, the primary drawback is subjectivity, leading to inter-reader variability, particularly among less-experienced readers. Visual interpretation also only provides a qualitative assessment—"positive" or "negative"—which makes it more difficult to track disease progression [14].

CL quantification can address the limitations of visual interpretation by offering a standardized, quantitative approach for Aβ PET assessment; it normalizes measurements across tracers and platforms, providing a continuous metric for precise, reproducible evaluations [14, 40]. The objectivity of a quantitative scale reduces inter-reader variability and allows for monitoring of disease progression and treatment efficacy [14, 41]. However, in our study, there were several cases in which the use of only global CL units can miss early, focal areas of cortical Aβ deposition. Particularly those who are CI could be misclassified as non-AD dementia, losing access to monoclonal antibody therapy. Therefore, a combined approach using both visual interpretation and Cl quantification is advised. Other drawbacks of CL quantification include the need for advanced software and expertise to develop these post-processing pipelines, potentially limiting its availability in some clinical settings. Vendors, including MIMneuro, have already started incorporating CL quantification into clinical software packages, allowing for ease of use [42].

Our proposed framework for reviewing Aβ PET cases includes three steps. First, the radiologist should provide a visual read per FDA guidelines for each tracer [9-11]. This allows for detection of focal areas of cortical Aβ deposition in key AD-relevant regions. Second, we recommend Centiloid quantification according to GAAIN guidelines, which will allow for a global quantitative assessment of beta-Aβ burden on a standardized scale, even if the patient is imaged on different scanners or with different tracers. Again, this is already being incorporated into vendor software for clinical use [42] and will be especially important for longitudinal quantitative monitoring on anti-Aβ therapy. For less experienced readers, this could also provide a useful adjunct to visual assessment. Finally, FDA guidelines specify that the area of positivity needs to be greater than a single gyrus or encompass the majority of slices in one of the key regions [9, 10]. Cases with focal hotspots that are below that threshold could be categorized as "borderline." Per the literature, this likely represents the earliest stages of Aβ deposition [8, 43], but is currently below the threshold for initiation of anti-Aβ therapy. These patients could obtain follow-up Aβ PET imaging to assess for Aβ accumulation.

Two prior studies have shown that Aβ quantified with CL units correlate with post-mortem evidence of Aβ deposition [44, 45]. One study reported that CL units below 12 excluded the presence of Aβ postmortem, whereas CL units above 24 detected intermediate-to-high Aβ deposition, with CL units between 12 and 24 denoting low levels of Aβ deposition [44]. Amadoru et al similarly reported thresholds of 10 and 30 for the lower and upper thresholds of Aβ negativity and positivity [45]. Our reported CL threshold of 18 for highest concordance between visual reads and CL quantification fits within this reported range for detection of low levels of Aβ postmortem.

LIMITATIONS

This study has several limitations. First, participants were recruited from multiple NIH-funded studies on aging and AD, encompassing the continuum from cognitively normal to MCI and dementia, but focused on the early stages of disease. Consequently, a large portion of the cohort was CU, which differs from the typical clinical population undergoing Aβ PET—as Aβ PET is not routinely performed in CU individuals. However, including this CU cohort allowed us to evaluate the utility and concordance of visual interpretation and CL quantification even in the earliest stages of Aβ accumulation, which may be relevant for research settings and dementia prevention trials. Furthermore, some clinical patients could have subjective memory concerns, which could place them earlier in the AD continuum. Second, as these participants were imaged within research studies, many underwent PiB PET, despite it not being FDA-approved due to the requirement for an on-site cyclotron. While PiB PET has been widely used in large, multicenter AD studies and clinical trials, its performance may therefore not directly translate to clinical practice. However, [18F]NAV4694 is a novel radiotracer with a dynamic range comparable to PiB that is advancing toward FDA approval and could provide similar diagnostic performance [46]. Additionally, there were notable differences between the FBB and PiB cohorts, including a lower number of CI participants imaged with FBB, which may have influenced our findings. Genetics could also potentially influence CL thresholds, particularly individuals with autosomal dominant AD. However, only one participant in our study was known to have a presenilin mutation. This PET scan was Aβ positive both visually and quantitatively, but our study is underpowered to assess how such mutations can affect thresholds. Finally, our study focused on defining thresholds to optimize diagnosis, but future work could investigate the optimal thresholds for predicting who would most benefit from

anti-amyloid therapies [47].

Table 1: Baseline demographic information.

	Whole cohort (n=221)	Participants who underwent Aß PET imaging using [18F] Florbetaben (n=90)	Participants who underwent Aß PET imaging using [11C]Pittsburgh Compound B (n=131)
Age (yr)	67.5 ± 12.5	65.2 ± 12.2	69.0 ± 12.4
Gender:			
Female/Male (n)	136/85	56/34	80/51
Education (years)	16.6 ± 2.4	16.5 ± 2.3	16.7 ± 2.5
Clinical diagnosis:			
Cognitively Unimpaired (n)	154	74	80
Cognitively Impaired (n)	67	16	51
Centiloid quantification value	20.2 ± 40.2	11.9 ± 34.7	25.8 ± 42.6
Race:			
Black (n)	20	7	13
Asian (n)	10	4	6
Caucasian (n)	181	74	107
Native Hawaiian (n)	1	0	1
Other (n)	6	3	3
Declined to answer (n)	3	2	1
Ethnicity:			
Hispanic (n)	12	4	8
Non-Hispanic (n)	207	86	121
Declined to answer (n)	2	0	2
ApoE ε4 carrier:noncarrier (n)† (% carrier)	51:130 (28%)	14:50 (22%)	37:80 (32%)

Data shown are means (± standard deviation) for continuous variables or the number of participants for categorical variables. n = number of participants. †ApoE ε4 genotyping was only available for 183 participants.

Table 2: Concordance between visual and Centiloid-based amyloid PET interpretation at different Centiloid (CL) cut-offs.

	Cut-off value for a positive Aß PET scan based on Centiloid quantification	Cohen's Kappa statistic (κ)	P-value	Number of concordant cases between Centiloid quantification and visual read (% agreement)	Number of cases that were classified as negative by Centiloid quantification (% agreement with visual read)	Number of cases that were classified as positive by Centiloid quantification (% agreement with visual read)
Whole cohort	CL 24	0.83	<0.0001	229 (93%)	165 (98%)	64 (81%)
	CL 20	0.83	<0.0001	229 (93%)	164 (98%)	65 (82%)
	CL 18	0.84	<0.0001	230 (93%)	160 (95%)	70 (89%)
	CL 16	0.83	<0.0001	228 (92%)	157 (93%)	71 (90%)
	CL 12	0.73	<0.0001	216 (87%)	145 (86%)	71 (90%)
[18F] Florbetaben PET scans	CL 24	0.89	<0.0001	87 (97%)	72 (97%)	15 (94%)
	CL 20	0.86	<0.0001	86 (96%)	71 (96%)	15 (94%)
	CL 18	0.76	<0.0001	83 (92%)	68 (92%)	15 (94%)
	CL 16	0.73	<0.0001	82 (91%)	67 (91%)	15 (94%)
	CL 12	0.51	<0.0001	72 (80%)	57 (77%)	15 (94%)
[11C]Pittsburgh Compound B PET scans	CL 24	0.79	<0.0001	142 (90%)	93 (99%)	49 (78%)
	CL 20	0.81	<0.0001	143 (91%)	93 (99%)	50 (79%)
	CL 18	0.87	<0.0001	147 (94%)	92 (98%)	55 (87%)
	CL 16	0.85	<0.0001	146 (93%)	90 (96%)	56 (89%)
	CL 12	0.83	<0.0001	144 (92%)	88 (94%)	56 (89%)

CONCLUSIONS

Our study demonstrated high concordance between visual and quantitative assessments of Aβ PET, particularly in negative Aβ PET studies and CI individuals. Discordant PiB cases usually occurred when visual interpretation identified focal, early areas of cortical Aβ deposition, which were averaged out in global CL units. Discordant FBB cases usually occurred when visual interpretation missed early Aβ deposition in cognitively normal individuals, possibly due to greater off-target white matter binding. Combining both qualitative and

quantitative methods may enhance diagnostic accuracy and reliability, as well as provide a robust method for monitoring disease progression.

ACKNOWLEDGMENTS

REFERENCES

1. Cummings J. Anti-Amyloid Monoclonal Antibodies are Transformative Treatments that Redefine Alzheimer's Disease Therapeutics. *Drugs* 2023;83:569-576

2. Karran E, De Strooper B. The amyloid hypothesis in Alzheimer disease: new insights from new therapeutics. *Nature Reviews Drug Discovery* 2022;21:306-318

3. Chen KT, Gong E, de Carvalho Macruz FB, et al. Ultra–low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology* 2019;290:649-656

4. Heeman F, Hendriks J, Lopes Alves I, et al. [11 C] PIB amyloid quantification: effect of reference region selection. *EJNMMI research* 2020;10:1-10

5. Okazawa H, Ikawa M, Jung M, et al. Multimodal analysis using [11 C] PiB-PET/MRI for functional evaluation of patients with Alzheimer's disease. *EJNMMI research* 2020;10:1-12

6. Sabri O, Seibyl J, Rowe C, Barthel H. Beta-amyloid imaging with florbetaben. *Clinical and translational imaging* 2015;3:13-26

7. van Berckel BN, Ossenkoppele R, Tolboom N, et al. Longitudinal amyloid imaging using 11C-PiB: methodologic considerations. *Journal of Nuclear Medicine* 2013;54:1570-1576

8. Collij LE, Salvadó G, Shekari M, 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;48:2169-2182

9. Amyvid. FDA-Approved Drugs: FDA; 2012. https://www.accessdata.fda.gov/drugsatfda_docs/label/2012/202008s000lbl.pdf

10. NEURACEQ. *FDA-Approved Drugs*. FDA; 2014. https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/204677s000lbl.pdf

11. Vizamyl™. *FDA-Approved Drugs*. FDA; 2013. https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/203137s008lbl.pdf

12. Takenaka A, Nihashi T, Sakurai K, et al. Interrater agreement and variability in visual reading of [18F] flutemetamol PET images. *Annals of Nuclear Medicine* 2024:1-9

13. Yamane T, Ishii K, Sakata M, et al. Inter-rater variability of visual interpretation and comparison with quantitative evaluation of 11 C-PiB PET amyloid images of the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) multicenter study. *European journal of nuclear medicine and molecular imaging* 2017;44:850-857

14. Pemberton HG, Collij LE, Heeman F, et al. Quantification of amyloid PET for future clinical use: a state-of-the-art review. *Eur J Nucl Med Mol Imaging* 2022;49:3508-3528

15. Su Y, Flores S, Hornbeck RC, et al. Utilizing the Centiloid scale in cross-sectional and longitudinal PiB PET studies. *Neuroimage Clin* 2018;19:406-416

16. Royse SK, Minhas DS, Lopresti BJ, et al. Validation of amyloid PET positivity thresholds in centiloids: a multisite PET study approach. *Alzheimer's Research & Therapy* 2021;13:99

17. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's & dementia* 2015;11:1-15. e14

18. Jeong YJ, Yoon HJ, Kang D-Y, Park KW. Quantitative comparative analysis of amyloid PET images using three radiopharmaceuticals. *Annals of Nuclear Medicine* 2023;37:271-279

19. The standard clinical method for PET S, Montandon ML, Lilja J, et al. PET amyloid in normal aging: direct comparison of visual and automatic processing methods. *Sci Rep* 2020;10:16665

20. Schmidt ME, Chiao P, Klein G, et al. The influence of biological and technical factors on quantitative analysis of amyloid PET: Points to consider and recommendations for controlling variability in longitudinal data. *Alzheimers Dement* 2015;11:1050-1068

21. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. *Neuroimage* 1996;4:153-158

22. Jenkinson M, Beckmann CF, Behrens TE, et al. FSL. *Neuroimage* 2012;62:782-790

23. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight j* 2009;2:1-35

24. Bollack A, Collij LE, García DV, et al. Investigating reliable amyloid accumulation in Centiloids: Results from the AMYPAD Prognostic and Natural History Study. *Alzheimers Dement* 2024;20:3429-3441

25. Quenon L, Collij LE, Garcia DV, et al. Amyloid-PET imaging predicts functional decline in clinically normal individuals. *Alzheimers Res Ther* 2024;16:130

26. Navitsky M, Joshi AD, Kennedy I, et al. Standardization of amyloid quantitation with florbetapir standardized uptake value ratios to the Centiloid scale. *Alzheimers Dement* 2018;14:1565-1571

27. Revelle W, Revelle MW. Package 'psych'. *The comprehensive R archive network* 2015;337:161-165

28. Okoye K, Hosseini S. Chi-Squared (X2) Statistical Test in R. *R Programming: Statistical Data Analysis in Research*: Springer; 2024:211-223

29. Haller S, Montandon M-L, Lilja J, et al. PET amyloid in normal aging: direct comparison of visual and automatic processing methods. *Scientific Reports* 2020;10:16665

30. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement* 2015;11:1-15.e11-14

31. Navitsky M, Joshi AD, Kennedy I, et al. Standardization of amyloid quantitation with florbetapir standardized uptake value ratios to the Centiloid scale. *Alzheimer's & Dementia* 2018;14:1565-1571

32. Zeydan B, Johnson DR, Schwarz CG, et al. Visual assessments of 11 C-Pittsburgh compound-B PET vs. 18 F-flutemetamol PET across the age

spectrum. *Nucl Med Commun* 2024;45:1047-1054

33.   Richards D, Sabbagh MN. Florbetaben for PET Imaging of Beta-Amyloid Plaques in the Brain. *Neurol Ther* 2014;3:79-88

34.   Su Y, Flores S, Wang G, et al. Comparison of Pittsburgh compound B and florbetapir in cross-sectional and longitudinal studies. *Alzheimers Dement (Amst)* 2019;11:180-190

35.   Lowe VJ, Kemp BJ, Jack CR, et al. Comparison of ^{18}F-FDG and PiB PET in Cognitive Impairment. *Journal of Nuclear Medicine* 2009;50:878-886

36.   Lowe VJ, Lundt E, Knopman D, et al. Comparison of [(18)F]Flutemetamol and [(11)C]Pittsburgh Compound-B in cognitively normal young, cognitively normal elderly, and Alzheimer's disease dementia individuals. *Neuroimage Clin* 2017;16:295-302

37.   Rowe CC, Doré V, Jones G, et al. (18)F-Florbetaben PET beta-amyloid binding expressed in Centiloids. *Eur J Nucl Med Mol Imaging* 2017;44:2053-2059

38.   Villemagne VL, Mulligan RS, Pejoska S, et al. Comparison of 11C-PiB and 18F-florbetaben for Aβ imaging in ageing and Alzheimer's disease. *Eur J Nucl Med Mol Imaging* 2012;39:983-989

39.   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-19s

40.   Kim SJ, Ham H, Park YH, et al. Development and clinical validation of CT-based regional modified Centiloid method for amyloid PET. *Alzheimers Res Ther* 2022;14:157

41.   Matsuda H, Soma T, Okita K, et al. Development of software for measuring brain amyloid accumulation using 18F-florbetapir PET and calculating global Centiloid scale and regional Z-score values. *Brain and Behavior* 2023;13:e3092

42.   Hall J. FDA Clears PET-Based Software for Centiloid Scaling of Amyloid Plaque Density. *SPOTLIGHT-RSNA*; 2024

43.   Murray ME, Lowe VJ, Graff-Radford NR, et al. Clinicopathologic and 11C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum. *Brain* 2015;138:1370-1381

44.   La Joie R, Ayakta N, Seeley WW, 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;15:205-216

45.   Amadoru S, Doré V, McLean CA, et al. Comparison of amyloid PET measured in Centiloid units with neuropathological findings in Alzheimer's disease. *Alzheimer's Research & Therapy* 2020;12:22

46.   Rowe CC, Pejoska S, Mulligan RS, 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;54:880-886

47.   Rabinovici GD, Knopman DS, Arbizu J, et al. Updated appropriate use criteria for amyloid and tau PET: A report from the Alzheimer's Association and Society for Nuclear Medicine and Molecular Imaging Workgroup. *Alzheimer's & Dementia* 2025;21:e14338

SUPPLEMENTAL FILES

SUPPLEMENTAL FIGURES

Supplemental Figure-1 Flowchart depicting participant inclusion and classification using 18 Centiloids as the test for positivity using the visual read as the reference standard, based on Standards for Reporting Diagnostic Accuracy studies (STARD) guidelines [1].
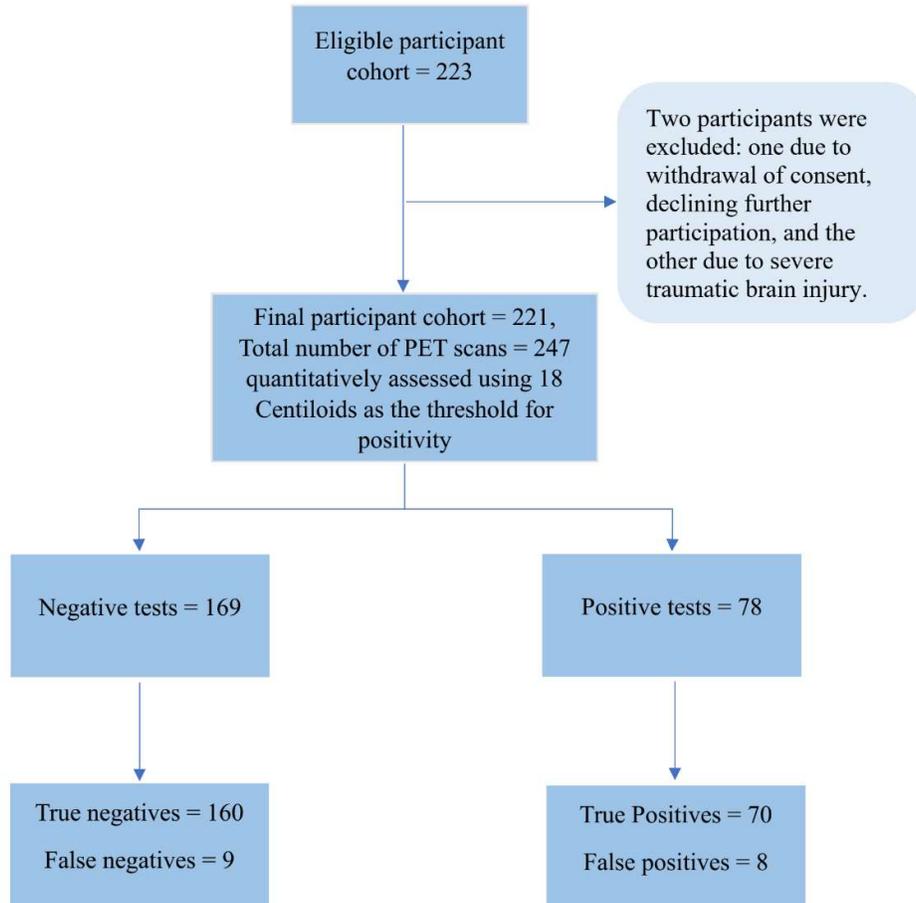
Supplemental Figure 2 Distribution of Centiloid values based on the visual amyloid PET read: A) Distribution of Centiloid values for the entire cohort. B) Distribution of Centiloid values for participants who underwent [11C]Pittsburgh Compound B PET. C) Distribution of Centiloid values for participants who underwent [18F]Florbetaben PET.

A)
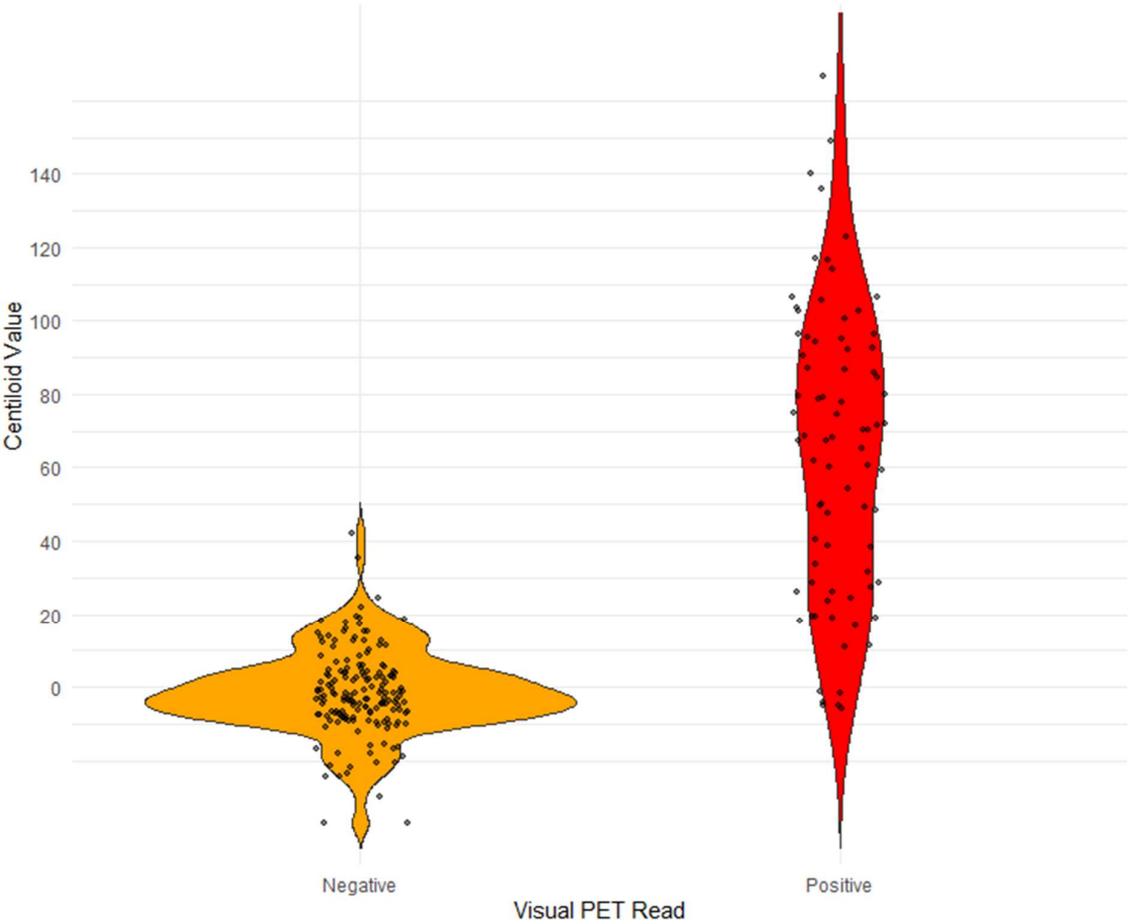
B)
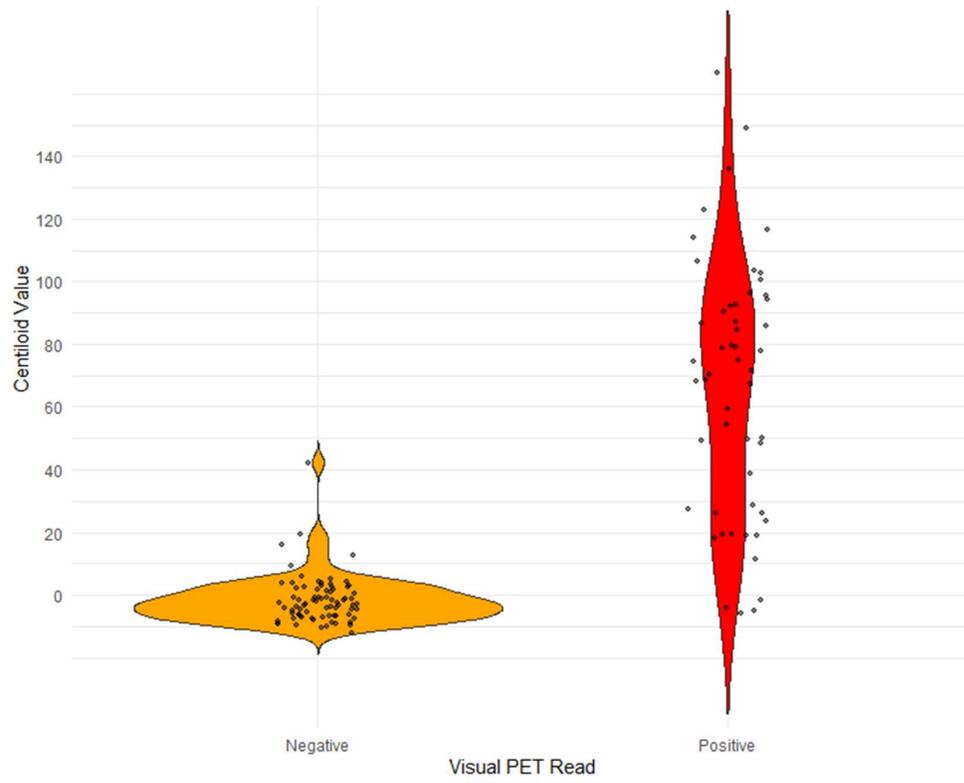
C)
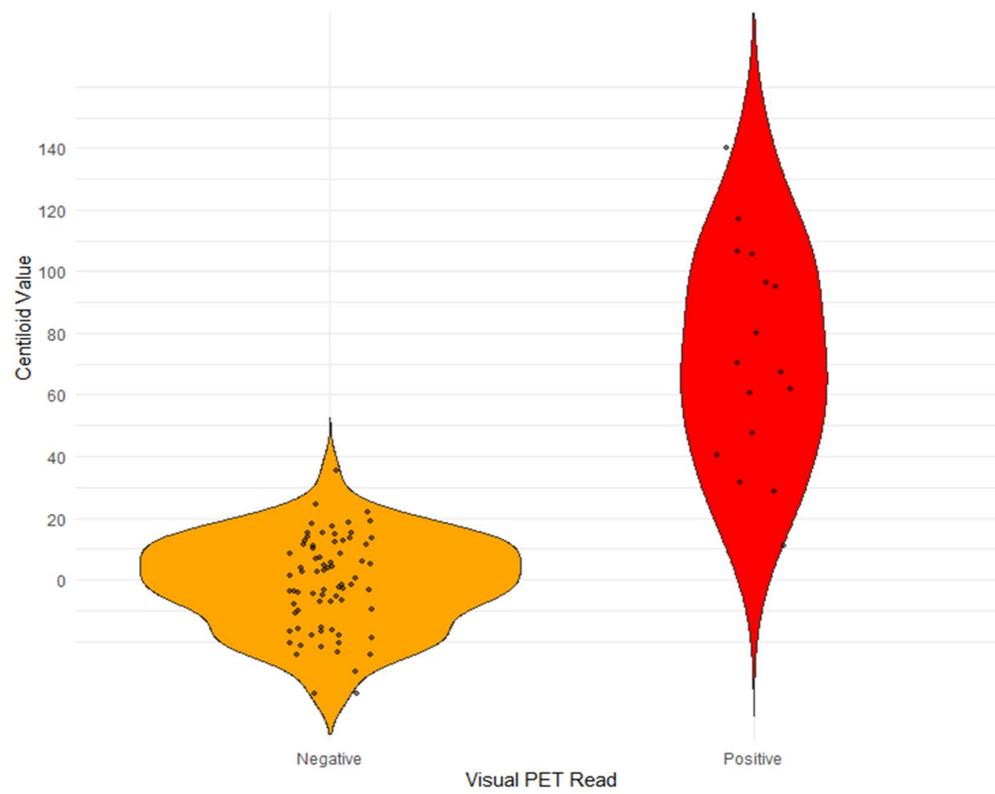
Supplemental Figure 3 Receiver operating characteristic (ROC) curves illustrating the area under the curve (AUC), sensitivity, specificity, and cut-off points for Centiloid-assessed PET quantification versus visual interpretation are presented for three groups: A) the entire cohort, B) cognitively impaired individuals, and C) cognitively unimpaired individuals.

A)

Centiloid-Assessed PET Quantification vs Visual Interpretation Across the Whole Cohort

AUC: 0.948
Optimal Threshold: 18.27
Sensitivity: 0.886
Specificity: 0.958

B)

Centiloid-Assessed PET Quantification vs Visual Interpretation Across the Cognitively Impaired Individuals

AUC: 0.983
Optimal Threshold: 17.498
Sensitivity: 0.927
Specificity: 1

C)

Centiloid-Assessed PET Quantification vs Visual Interpretation Across the Cognitively Unimpaired Individuals

AUC: 0.926
Optimal Threshold: 18.27
Sensitivity: 0.893
Specificity: 0.944

Supplemental Figure 4. Discordant case of [18F]Florbetaben PET, interpreted as a negative Aß PET scan visually, but positive by quantification (35 Centiloid units). Axial 3D T1 MPRAGE image through the frontal and parietal lobes (A) and [18F]Florbetaben PET, shown in grayscale (B), inverse grayscale (C), and rainbow fused with the coregistered T1 (D). This study was visually interpreted as negative. However, there is confluent cortical tracer uptake primarily in the left hemisphere (yellow arrow).

Supplemental Figure 5. Discordant case of [11C]Pittsburgh compound-B PET, interpreted as a positive Aß PET scan visually, but negative by quantification (17 Centiloid units). Axial 3D T1 MPRAGE image through the temporal lobes (A) and [11C]Pittsburgh compound B PET, shown in grayscale (B), inverse grayscale (C), and rainbow fused with the coregistered 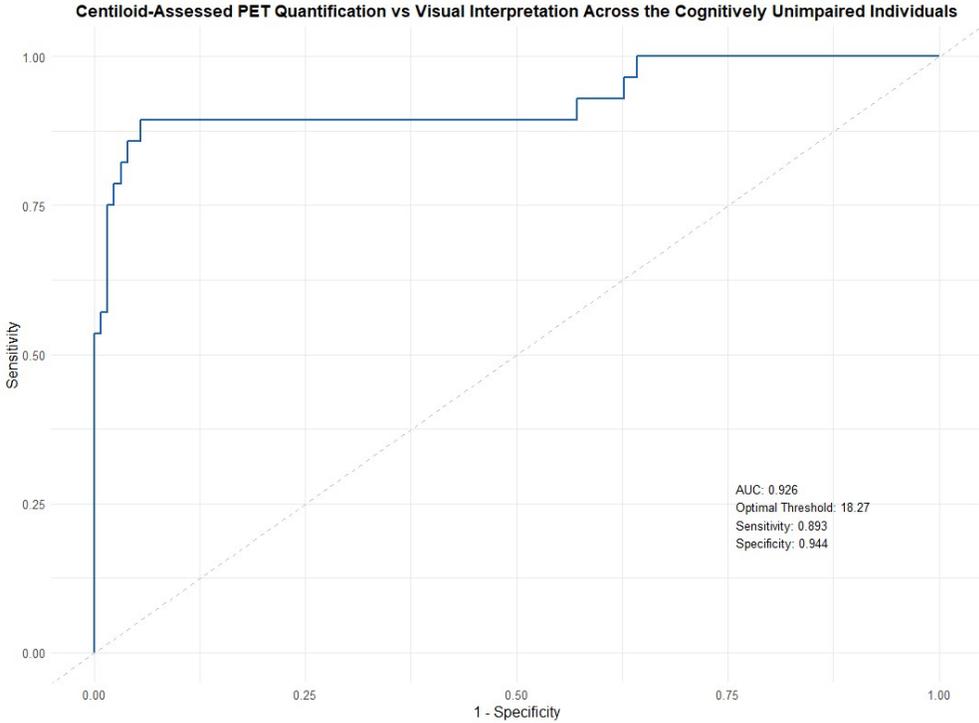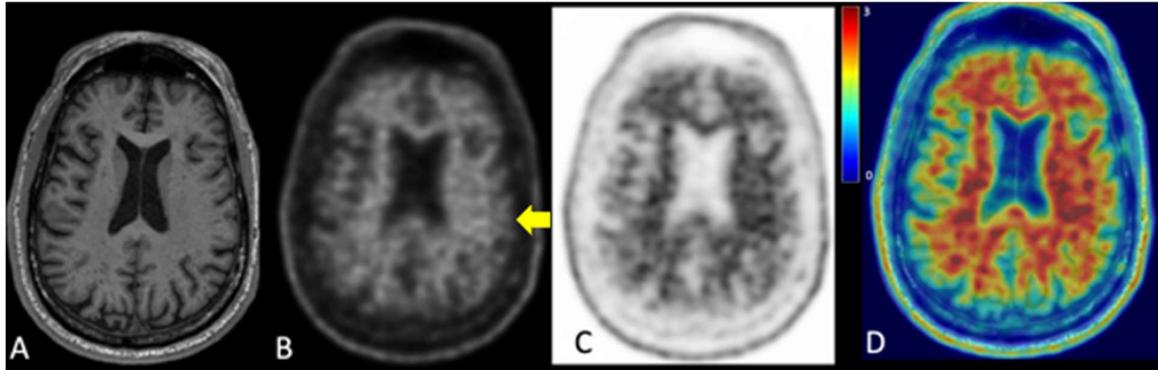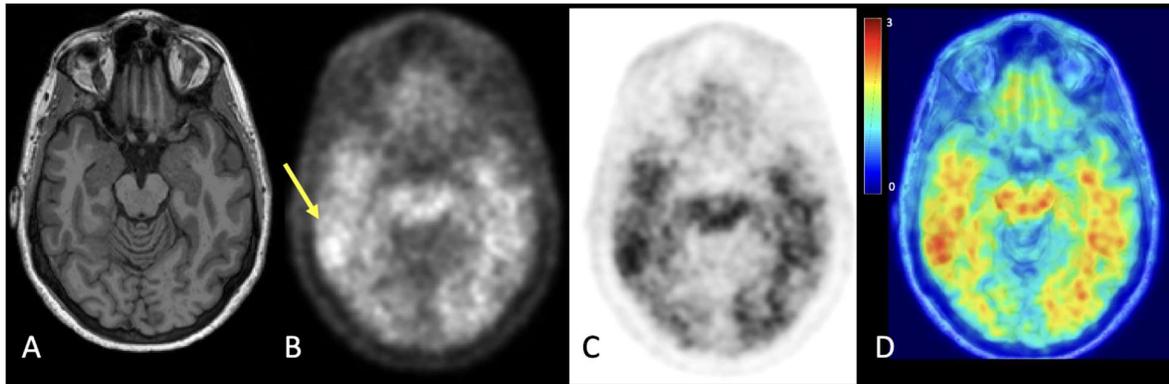T1 (D). This study was visually interpreted as positive due to several areas of confluent cortical tracer uptake, most prominently in the right temporal lobe (yellow arrow). Of note, a threshold of 17 Centiloid units is close to most thresholds for Aß positivity, including our reported optimal threshold of 18 Centiloid units.

Supplemental Figure 6. Discordant case of [11C]Pittsburgh compound B (PiB) PET, initially interpreted as a negative Aß PET scan visually by one reader and positive by another reader, due to differences in whether the majority of the temporal lobe region showed uptake. This PET was considered positive by quantification (42 centiloid units). Axial 3D T1 MPRAGE image through the temporal lobes (A) and [11C]Pittsburgh compound B PET, shown in grayscale (B), inverse grayscale (C), and rainbow fused with the coregistered T1 (D). There were focal areas of cortical tracer uptake in the temporal lobes (yellow arrows) and an atypical area of tracer uptake in the left basal ganglia (double arrows). Of note, this was the only case in our series for which a PiB PET scan was interpreted as negative by one reader, but positive by quantification.
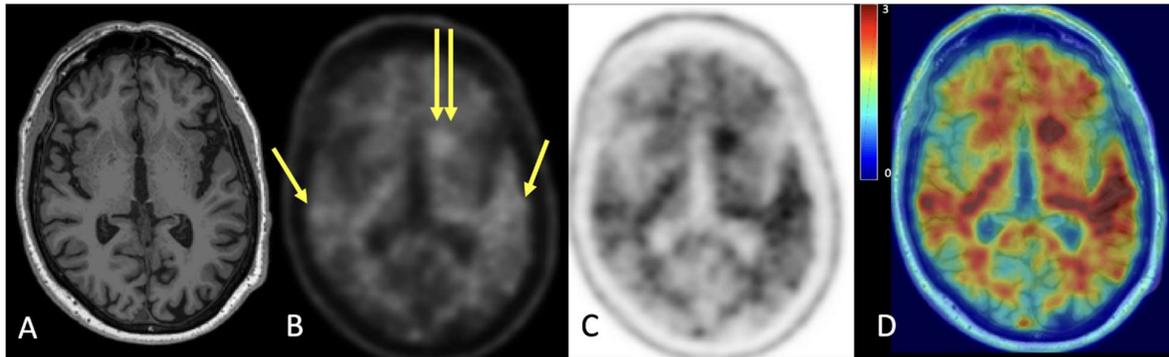
Supplemental Table 2: Concordance between visual and Centiloid-based amyloid PET interpretation at different Centiloid cutoffs, stratified by cognitive status.

	Cut-off value for a positive Aß PET scan based on Centiloid quantification	Cohen's Kappa statistic	P-value	Number of concordant cases between Centiloid quantification and visual read (% agreement)
All Cognitively impaired participants (n=67)	CL 24	0.88	<0.0001	63 (94%)
	CL 20	0.88	<0.0001	63 (94%)
	CL 18	0.91	<0.0001	64 (96%)
	CL 16	0.88	<0.0001	63 (94%)
	CL 12	0.84	<0.0001	62 (93%)
Cognitively impaired participants who underwent [18F]Florbetaben PET (n=16)	CL 24	0.88	<0.0001	15 (94%)
	CL 20	0.88	<0.0001	15 (94%)
	CL 18	0.88	<0.0001	15 (94%)
	CL 16	0.88	<0.0001	15 (94%)
	CL 12	0.75	<0.0001	14 (88%)
Cognitively impaired participants who underwent [11C]Pittsburgh Compound B PET (n=51)	CL 24	0.88	<0.0001	48 (94%)
	CL 20	0.88	<0.0001	48 (94%)
	CL 18	0.92	<0.0001	49 (96%)
	CL 16	0.88	<0.0001	48 (94%)
	CL 12	0.88	<0.0001	48 (94%)
All Cognitively unimpaired participants (n=154)	CL 24	0.75	<0.0001	143 (93%)
	CL 20	0.76	<0.0001	143 (93%)
	CL 18	0.76	<0.0001	142 (92%)
	CL 16	0.76	<0.0001	142 (92%)
	CL 12	0.61	<0.0001	132 (86%)
Cognitively unimpaired participants who underwent [18F]Florbetaben PET (n=74)	CL 24	0.86	<0.0001	72 (97%)
	CL 20	0.80	<0.0001	71 (96%)
	CL 18	0.66	<0.0001	68 (92%)
	CL 16	0.62	<0.0001	67 (91%)
	CL 12	0.38	<0.0001	58 (78%)
Cognitively unimpaired participants who underwent [11C]Pittsburgh Compound B PET (n=80)	CL 24	0.69	<0.0001	71 (89%)
	CL 20	0.73	<0.0001	72 (90%)
	CL 18	0.81	<0.0001	74 (93%)
	CL 16	0.84	<0.0001	75 (94%)
	CL 12	0.81	<0.0001	74 (93%)

CL = Centiloids

Supplemental Table 2: The Chi-squared test results for each Centiloid cutoff for all cognitively impaired participants, stratified by participants who underwent [18F]Florbetaben PET and those who underwent [11C]Pittsburgh Compound B PET. P-values show whether there was a statistically significant association between the classification based on the Centiloid threshold and visual interpretation. CL = Centiloid.

	Centiloid assessed cut-offs	Chi-squared	P-value (CS)
All Cognitively impaired participants (n=67)	CL 24	48.816	<0.0001
	CL 20	48.816	<0.0001
	CL 18	51.962	<0.0001
	CL 16	48.030	<0.0001
	CL 12	44.302	<0.0001
Cognitively impaired participants who underwent [18F]florbetaben PET (n=16)	CL 24	9.143	<0.0001
	CL 20	9.143	<0.0001
	CL 18	9.143	<0.0001
	CL 16	9.143	<0.0001
	CL 12	6.132	<0.0001
Cognitively impaired participants who underwent [11C]Pittsburgh Compound B PET (n=51)	CL 24	36.307	<0.0001
	CL 20	36.307	<0.0001
	CL 18	39.475	<0.0001
	CL 16	35.536	<0.0001
	CL 12	35.536	<0.0001

Supplemental Table 3: The Chi-squared test results for each Centiloid cutoff for all cognitively unimpaired participants, stratified by participants who underwent [18F]Florbetaben PET and those who underwent [11C]Pittsburgh Compound B PET. P-values show whether there was a statistically significant association between the classification based on the Centiloid threshold and visual interpretation.

	Centiloid assessed cut-offs	Chi-squared	P-value (CS)
All Cognitively unimpaired participants (n=154)	CL 24	92.273	<0.0001
	CL 20	81.710	<0.0001
	CL 18	8.731	<0.0001
	CL 16	85.141	<0.0001
	CL 12	58.231	<0.0001
Cognitively unimpaired participants who underwent [18F]Florbetaben PET (n=74)	CL 24	42.125	<0.0001
	CL 20	41.645	<0.0001
	CL 18	30.263	<0.0001
	CL 16	27.554	<0.0001
	CL 12	13.774	<0.0001
Cognitively unimpaired participants who underwent [11C]Pittsburgh Compound B PET (n=80)	CL 24	42.813	<0.0001
	CL 20	38.756	<0.0001
	CL 18	51.676	<0.0001
	CL 16	51.676	<0.0001
	CL 12	47.927	<0.0001

Supplemental Table 4: Descriptive information for all 17 discordant cases, including the participant's age, APOE genotype, cognitive status, the amyloid tracer used for PET imaging, the Centiloid quantification value, and visual interpretations.

	Amyloid tracer	Age (yr)	Cognitive status	APOE genotype	Centiloid value of amyloid deposition	Visual read of amyloid PET	Visual read comment
Case 1	[18F]Florbetaben	69	Cognitively unimpaired	APOE ε3/4	18.85	Negative	
Case 2	[18F]Florbetaben	72	Cognitively unimpaired	APOE ε3/3	18.67	Negative	
Case 3	[18F]Florbetaben	53	Cognitively unimpaired	APOE ε3/3	18.26	Negative	
Case 4	[18F]Florbetaben	70	Cognitively unimpaired	APOE ε3/3	35.44	Negative	
Case 5	[18F]Florbetaben	72	Cognitively unimpaired	APOE ε2/3	21.75	Negative	
Case 6	[18F]Florbetaben	70	Cognitively unimpaired	APOE ε3/3	24.54	Negative	
Case 7	[18F]Florbetaben	68	Mild cognitive impairment	APOE ε3/4	10.96	Positive	Focally positive
Case 8	[11C]Pittsburgh Compound B	74	Cognitively unimpaired	APOE ε3/4	19.45	Negative	
Case 9	[11C]Pittsburgh Compound B	76	Cognitively unimpaired	APOE ε3/3	42.21	Negative	Atypical signal in basal ganglia
Case 10	[11C]Pittsburgh Compound B	56	Cognitively unimpaired	APOE ε3/4	-5.87	Positive	Focally positive
Case 11	[11C]Pittsburgh Compound B	58	Cognitively unimpaired	APOE ε3/4	-5.10	Positive	Focally positive
Case 12	[11C]Pittsburgh Compound B	76	Mild cognitive impairment	APOE ε3/3	-1.78	Positive	Focally positive
Case 13	[11C]Pittsburgh Compound B	78	Mild cognitive impairment	APOE ε3/3	-1.24	Positive	Focally positive
Case 14	[11C]Pittsburgh Compound B	68	Cognitively unimpaired	APOE ε3/4	16.70	Positive	Diffusely positive
Case 15	[11C]Pittsburgh Compound B	73	Alzheimer's dementia	APOE ε3/4	11.25	Positive	Focally positive
Case 16	[11C]Pittsburgh Compound B	85	Cognitively unimpaired	APOE ε3/3	-4.16	Positive	
Case 17	[11C]Pittsburgh Compound B	57	Cognitively unimpaired	APOE ε3/3	-5.04	Positive	Mildly blurry image

Supplemental Table 5: Characteristics of participants who underwent more than one amyloid PET scan. Centiloid quantification units and visual reads for baseline and follow-up PET scans are shown in each row.

Participant	Age at baseline (yr)	Sex	Clinical Dementia Rating (CDR)	APOE genotype	Centiloid value	Visual read
Case 1;	58	Male		APOE ε3/4		
Baseline			0		-5.87	Positive
Follow-up			0		-5.10	Positive
Case 2;	70	Male		APOE ε3/3		
Baseline			0		-10.61	Negative
Follow-up			0		-7.52	Negative
Case 3;	41	Female		APOE ε2/3		
Baseline			0		-0.87	Negative
Follow-up			0		0.076	Negative
Case 4;	71	Female		Not available		
Baseline			0		2.48	Negative
Follow-up			0		3.01	Negative
Case 5;	68	Female		APOE ε3/4		
Baseline			0		-6.39	Negative
Follow-up			0		-4.60	Negative
Case 6;	79	Male		APOE ε3/4		
Baseline			0		49.50	Positive
Follow-up			0		60.18	Positive
Case 7;	89	Male		APOE ε3/3		
Baseline			0		-4.05	Negative
Follow-up			0		1.88	Negative
Case 8;	82	Female		APOE ε3/3		
Baseline			0		-4.47	Negative
Follow-up			0		-0.24	Negative
Case 9;	78	Male		APOE ε3/3		
Baseline			0.5		-1.78	Positive
Follow-up			0.5		-1.24	Positive
Case 10;	76	Male		APOE ε3/3		
Baseline			0.5		4.38	Negative
Follow-up			0.5		5.85	Negative
Case 11;	89	Female		APOE ε3/4		
Baseline			0		9.17	Negative
Follow-up			0		17.87	Negative
Case 12;	69	Female		APOE ε2/3		
Baseline			0		-3.69	Negative
Follow-up			0		-3.28	Negative
Case 13;	68	Female		APOE ε3/4		
Baseline			0		18.81	Positive
Follow-up			0		24.23	Positive
Case 14;	63	Male		APOE ε3/3		
Baseline			0		-9.55	Negative
Follow-up			0		-9.41	Negative
Case 15;	71	Female		APOE ε3/4		
Baseline			0		23.79	Positive
Follow-up			0		33.59	Positive

Case	Age	Sex		APOE		
Case 16; Baseline Follow-up	68	Female	0 0	APOE ε3/4	12.66 16.70	Negative Positive
Case 17; Baseline Follow-up	71	Female	0 0	APOE ε3/4	-8.40 -7.76	Negative Negative
Case 18; Baseline Follow-up	69	Female	0 0	Not available	-8.51 -6.98	Negative Negative
Case 19; Baseline Follow-up	67	Female	0 0	Not available	3.93 12.73	Negative Negative
Case 20; Baseline Follow-up	74	Male	0 0	Not available	68.74 71.77	Positive Positive
Case 21; Baseline Follow-up 1 Follow-up 2	82	Male	0 0 0	APOE ε3/4	38.52 65.44 67.24	Positive Positive Positive
Case 22; Baseline Follow-up	73	Female	0 0	APOE ε3/3	-1.28 -0.55	Negative Negative
Case 23; Baseline Follow-up	86	Male	0 0	APOE ε3/3	-11.21 -8.97	Negative Negative
Case 24; Baseline Follow-up	87	Male	0.5 0.5	APOE ε3/3	90.57 102.85	Positive Positive
Case 25; Baseline Follow-up	82	Female	0 0	APOE ε3/3	-9.86 -9.17	Negative Negative

SUPPLEMENTAL REFERENCES

1.   Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ open* 2016;6:e012799