



Validation of quantitative assessment of florbetaben PET scans as an adjunct to the visual assessment across 15 software methods

Aleksandar Jovalekic¹ · Núria Roé-Vellvé¹ · Norman Koglin¹ · Mariana Lagos Quintana¹ · Aaron Nelson² · Markus Diemling³ · Johan Lilja³ · Juan Pablo Gómez-González⁴ · Vincent Doré⁵ · Pierrick Bourgeat⁶ · Alex Whittington⁷ · Roger Gunn⁷ · Andrew W. Stephens¹ · Santiago Bullich¹

Received: 27 January 2023 / Accepted: 19 May 2023 / Published online: 10 June 2023
© The Author(s) 2023

Abstract

Purpose Amyloid positron emission tomography (PET) with [¹⁸F]florbetaben (FBB) is an established tool for detecting A β deposition in the brain in vivo based on visual assessment of PET scans. Quantitative measures are commonly used in the research context and allow continuous measurement of amyloid burden. The aim of this study was to demonstrate the robustness of FBB PET quantification.

Methods This is a retrospective analysis of FBB PET images from 589 subjects. PET scans were quantified with 15 analytical methods using nine software packages (MIMneuro, Hermes BRASS, Neurocloud, Neurology Toolkit, statistical parametric mapping (SPM8), PMOD Neuro, CapAIBL, non-negative matrix factorization (NMF), Amyloid^{IQ}) that used several metrics to estimate A β load (SUVR, centiloid, amyloid load, and amyloid index). Six analytical methods reported centiloid (MIMneuro, standard centiloid, Neurology Toolkit, SPM8 (PET only), CapAIBL, NMF). All results were quality controlled.

Results The mean sensitivity, specificity, and accuracy were $96.1 \pm 1.6\%$, $96.9 \pm 1.0\%$, and $96.4 \pm 1.1\%$, respectively, for all quantitative methods tested when compared to histopathology, where available. The mean percentage of agreement between binary quantitative assessment across all 15 methods and visual majority assessment was $92.4 \pm 1.5\%$. Assessments of reliability, correlation analyses, and comparisons across software packages showed excellent performance and consistent results between analytical methods.

Conclusion This study demonstrated that quantitative methods using both CE marked software and other widely available processing tools provided comparable results to visual assessments of FBB PET scans. Software quantification methods, such as centiloid analysis, can complement visual assessment of FBB PET images and could be used in the future for identification of early amyloid deposition, monitoring disease progression and treatment effectiveness.

Keywords Alzheimer's disease · Amyloid-beta · Florbetaben · Quantification · Centiloid · Mild cognitive impairment

Introduction

Alzheimer's disease (AD) is the leading cause of dementia and constitutes 60–80% of dementia cases over the age of 65 years. There is a protracted preclinical (asymptomatic)

period during which abnormal traits manifest in the central nervous system detectable via biomarkers of the disease. It is characterized by the persistent formation of amyloid plaques and the subsequent development of amyloid-dependent tau pathology, neuroinflammation, and synaptic dysfunction, followed by clinical symptoms [1]. As such, the protracted AD continuum is accompanied by increasing severity of symptoms and is ultimately fatal. Detecting and quantifying such biomarkers are critical for early diagnosis of AD, for the stratification of patients for targeted disease-modifying drugs, and to monitor the effects of potential treatments by providing information about the level of relevant neuropathology [2–5].

Biomarkers and methods for AD pathology detection have been established in recent years, which induced a shift towards a biomarker-assisted diagnosis. Guidelines emphasize the fundamental role of amyloid in the AD diagnostic

✉ Aleksandar Jovalekic
a.jovalekic@life-mi.com

¹ Life Molecular Imaging GmbH, Berlin, Germany

² MIM Software Inc., Cleveland, OH, USA

³ Hermes Medical Solutions, Stockholm, Sweden

⁴ QubioTech Health Intelligence, A Coruña, Spain

⁵ Department of Molecular Imaging & Therapy, Austin Health, Melbourne, Australia

⁶ CSIRO, Brisbane, Australia

⁷ Invicro, London, UK

process. Amyloid positron emission tomography (PET) with florbetaben (FBB) is an established tool for detecting A β deposition in vivo. FBB underwent a global multicenter development program and was approved by the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) in 2014, and other national agencies have subsequently approved its use [6, 7]. FBB has been robustly validated against histopathological confirmation of neuritic A β plaque density in the brain as the standard of truth (SoT) [8]. Visual inspection of the FBB PET scans is the approved and validated method for image interpretation and has proven to be very reliable in clinical practice [9].

Although visual assessment (VA) is an appropriate method for the vast majority of scans assessed in clinical practice, it has been proposed that certain situations would benefit from additional quantitative information gleaned from the amyloid images [10, 11]. In a heterogeneous clinical population, for example, VA can be challenging, especially for less-experienced readers: anatomical abnormalities, such as cortical thinning caused by atrophy and ventricular enlargement, may hamper VA; the presence of other neurodegenerative disorders can also confound the assessment on a purely visual basis; and the dichotomous readout of VA may lack the required sensitivity to assess longitudinal changes of amyloid load [12, 13]. A substantial part of the clinical population now includes patients with early-stage disease, which may only show isolated regional uptake and emerging amyloid deposits [14, 15]. Additionally, a small fraction of PET images is only assessed with low confidence, e.g., when amyloid levels are intermediate [16, 17]. In such situations, the continuous measurement of amyloid burden with quantification can provide additional information to the binary VA and increase confidence in such equivocal situations.

Indeed, quantification of PET images is commonly performed in research studies [18] and therapeutic clinical trials [3, 19–24]. The recent FDA accelerated approval of lecanemab, an anti-amyloid antibody, was based on the observed reduction of A β plaques as quantified and monitored by amyloid PET [25]. Detection of longitudinal changes using quantification methods is established in the research and development setting [14] with many PET software packages capable of calculating amyloid burden both on a composite and a regional level [26, 27].

In this context, this retrospective data analysis was conducted to evaluate FBB PET quantification as an adjunct to VA. This study was aimed at providing scientific evidence of the robustness and additional value of FBB PET quantification. FBB PET scans from previous clinical trials were quantitatively assessed with several analytical methods. The diagnostic performance (i.e., sensitivity and specificity) of quantification against the histopathological confirmation of A β load was estimated and compared to the effectiveness of the approved VA method. Additionally, the concordance

between visual and quantitative evaluation of FBB PET scans was assessed. The reliability and comparability of the different analytical methods were further tested.

Methods

Study design

This is a retrospective analysis of FBB PET images that had been obtained in previous clinical trials. All data had been acquired in accordance with the Declaration of Helsinki and after approval of local ethics committees; the informed consent from subjects included the future scientific analyses of acquired PET scans.

Participants

The study population consisted of 589 subjects with at least one available FBB PET scan from previously completed clinical studies. All subjects were grouped into 4 cohorts to address different study objectives (Table 1).

Cohort #1 included end-of-life subjects with histopathological confirmation of presence or absence of A β neuritic plaques ($n=81$) and young healthy controls ($n=10$). A β status in the young healthy controls (27.4 ± 5.5 years (mean \pm SD)) was considered absent in all brain regions.

Cohort #2 comprised 439 subjects, including 53 subjects from cohort #1 and 45 subjects from cohort #3 that were visually assessed by 5 independent newly trained blinded readers. Four different subsets were analyzed. Subset 1 excluded subjects used to generate the cut-off and comprised 386 subjects. Subset 2 represents the full sample, consisting of all 439 subjects. Subset 3 was based on subset 2 but included only subjects where the 5 independent blinded readers had consensus in the assessment, resulting in a subset of $n=336$ subjects. Subset 4 comprised randomly selected scans to assess intra-software reliability.

Cohort #3 included amnesic mild cognitively impaired (MCI) subjects ($n=45$) that underwent three FBB PET scans at baseline and a 4-year clinical follow-up.

Cohort #4 was the publicly available GAAIN (Global Alzheimer's Association Information Network) dataset (<http://www.gaain.org/centiloid-project>) ($n=35$) including FBB PET scans of patients with MCI, AD, and frontotemporal dementia (FTD) and elderly and young healthy controls.

Image acquisition and reconstruction

The PET image acquisition and reconstruction were performed as detailed in previous literature [8, 9]. Briefly, all subjects underwent a 20-min PET scan (4 \times 5-min dynamic

Table 1 Description of the cohorts included in this retrospective analysis

Cohort	Clinical data source	Clinical diagnosis	Objective for retrospective analysis	Subset analysis	Visual assessment
#1, $n=91$	NCT01020838 [8, 9]	59 AD 4 DLB 9 DEM 19 NDV	To assess the sensitivity and specificity of quantification methods against the histopathological confirmation To compare the diagnostic performance of quantification methods with previously established visual assessment performed by 3 expert blinded readers and 5 newly trained blinded readers		3 expert blinded readers 5 newly trained blinded readers
#2, $n=439$	Images from clinical phase 1, 2, and 3 studies were visually assessed by five independent blinded readers in a pooled read study	183 eHC 169 AD 9 DLB 11 FTLD 1 FTD 51 MCI 5 PD 4 VaD 6 other	To assess the concordance between visual assessment performed by 5 blinded readers and binary quantitative assessment of FBB PET scans To assess agreement and reliability between (inter-software) and within (intra-software) analytical methods	Subset 1 ($n=386$): excluded subjects from cohort #1 used to generate the cut-off Subset 2 ($n=439$): represents the full sample Subset 3 ($n=336$): based on subset 2 and included only subjects with consensus among 5 independent blinded readers Subset 4 ($n=84$): randomly selected scans to assess intra-software reliability	5 newly trained blinded readers
#3, $n=45$	NCT01138111 [28]	45 MCI	To assess the prognostic accuracy of quantitative assessment of amyloid-beta load using FBB PET for progression to AD To assess the capacity of quantitative assessment of amyloid-beta load using FBB PET to detect longitudinal changes		
#4, $n=35$	Global Alzheimer's Association Information Network (GAAIN) [29]	10 yHC 6 eHC 9 MCI 8 AD 2 FTD	To compare the results of FBB PET quantification against the quantitative results provided in a publicly available dataset		

AD Alzheimer's disease, DEM other dementia, DLB dementia with Lewy bodies, eHC elderly healthy control, FTD frontotemporal dementia, FTLD frontotemporal lobar degeneration, MCI amnesic mild cognitively impaired, NDV non-demented volunteer, PD Parkinson's disease, VaD vascular dementia, yHC young healthy control

frames) starting at 90 min after intravenous injection of 300 MBq \pm 20% of FBB. PET scans were then reconstructed using the ordered subset expectation maximization (OSEM) algorithm with 4 iterations and 16 subsets (zoom = 2) or comparable reconstruction. Corrections were applied for attenuation, scatter, randoms, and dead time. 3D volumetric T1-weighted brain MRI data was available for 497 subjects.

Post-mortem histopathology

Histopathological confirmation of A β presence or absence in the brain was available for subjects from cohort #1. A β was regarded as present in a given brain region when moderate or frequent plaques were present either by Bielschowsky silver staining (BSS) or immunohistochemistry. Detailed autopsy procedures and collection/analysis of these post-mortem histopathology specimens are described in detail in [8, 9].

Visual assessment

FBB PET scans from cohort #1 were visually assessed by 3 independent and blinded expert readers and 5 newly trained, independent blinded readers. PET scans from cohort #2 were visually assessed by independent blinded readers, which were not identical to those that assessed cohort #1. All blinded readers followed the same reading methodology as previously described by Seibyl et al. [9]. No structural information from CT or MRI was available to interpreters.

Quantification of amyloid load in PET images

All FBB PET scans were quantified with nine software packages using several metrics to estimate A β levels (Table 2). For some software packages, different analytical methods were tested using different reference regions or without using the T1-weighted MRI scan. All the scans were quantified in batch mode to minimize operator intervention. The operators were different for each software package and blinded to the diagnosis of subjects, demographics, visual PET assessment, histopathology results, and all other clinical data. All the results were quality controlled using the guidelines described in the supplementary material. In scans where quality control was not successful, the operator was allowed to use the available tools of the software to overcome quality control issues and to improve quantification. Figure 1 illustrates quality control issues that may impact the correct calculation of A β quantity from FBB PET scans. Subjects for whom quality control (QC) issues could not be resolved were excluded from the analysis of the individual analytical method (details and results of the QC procedure are described in the supplement). The quantification methods were assessed based on their continuous output, and binary classification of the continuous results was only used when comparing quantification to visual assessment.

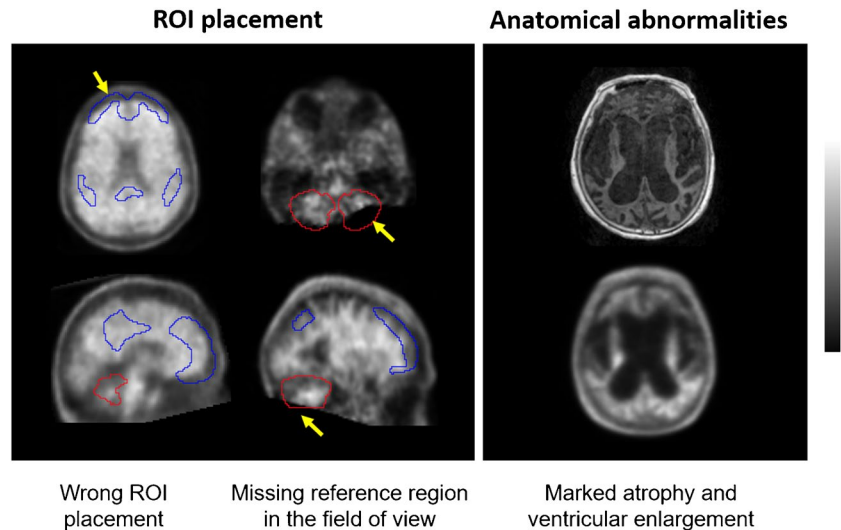
The non-commercial SPM-based methods using the T1-weighted MRI (analytical methods #7, #8, and #9) used spatial processing (PET-MRI registration and normalization to the standard template) according to the procedure specified by Klunk et al. [31]. ROIs were defined as the

Table 2 Summary of analytical methods included in the retrospective analysis

Analytical method id	Software	MRI required	Reference region	Metric	CE-marked	510(k)	References
#1	MIMneuro (v.7.1.2)	No	WC	CL	Yes	Yes	[30]
#2	Hermes BRASS (v.5.1.1)	No	WC	SUVR	Yes	Yes	
#3	Neurocloud (v.1.4)	No	WC	SUVR	Yes	No	
#4	Standard centiloid	Yes	WC	CL	No	No	[29, 31]
#5	Neurology Toolkit	No	WC	CL	No	No	[32]
#6	Neurology Toolkit	No	NA	Amyloid index	No	No	[32]
#7	SPM8	Yes	WC	SUVR	No	No	
#8	SPM8	Yes	WC*	SUVR	No	No	
#9	SPM8	Yes	CGM	SUVR	No	No	
#10	SPM8	No	WC	CL	No	No	
#11	PMOD Neuro (v.3.7)	Yes	CGM	SUVR	No	No	
#12	CapAIBL	No	WC	CL	No	No	[33–35]
#13	NMF CapAIBL	No	WC	CL	No	No	[36]
#14	Amyloid ^{IQ}	Yes	NA	Amyloid load	No	Submitted ¹	[37, 38]
#15	Amyloid ^{IQ}	No	NA	Amyloid load	No	Submitted ¹	[37, 38]

CGM cerebellar gray matter, CL centiloid, NA not applicable, NMF non-negative matrix factorisation, SPM statistical parametric mapping, SUVR standardized uptake value ratio, WC whole cerebellum (volume weighted average of CGM and WC), WC* whole cerebellum (average of CGM and cerebellar WM); ¹personal communication

Fig. 1 Examples of quality control issues that may impact the correct calculation of A β load from FBB PET scans. Left: yellow arrows indicate wrong cortical (blue) and reference region (red) ROI placement. Right: MRI (top) and FBB PET (bottom) scans illustrating marked atrophy and ventricular enlargement



intersection between the standard automated anatomic labeling (AAL) atlas and the normalized gray matter segmentation map thresholded at a probability level of 0.2 [14]. The PET-only SPM-based method (analytical method #10) used the SPM8 tools to normalize each FBB PET scan on the standard space with an adaptive template generated from the average of several AD patients ($n=53$) and amyloid-beta negative healthy controls ($n=55$). The standard CL ROIs were applied to the normalized scan.

Statistical methods

Sensitivity and specificity The diagnostic performance of FBB PET scan analysis of visual and quantitative assessment was evaluated by means of sensitivity, specificity, accuracy, and their 95% confidence interval using histopathological confirmation as SoT (cohort #1). The cut-off for the established A β pathology for each analytical method was derived using receiver operating characteristic (ROC) curve analysis to ascertain the optimal threshold for the sensitivity and specificity calculation of the sample with histopathology confirmation of A β . The quantitative value that provided the highest Youden index ($J = \text{sensitivity} + \text{specificity} - 1$) was selected as the cut-off.

Concordance between visual and quantitative assessments The percent agreement between visual read and binary quantitative assessments and the exact two-sided 95% Clopper-Pearson confidence interval in subjects from cohort #2 in subsets 1 and 2 were calculated for each reader assessment and the majority read (i.e., agreement of the majority of five readers). The concordance was also assessed in the subset of scans for which all the five independent blinded readers provided the same classification (“consensus” subset, subset 3).

Inter-software reliability The reliability of agreement across software packages (i.e., inter-rater reliability) for the binary assessment (normal/abnormal A β load) in cohort #2 in subset 2 was performed by means of the Fleiss’ kappa (κ) and correlation analysis calculated across all pairs of analytical methods.

Intra-software reliability A randomly selected subset of scans ($n=84$) of cohort #2 (subset 4) was analyzed twice. Reliability was assessed by means of scatter plots and determination coefficients (R^2).

Comparison of quantitative analytical methods using a public dataset A linear regression was fitted to the A β metrics for each analytical method and the published centiloid (CL_{GAAIN}) results [29] from the publicly available GAAIN dataset (cohort #4) as follows (A β metrics = $\alpha CL_{GAAIN} + \beta$, where α and β are the constants of the model).

Distribution of A β load in elderly cognitively normal subjects A β metrics were assessed in elderly cognitively normal subjects from cohort #2 (subset 3) that were assessed as A β -negative by consensus by five independent blinded readers. Histograms of the different A β metrics were generated and the 95th percentile of the A β metrics was calculated for each analytical method.

A β deposition over time A linear regression model was fitted to each subject’s data in cohort #3, A β metric = $\alpha \cdot t + \beta$, where A β metric is the metric used to assess A β load, α and β are the coefficients of the model, and t is the scan time in years. The annual A β load increase was obtained from α . Subsequently, the average annual A β metric increase (α) was tested statistically by means of a Wilcoxon rank sum test whether A β -positive subjects, as defined by pipeline

specific cut-offs derived from the pathology sample, are in the process of accumulating A β (i.e., ($H_0: \alpha=0; H_1: \alpha>0$)). Descriptive statistics are reported as arithmetic mean and sample standard deviation (mean \pm SD).

Prognostic accuracy of quantitative assessment Clinical follow-up was evaluated in MCI subjects from cohort #3 to assess conversion to AD. The proportion of subjects that progressed and not progressed to AD after 4-year clinical follow-up and 95% exact Clopper-Pearson confidence intervals was calculated in each group (i.e., A β -negative or A β -positive at baseline). Positive likelihood ratio (LR+) and negative likelihood ratio (LR-) were also calculated. Descriptive statistics are reported as arithmetic mean and sample standard deviation (mean \pm SD).

Results

More than 11,000 PET scan quantifications from 589 subjects were analyzed in this retrospective analysis. All scans underwent quality control to ensure correct positioning of regions of interest (see supplementary material); 97.8% (range 94.95–100%) of the cases of the full sample met the stringent quality control criteria and were included in the presented analyses.

Sensitivity and specificity (cohort #1)

Optimal quantitative cut-offs for each analytical method were developed using ROC curve analyses and histopathological confirmation as SoT (cohort #1). Centiloid values above 35 CL (average cut-off across analytical methods reporting centiloids of 36.2 ± 5.5 CL) indicate established A β pathology corresponding to moderate and frequent neuritic plaques by neuropathology. In order to make a fair comparison, a cut-off based on pathology was utilized in this study, aligning with the established method of VA. It is important to note that the threshold for identifying early amyloid pathology is continually evolving and is often determined using a group of young healthy controls [14].

The mean sensitivity, specificity, and accuracy were $96.1 \pm 1.6\%$, $96.9 \pm 1.0\%$, and $96.4 \pm 1.1\%$, respectively, across all quantitative methods (cohort #1). For those analytical methods reporting centiloids, the results were $96.1 \pm 1.6\%$, $97.4 \pm 1.2\%$, and $96.7 \pm 1.2\%$, respectively. The positive predictive value (PPV) and the negative predictive value (NPV) in this study population across all quantitative methods were $98.1 \pm 0.5\%$ and $93.7 \pm 2.3\%$, respectively. Analyses of additional SoTs and examining a subgroup of CE-marked methods produced comparable

results (see supplemental Table 1). The mean sensitivity, specificity, and accuracy of VA were 97.0 ± 1.7 , 92.6 ± 1.3 , and 95.2 ± 0.5 , respectively, when performed by expert readers and 94.5 ± 3.3 , 75.0 ± 18.2 , and 88.4 ± 5.4 respectively, when performed by newly trained readers. The PPV and NPV values for expert readers in this study population were $95.3 \pm 0.7\%$ and $95.3 \pm 2.4\%$, respectively, while newly trained readers had values of $87.4 \pm 8.2\%$ and $93.5 \pm 4.7\%$ for PPV and NPV, respectively. In this challenging end-of-life population, the overall accuracy of VA performed by expert readers was similar to quantitative assessment, but significantly lower for visual assessment performed by newly trained readers (Mann–Whitney test; $p=0.001$), as it was also reported earlier from this dataset [9].

Concordance between visual and quantitative assessment (cohort #2)

The mean percentage of agreement between binary quantitative assessment and visual majority assessment on the cohort #2 dataset was $92.4 \pm 1.5\%$ (range 88.9–94.9%) (subset 1: excluding subjects with autopsy that were used to generate the cut-offs) or $92.5 \pm 1.5\%$ (range 89.2–94.8%) (subset 2: whole cohort). For the consensus dataset (subset 3: reads had consensus VA), the mean percentage agreement was $97.4 \pm 1.3\%$ (range 93.8–99.1%). Analysis of a subgroup of CE-marked methods produced comparable results (see supplementary material).

Binarization of quantitative assessments was based on individual abnormality cut-offs derived from the pathology sample. Concordance between quantitative assessment and VA was highest for cases with either absence or substantial presence of amyloid. The agreement rate between methods dropped for cases with intermediate amyloid accumulation or close to pathology-derived cut-offs (see Fig. 2).

Inter- and intra-software reliability (cohort #2)

Substantial agreement was observed across software packages. Fleiss' κ for the inter-software reliability was 0.90 for the full cohort (cohort #2, subset 2), and 0.94 for the consensus cohort dataset (cohort #2, subset 3). The Fleiss κ of the VA across 5 independent blinded readers using the same subsets was 0.79 for the full cohort (cohort #2, subset 2). The inter-software agreement was also high for all possible software pairs ranging from 0.78 to 0.99 (cohort #2, subset 2). All pairs of analytical methods were linearly associated with an average correlation coefficient of 0.95 ± 0.03 (mean \pm SD) ranging between 0.875 and 0.997 (cohort #2, subset 2).

The R^2 value for all analytical methods' re-analysis (intra-software reliability) ranged between 0.98 and 1.00 (cohort #2, subset 4).

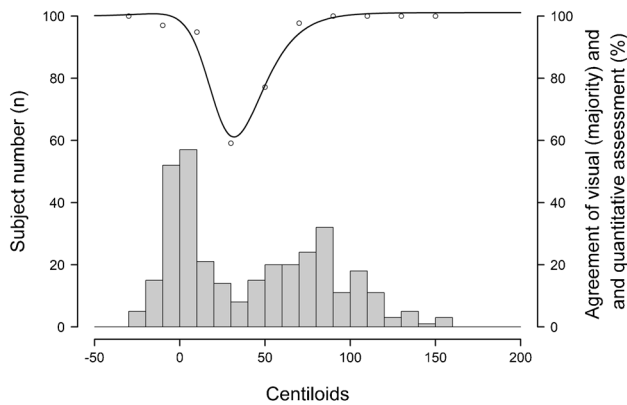


Fig. 2 Agreement rate (circles) between visual and quantitative assessments (standard centiloid method). Histogram shows the subject numbers as a function of centiloid levels obtained from the standard centiloid method. The black line is the interpolation of the agreement rate. Concordance of the standard centiloid method with VA is 92.5% and is highest for cases with either absence or substantial presence of amyloid. Discordance between visual reads and quantification is highest at intermediate amyloid accumulation stages or close to pathology-derived cut-offs

Comparison of quantitative analytical methods using a public dataset (cohort #4)

All the software methods that reported centiloids passed the validation criteria described by Klunk et al. (i.e., slope will be between 0.98 and 1.02, the intercept between -2 and 2 CL, and the $R^2 > 0.98$.) [31]. Quantification results of all analytical methods were also tightly correlated with the centiloid values published in the literature [29], with determination coefficients R^2 ranging from 0.92 to 0.99 for pipelines that do not report CLs. The R^2 of the standard centiloid pipeline was 1.00, as expected. Other pipelines reporting centiloids that do not follow exactly the procedure described in Klunk et al. [31] had R^2 values ranging from 0.98 to 0.99.

Distribution of quantitative metrics in elderly healthy control subjects (cohort #2)

The distribution of A β metrics was assessed in a subset of cognitively normal controls ($n = 122$ for PET-only analytical methods; $n = 88$ for analytical methods using T1-weighted MRI) from cohort #2, subset 3, who were classified by consensus as A β -negative by the five independent blinded readers. Table 3 lists the median and the 95th percentile of A β metrics calculated for each analysis method. The average 95th percentile for the six analytical methods that provided centiloid estimates was 21.2 ± 2.9 CL (mean \pm SD, $n = 6$). Figure 3 shows the corresponding distribution histograms for centiloid methods.

Table 3 Median and 95th percentile of A β metrics calculated in elderly healthy controls ($n = 122$ for PET-only analytical methods; $n = 88$ for analytical methods using T1-weighted MRI) that were visually assessed as A β -negative by five independent blinded readers

Software	Metric	Median	95th percentile
#1 MIMneuro	Centiloid	5.01	25.09
#2 Hermes BRASS	SUVr	1.17	1.29
#3 Neurocloud	SUVr	1.08	1.28
#4 Standard centiloid	Centiloid	2.07	20.94
#5 Neurology Toolkit	Centiloid	5.56	22.97
#6 Neurology Toolkit	Amyloid index	-0.44	-0.30
#7 SPM8 (WC)	SUVr	1.09	1.21
#8 SPM8 (WC*)	SUVr	0.88	0.98
#9 SPM8 (CGM)	SUVr	1.18	1.33
#10 SPM8-PET only (WC)	Centiloid	4.57	22.52
#11 PMOD Neuro	SUVr	1.18	1.32
#12 CapAIBL	Centiloid	1.81	18.32
#13 NMF CapAIBL	Centiloid	2.95	17.59
#14 Amyloid ^{IQ} (PET only)	Amyloid load	18.30	28.14
#15 Amyloid ^{IQ} (MR)	Amyloid load	16.30	24.36

CGM cerebellar gray matter, NMF non-negative matrix factorisation, SPM statistical parametric mapping, SUVr standardized uptake value ratio, WC whole cerebellum (volume weighted average of CGM and WC), WC* whole cerebellum (average of CGM and cerebellar WM)

Impact of amyloid status on (non)progression from MCI to AD dementia (cohort #3)

Patients in cohort #3 underwent repeated clinical evaluations over a period of up to 55 months after an initial baseline assessment [28]. On average across all quantification methods, $86.7 \pm 4.1\%$ of the A β -positive subjects by quantitative assessment at baseline progressed to AD dementia as revealed at the subsequent clinical follow-up. $1.7 \pm 3.1\%$ of A β -negative scans at baseline did progress to AD dementia. This relates to a positive likelihood ratio (LR+) of 8.2 ± 2.6 and a negative likelihood ratio (LR-) 0.02 ± 0.04 .

The AD progression numbers based on VA at baseline were very similar. On average, $80.9 \pm 5.8\%$ of the A β -positive patients progressed to AD dementia within 55 months clinical follow-up. $1.9 \pm 2.6\%$ of subjects with a negative scan at baseline did progress to AD dementia. This relates to a positive likelihood ratio (LR+) of 5.3 ± 1.8 and a negative likelihood ratio 0.02 ± 0.03 .

Positive A β scans in an MCI population were associated with significant increase in risk of clinical progression, as shown exemplarily for standard centiloid (Fig. 4).

Assessment of A β deposition over time (cohort #3)

Amyloid-beta accumulation was assessed in MCI subjects with longitudinal FBB PET scans (cohort #3). The annual

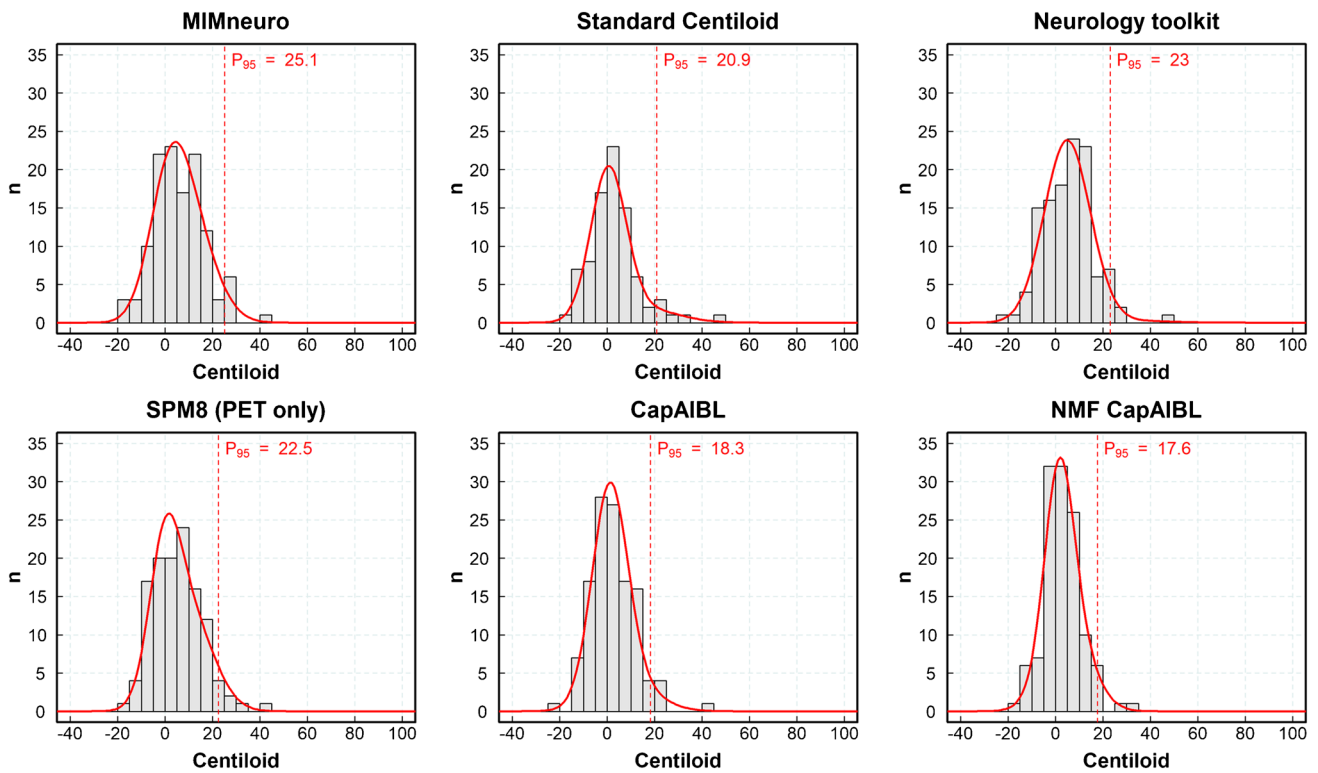


Fig. 3 Distribution of CL values in elderly cognitively normal subjects with amyloid-beta negative visual assessment for different analytical methods and 95th percentile. The figure includes, for illustra-

tive purposes, the Gaussian mixture model obtained from the sum of 2 Gaussian functions fitted to the data

A β increase was estimated with a linear regression model fitted to each subject’s data. For all 15 analytical pipelines, subjects with established A β deposition at baseline had annual rates of A β accumulation statistically different from zero ($p < 0.05$). The average rate of CL increase per year in centiloid pipelines ($n = 6$) was 3.49 ± 1.19 CL for A β -positive subjects at baseline and 0.68 ± 0.77 CL for A β -negative subjects at baseline.

Discussion

The results of this comprehensive analysis show that all investigated software quantification methods generate homogeneous and robust data and that quantification methods can complement visual assessment of FBB PET images. Adjunct use of quantification tools could be beneficial for newly trained or inexperienced operators, in instances when images are assessed with relatively low confidence based on visual assessment alone, to detect early amyloid deposition, or when amyloid levels of patients are close to “pathology” thresholds.

Application of A β PET quantification Excellent diagnostic efficacy of FBB PET quantification was demonstrated when

compared to histopathology assessment as SoT. Mean sensitivity and specificity for all the software packages were 96.1% and 96.9%, respectively. The VA results suggest that less-experienced readers will likely benefit (in terms of improved specificity) of adjunct information obtained

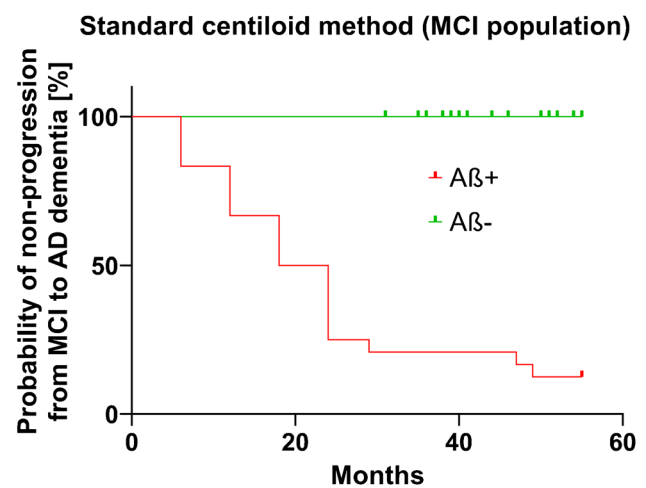


Fig. 4 Kaplan-Meier survival analysis based on amyloid status assessed by the standard centiloid quantification method in an MCI population (cohort #3). A β positivity is defined by the pathology-derived cut-off specific for the standard centiloid method

from quantitation, especially with difficult cases when the reader may have decreased confidence in the assessment. Similarly, the added value of adjunct use of quantification has been demonstrated for other amyloid tracers. Pontecorvo et al. suggested that access to quantitative information of the scans can improve the performance and confidence of some readers of florbetapir PET scans, particularly inexperienced readers [13]. Similarly, Kim et al. reported for florbetaben PET that when VA is supported by adjunct quantification, confidence in VA and inter-reader agreement is improved [39]. Finally, Bucci et al. reported that while quantitation of flutemetamol PET shows generally high agreement with VA, discordant cases with quantitative amyloid positivity are more likely to progress to AD, suggesting that quantification may be a tool to potentially detect earlier pathologic deposition [40].

The dichotomous quantitative assessment of FBB PET had a very high agreement of >92% to the majority VA. This result is consistent with previously published FBB data where percent agreement was reported between 88 and 97% [9, 10, 30, 41–47]. The average concordance of quantification and VA was higher (97.4%) in the consensus subset where images were only included when all five independent readers provided the same visual assessment. Very high agreement rates between quantification and VA have previously been shown for all amyloid PET tracers [5]. It is worth noting that typically quantitative methods tend to have higher concordance rates with expert readers than with non-expert readers [48].

While some details of individual quantification methods may differ, the software evaluation revealed homogenous performances in the current study. All software packages achieved excellent diagnostic efficacy when compared to histopathology and high concordance with visual assessments. Fleiss' kappa between analytical methods was almost perfect and superior to Fleiss' kappa of VA (0.9 and 0.79, respectively). This substantial inter-software agreement confirms good reproducibility of the results independent of the quantification method. The pairwise kappa agreement analysis led to similarly high agreement rates. Furthermore, all the quantitative metrics provided by the different analytical tools showed a tight linear agreement with an average correlation coefficient between pairs of analytical methods of 0.95 ± 0.03 . This substantial correlation between different methods substantiates the generalisability of the obtained results, and this was confirmed by an additional analysis with a publicly available dataset (GAAIN) in which quantification results from the evaluated methods correlated closely with the standard centiloid method published previously [29]. Intra-software reproducibility was excellent for all methods and correlation coefficients for the individual method re-analysis ranged

between 0.98 and 1.00. Overall, the relatively heterogeneous analysis approaches of the different methods yielded very homogenous and robust results.

While all approved [^{18}F]-labeled amyloid PET tracers have been validated based on their performance in late-stage disease patients against histopathology [8, 49, 50], the performance of quantitative methods in earlier stages of disease is important for the clinical routine situation. This was tested in cohort #3 (MCI), and the results seem particularly noteworthy, as they allow direct assessment of PET quantification in patients with early disease. Positive A β MCI patients in this cohort were associated with significant increase in risk of clinical progression, while negative A β patients, defined by the pathology-derived cut-off specific to each pipeline, did not progress to AD. The results of the different quantification methods and visual assessment were comparable demonstrating very high reliability and reproducibility and in concordance to the previously reported values for FBB [7, 28, 51, 52]. This confirms that quantification and visual assessment of FBB PET scans in early AD stages provide important information to predict progression to AD in MCI subjects over a 4-year observation period. This data is in line with results of the other amyloid tracers [53–56].

For all 15 analytical pipelines, subjects with established A β deposition at baseline had annual rates of A β accumulation statistically different from zero. Thus, quantification of amyloid load enables the evaluation of increased amyloid accumulation over time in a sensitive, objective manner; this can be particularly useful for low amyloid level subjects at an early disease stage and may help to identify groups of subjects most likely to benefit from early disease detection and possible therapeutic intervention. Indeed, selection by amyloid PET imaging has been a requirement to participate in therapeutic trials for many years [3, 4, 19–24].

Use of quantitative information to supplement visual assessment The correct interpretation of quantitative values is important for the appropriate use of quantitation to supplement VA. Interpretation of quantitative information of A β burden is conventionally performed by comparing the values against software-specific abnormality cut-offs or comparing against the typical distribution of negative and positive scans.

Those software packages that provided centiloid values reported similar cut-offs and interpretation of the results. Centiloid values above 35 CL indicate established A β pathology corresponding to moderate and frequent neuritic plaques by neuropathology. Subjects above this cut-off have a high probability of moderate or frequent plaques. Likewise, the agreement between VA and quantitative methods was extremely high for subjects with substantial amyloid levels. A lower cut-off of around 20 CL derived from elderly A β -negative, cognitively normal subjects provided a high

specificity to rule out amyloid pathology. As for subjects with substantial amyloid deposition, agreement between VA and quantitative methods was very high for low amyloid levels. The level of agreement between VA and quantitative methods decreased for cases with intermediate amyloid accumulation or those near the pathology-derived cut-offs. This observation is consistent with the findings of Zeltzer et al., who reported a higher frequency of visual-quantitative discrepancies in scans near the amyloid positivity threshold [48]. Furthermore, subjects with CL values indicating intermediate amyloid deposition had a higher rate of not reaching full concordance between all 5 readers (see supplemental Fig. 1). These findings indicate that subjects with emerging or intermediate amyloid burden may need more careful inspection, benefiting from additional quantification derived information. Figure 5 is an illustrative example of the use of quantitative information to supplement visual assessment of FBB PET scans using centiloids.

As illustrated in Fig. 5, focal uptake is one factor that should be considered when interpreting global quantitative measures. Multiple studies and cohort types have demonstrated the robustness of centiloid quantification as a reliable measure [5, 52]. In the clinical implementation context, it is crucial to emphasize that the interpretation of quantitation should always be done in conjunction with a visual read. This ensures data quality, allows for understanding of possible causes of discordance, and helps prevent false-negative

and false-positive classifications based solely on quantification or visual assessment. Figure 6 showcases examples of challenging cases characterized by emerging or focal amyloid deposition, which may result in discordance between visual readers and centiloid quantification. Most of the readers assessed these scans as positive, which were negative based on global binary quantitative assessments even though three of the cases present CL values pointing towards emerging A β pathology. In all these cases, regional information, e.g., as provided by z-score information, may provide additional incremental value and thus assist VA. As illustrated with these difficult and discordant patients in this work, focal uptake in regions less represented in the CL mask can result in lower quantitative values. In all these cases, regional information, e.g., as provided by z-score information, would provide an incremental value and thus assist correct classification of scans. This is in line with previous work [15, 26, 52, 57, 58].

The obtained centiloids of the current study agree with several recent reports across different tracers converging to the use of two cut-offs for amyloid PET abnormality: an early cut-off around CL = 11–20 where pathology may be emerging and a second around CL = 29–36 where amyloid burden levels correspond to moderate and frequent neuritic plaques. Bullich et al. developed a lower FBB PET cut-off of 13.5 CL when derived from young healthy controls indicating emerging A β pathology and derived a higher cut-off of 35.7 CL

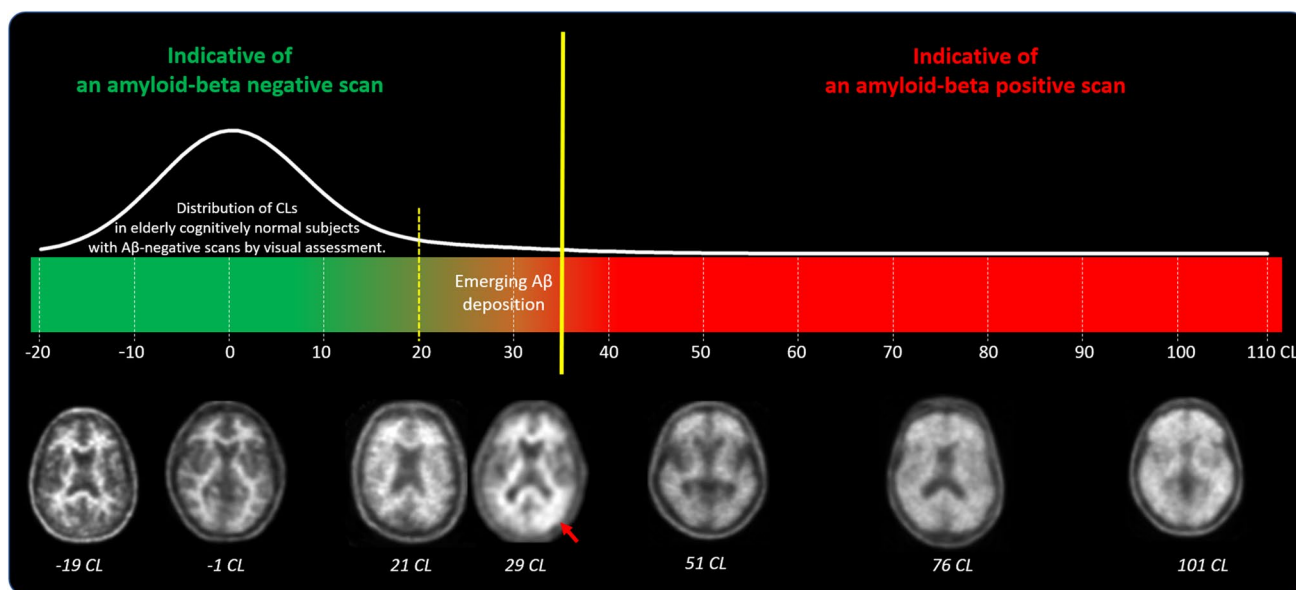


Fig. 5 Illustrative display of the use of quantitative information to supplement visual assessment of FBB PET scans. Centiloid values above 35 CL indicate established A β pathology corresponding to a density of moderate and frequent neuritic plaques by neuropathology. Centiloid values below 20 represent elderly cognitively normal subjects with negative amyloid-beta scans by visual assess-

ment. Centiloid values in the range between 20 and 35 CL are more likely to be ambiguous, can be either negative or positive by visual assessment, and correspond to subjects with emerging A β deposition. The readers should review such scans carefully to identify subtle amyloid accumulation that can be focal and/or unilateral (red arrow)

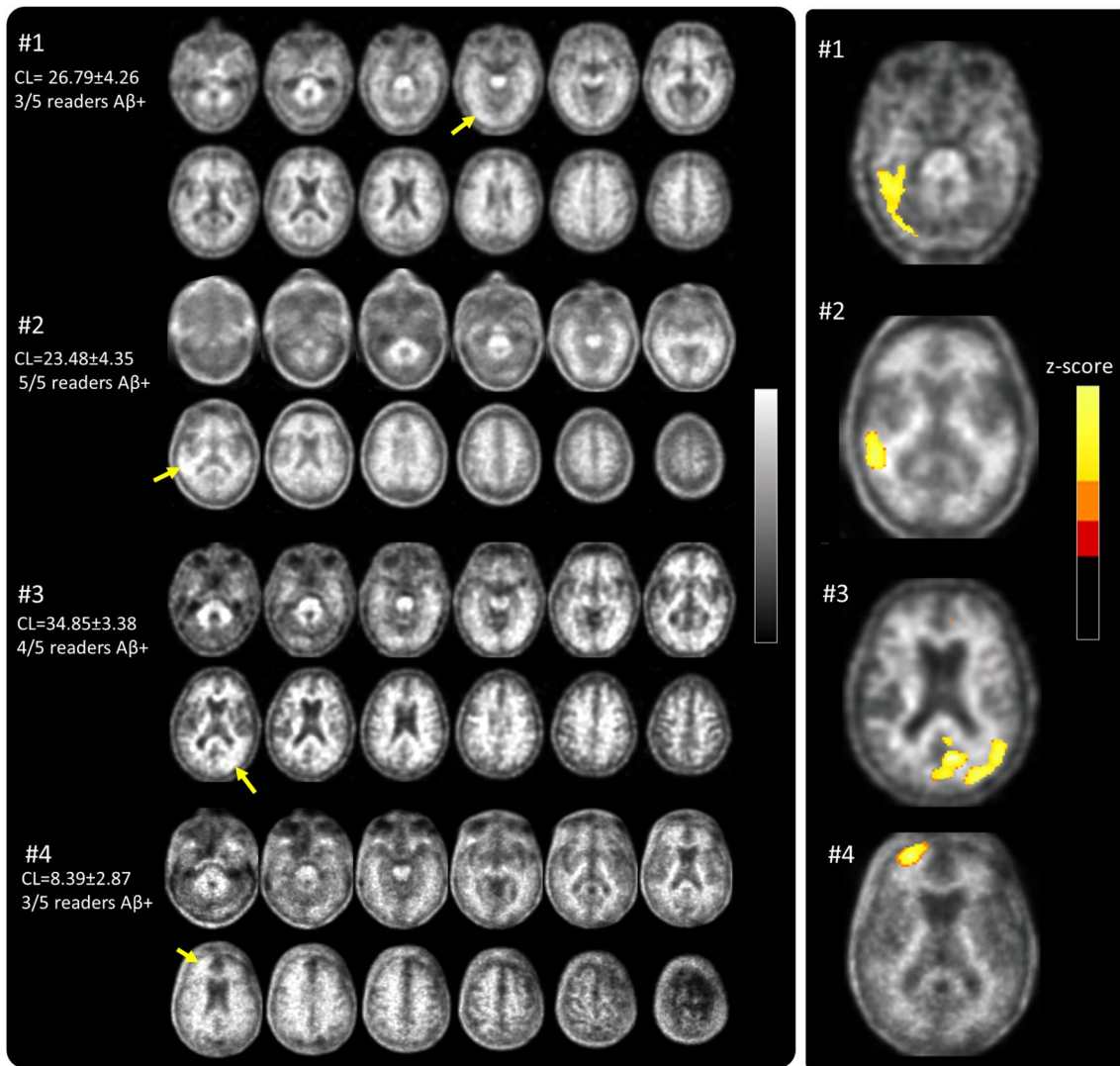


Fig. 6 Illustrative examples of 4 cases (#1 to #4) that showed focal tracer uptake (arrow) and relatively low CL values (mean \pm SD) obtained from the analytical pipelines analyzed. Also shown on the left is the number of independent blinded readers (out of 5 readers) that assessed the scans as positive ($A\beta+$). Focal tracer uptake with the cor-

responding z-scores is shown in the right panel. Only areas containing voxels with z-scores above 3 (in red, orange, and yellow color) are displayed. All 4 cases are clearly positive based on regional z-scores

where amyloid burden levels correspond to established neuropathology findings [14]. The publication showed that these two cut-offs define a subset of subjects (13.5–35.7 CL) characterized by pre-AD dementia levels of amyloid burden that precede other biomarkers, such as tau deposition or clinical symptoms and accelerated amyloid accumulation. Other early cut-offs of 11, 14, or 17 CL have been reported for the FACE-HBI, ALFA+, and AMYPAD PNHS (Prognostic and Natural History Study) studies using Gaussian mixture models [59]. Similarly, Salvadó et al. identified two cut-offs based on a direct comparison with established CSF $A\beta_{42}$ thresholds: CL = 12 to rule out amyloid pathology and CL = 29 to denote established pathology [60]. Mormino et al. also showed the

biological relevance of slight ^{11}C -PIB elevations in elderly normal control subjects and provided an estimate for the cut-offs defining the “gray zone” using distribution volume ratios [17]. Using histopathological confirmation, Doré et al. and La Joie et al. reported gray zones from 12.2 to 24.4 and 19 to 28 CLs, respectively [61, 62]. Although this study achieved tightly correlated results between pipelines, especially between the different centiloid methods, it should be highlighted that even for the different centiloid approaches, a range of different cut-offs was observed, which is in line with previous reports that investigated the sensitivity of centiloid quantification to pipeline design [63, 64]. The variability in the cut-offs is not attributable to random test-retest errors,

but to differences in pipeline-specific factors such as spatial processing, ROI definition, and SUVR to CL calibration bias. One centiloid pipeline used in the manuscript, the standard centiloid, adopts the approach developed by Klunk et al. (2015) and employs SPM8 for PET-MRI coregistration and normalization on the standard template, followed by the use of standard centiloid ROIs [31]. However, the other centiloid pipelines may employ different spatial processing algorithms and ROI definitions, leading to increased variability beyond the expected test–retest variability and resulting in different cut-offs for each pipeline. Additionally, the number of subjects excluded by quality control measures was low but slightly varied between pipelines, possibly leading to further increases in variability.

Limitations As a study limitation, an end-of-life population was used to estimate sensitivity and specificity with histopathological confirmation as SoT. This advanced population is characterized by anatomical abnormalities such as marked atrophy or ventricular enlargement, which hinders quantification and can pose a challenge for VA. Despite these challenges, the diagnostic efficacy of quantitative and VA was comparably good or even slightly higher for quantification in certain situations (i.e., for inexperienced readers).

Another limitation of the present study is that only indirect evidence on the adjunct value of quantification to visual assessment is provided. Such additional value can only be demonstrated in a prospective study. Furthermore, we acknowledge that an end-of-life cohort is less representative for early disease stages with lower amyloid burden. The derived sensitivity, specificity, and accuracy both for quantitative methods and readers were based on histopathology derived from such an end-of-life cohort, and these numbers should be considered in this context. However, similar or even better quantification performance is expected in an earlier study population, as fewer structural brain abnormalities are typically observed that could interfere with the assessments.

This study focused on simple binary reads for VA, which is the routine clinical method. However, recent research has shown the potential significance of amyloid PET regional quantification in staging AD [26, 27, 65], determining the risk of subsequent cognitive decline [27, 53], for optimal patient selection for anti-amyloid intervention trials [65, 66], and for reducing the sample size in anti-amyloid intervention trials [67, 68]. Pascoal et al. also showed that the topographical pattern of the PET signals in individuals with MCI who progress to dementia is “traditionally AD-like,” while that of non-converters includes more temporal and occipital regions instead [69]. The use of CL and composite SUVR is limited in determining the topographical distribution of A β load. Regional information may provide additional information to supplement

visual assessment, but it has not been assessed in this analysis and is not widely available in all the quantitative tools. Such topographical information of A β load may enable earlier identification of subjects in the early-stage AD pathological continuum and may overcome simple dichotomous measures [15].

Finally, this study has also assessed global measurement of A β load such as SUVR, centiloids, amyloid load, or amyloid index. Some amyloid PET analysis software provides additional tools, such as z-scores obtained when comparing FBB PET signal from a scan with a normal database, which may assist with the clinical diagnosis. Such tools may offer additional information to supplement VA but were not evaluated in this study. Figure 6 presents several cases where regional z-score information may provide incremental value and thus assist VA.

Conclusions

Visual binary reads of amyloid load provide the essential information for clinical routine diagnosis of AD but do not consider the wealth of information that brain PET scans provide. This study demonstrated that quantitative methods, using both CE marked software and other widely available processing tools, perform very homogeneously and robustly, providing comparable results to visual assessments of FBB PET scans. Adjunct use of quantification could be beneficial in certain situations, e.g., for newly trained or inexperienced readers, in instances when images are visually assessed with relatively low confidence, for the early detection of amyloid load or when amyloid levels are close to “pathology” thresholds, for which assessment based on VA can be difficult. Physicians should retain the careful visual inspection of the images, but quantification methods may provide additional insights in cases of doubt or for research purposes.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00259-023-06279-0>.

Acknowledgements Medical writing support was provided by Heather Gaile Marco-Gäde. We would like to express our gratitude to all study participants and investigators of the previous clinical studies.

Author contribution All authors made a substantial contribution in the interpretation of the study results and revising the manuscript critically for important intellectual content and approved the final version to be published. SB, NRV, AN, JL, JPGG, VD, PB, AW, and RG performed image analysis for individual methods. SB and AJ performed data analysis and prepared the first versions of the manuscript. SB, AJ, NRV, NK, MLQ, and AS contributed to the concept, design, and interpretation of the study.

Funding Part of the analyzed data were acquired in previous clinical studies funded by Bayer Pharma AG or Piramal Imaging. Medical writing support was funded by Life Molecular Imaging GmbH.

Data Availability Data can be made available for research purposes upon reasonable request.

Declarations

Competing interests Aleksandar Jovalekic, Núria Roé-Vellvé, Norman Koglin, Mariana Lagos Quintana, Andrew W. Stephens and Santiago Bullich are employees of Life Molecular Imaging GmbH (formerly Piramal Imaging GmbH).

Aaron Nelson is an employee of MIM Software Inc.

Markus Diemling and Johan Lilja are employees of Hermes Medical Solutions.

Juan Pablo Gómez-González is an employee of Qubitech Health Intelligence.

Vincent Doré and Pierrick Bourgeat declare no competing interests.

Alex Whittington and Roger Gunn are employees of Invicro.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Long JM, Holtzman DM. Alzheimer disease: an update on pathobiology and treatment strategies. *Cell*. 2019;179(2):312–39.
- Milà-Alomà M, et al. Comparative analysis of different definitions of amyloid- β positivity to detect early downstream pathophysiological alterations in preclinical Alzheimer. *J Prev Alzheimer's Dis*. 2021;8(1):68–77.
- Sevigny J, et al. Amyloid PET screening for enrichment of early-stage Alzheimer disease clinical trials: experience in a phase 1b clinical trial. *Alzheimer Dis Assoc Disord*. 2016;30(1):1–7.
- Roberts C, et al. Amyloid and APOE status of screened subjects in the elenbecestat MissionAD phase 3 program. *J Prev Alzheimer's Dis*. 2021;8(2):218–23.
- Pemberton HG, et al. Quantification of amyloid PET for future clinical use: a state-of-the-art review. *Eur J Nucl Med Mol Imaging*. 2022;49(10):3508–28.
- Life Molecular Imaging, Neuraceq USPI. https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/204677Orig1s021lbl.pdf. Accessed 25 Jan 2023.
- Life Radiopharma Berlin, Neuraceq SmPC. https://www.ema.europa.eu/en/documents/product-information/neuraceq-epar-product-information_en.pdf. Accessed 25 Jan 2023.
- Sabri O, et al. Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: phase 3 study. *Alzheimers Dement*. 2015;11(8):964–74.
- Seibyl J, et al. Impact of training method on the robustness of the visual assessment of 18F-florbetaben PET scans: results from a phase-3 study. *J Nucl Med*. 2016;57(6):900–6.
- Bullich S, et al. Optimized classification of (18)F-florbetaben PET scans as positive and negative using an SUVR quantitative approach and comparison to visual assessment. *Neuroimage Clin*. 2017;15:325–32.
- Perani D. FDG-PET and amyloid-PET imaging: the diverging paths. *Curr Opin Neurol*. 2014;27(4):405–13.
- Lilja J, et al. Spatial normalization of (18)F-flutemetamol PET images using an adaptive principal-component template. *J Nucl Med*. 2019;60(2):285–91.
- Pontecorvo MJ, et al. Quantitation of PET signal as an adjunct to visual interpretation of florbetapir imaging. *Eur J Nucl Med Mol Imaging*. 2017;44(5):825–37.
- Bullich S, et al. Early detection of amyloid load using (18)F-florbetaben PET. *Alzheimers Res Ther*. 2021;13(1):67.
- Fantoni E, et al. The spatial-temporal ordering of amyloid pathology and opportunities for PET imaging. *J Nucl Med*. 2020;61(2):166–71.
- Hosokawa C, et al. Performance of 11C-Pittsburgh compound B PET binding potential images in the detection of amyloid deposits on equivocal static images. *J Nucl Med*. 2015;56(12):1910–5.
- Mormino EC, et al. Not quite PIB-positive, not quite PIB-negative: slight PIB elevations in elderly normal control subjects are biologically relevant. *Neuroimage*. 2012;59(2):1152–60.
- Barthel H, Seibyl J, Sabri O. Yes we can analyse amyloid images - now what? *Eur J Nucl Med Mol Imaging*. 2017;44(5):822–4.
- Klein G, et al. Gantenerumab reduces amyloid-beta plaques in patients with prodromal to moderate Alzheimer's disease: a PET substudy interim analysis. *Alzheimers Res Ther*. 2019;11(1):101.
- Mintun MA, et al. Donanemab in early Alzheimer's disease. *N Engl J Med*. 2021;384(18):1691–704.
- Shcherbinin S, et al. Association of amyloid reduction after donanemab treatment with tau pathology and clinical outcomes: the TRAILBLAZER-ALZ randomized clinical trial. *JAMA Neurol*. 2022;79(10):1015–24.
- Swanson CJ, et al. A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-A β protofibril antibody. *Alzheimers Res Ther*. 2021;13(1):80.
- van Dyck CH, et al. Lecanemab in early Alzheimer's disease. *N Engl J Med*. 2022;388(1):9–21.
- Budd Haeberlein S, et al. Two randomized phase 3 studies of aducanumab in early Alzheimer's disease. *J Prev Alzheimer's Dis*. 2022;9(2):197–210.
- FDA. FDA grants accelerated approval for Alzheimer's disease treatment. <https://www.fda.gov/news-events/press-announcements/fda-grants-accelerated-approval-alzheimers-disease-treatment>. Accessed 6 Jan 2023.
- Grothe MJ, et al. In vivo staging of regional amyloid deposition. *Neurology*. 2017;89(20):2031–8.
- Collij LE, et al. Multitracer model for staging cortical amyloid deposition using PET imaging. *Neurology*. 2020;95(11):e1538–53.
- Ong KT, et al. Abeta imaging with 18F-florbetaben in prodromal Alzheimer's disease: a prospective outcome study. *J Neurol Neurosurg Psychiatry*. 2015;86(4):431–6.
- Rowe CC, et al. (18)F-Florbetaben PET beta-amyloid binding expressed in CENTILOIDS. *Eur J Nucl Med Mol Imaging*. 2017;44(12):2053–9.
- Piper J, Nelson A, Javorek A. Evaluation of a quantitative method for florbetaben (FBB) PET using SUVR. in EANM. 2014.
- Klunk WE, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement*. 2015;11(1):1–15 e1–4.
- Leuzy A, et al. Derivation and utility of an Abeta-PET pathology accumulation index to estimate Abeta load. *Neurology*. 2020;95(21):e2834–44.
- Zhou L, et al. MR-less surface-based amyloid assessment based on 11C PiB PET. *PLoS ONE*. 2014;9(1): e84777.
- Bourgeat P, et al. Comparison of MR-less PiB SUVR quantification methods. *Neurobiol Aging*. 2015;36(Suppl 1):S159–66.
- Bourgeat P, et al. Implementing the centiloid transformation for (11)C-PiB and beta-amyloid (18)F-PET tracers using CapAIBL. *Neuroimage*. 2018;183:387–93.

36. Bourgeat P, et al. Non-negative matrix factorisation improves Centiloid robustness in longitudinal studies. *Neuroimage*. 2021;226:117593.
37. Whittington A, et al. Spatiotemporal distribution of beta-amyloid in Alzheimer disease is the result of heterogeneous regional carrying capacities. *J Nucl Med*. 2018;59(5):822–7.
38. Whittington A, Gunn RN; I. Alzheimer's Disease Neuroimaging. Amyloid load: a more sensitive biomarker for amyloid imaging. *J Nucl Med*. 2019;60(4):536–540.
39. Kim JY, et al. Visual interpretation of [(18)F]florbetaben PET supported by deep learning-based estimation of amyloid burden. *Eur J Nucl Med Mol Imaging*. 2021;48(4):1116–23.
40. Bucci M, et al. A multisite analysis of the concordance between visual image interpretation and quantitative analysis of [(18)F]flutemetamol amyloid PET images. *Eur J Nucl Med Mol Imaging*. 2021;48(7):2183–99.
41. Cattell L, et al. Classification of amyloid status using machine learning with histograms of oriented 3D gradients. *Neuroimage Clin*. 2016;12:990–1003.
42. Cho K, et al. Classification of 18F-florbetaben amyloid brain PET image using PCA-SVM. *Biomed Sci Lett*. 2019;25(1):99–106.
43. Cho SH, et al. Concordance in detecting amyloid positivity between (18)F-florbetaben and (18)F-flutemetamol amyloid PET using quantitative and qualitative assessments. *Sci Rep*. 2020;10(1):19576.
44. Giovacchini G, et al. Impact of tracer retention levels on visual analysis of cerebral [(18)F]-florbetaben PET images. *Curr Radiopharm*. 2021;14(1):70–7.
45. Kim H-R, et al. Finding the optimal cutoff value for amyloid β positivity using the iterative outlier method and concordance rate. *Precis Futur Med*. 2021;5(2):83–9.
46. Kim JP, et al. Staging and quantification of florbetaben PET images using machine learning: impact of predicted regional cortical tracer uptake and amyloid stage on clinical outcomes. *Eur J Nucl Med Mol Imaging*. 2020;47(8):1971–83.
47. Whittington A, et al. AmyloidQ – an automated algorithm for classifying amyloid- β PET scans. In: *Human Amyloid Imaging*. 2019. <https://hai.worldeventsforum.com/wp-content/uploads/2019/01/HAI-Book-Print-Jan-8.pdf>. Accessed 25 Jan 2023.
48. Zeltzer E, et al. Quantitative analysis of 6,150 real-world amyloid positron emission tomography (PET) scans from the imaging dementia—evidence for amyloid scanning (IDEAS) study. *Alzheimer's Dement*. 2022;18(S6): e066217.
49. Clark CM, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid- β plaques: a prospective cohort study. *Lancet Neurol*. 2012;11(8):669–78.
50. Curtis C, et al. Phase 3 trial of flutemetamol labeled with radioactive fluorine 18 imaging and neuritic plaque density. *JAMA Neurol*. 2015;72(3):287–94.
51. Martínez G, et al. 18F PET with florbetaben for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev*. 2017;11(11):Cd012883.
52. Collij LE, et al. Quantification of [(18)F]florbetaben amyloid-PET imaging in a mixed memory clinic population: the ABIDE project. *Alzheimers Dement*. 2022. <https://doi.org/10.1002/alz.12886>.
53. Hanseeuw BJ, et al. Defining a Centiloid scale threshold predicting long-term progression to dementia in patients attending the memory clinic: an [18F]flutemetamol amyloid PET study. *Eur J Nucl Med Mol Imaging*. 2021;48(1):302–10.
54. van der Kall LM, et al. Association of β -amyloid level, clinical progression, and longitudinal cognitive change in normal older individuals. *Neurology*. 2021;96(5):e662–70.
55. Ossenkoppele R, et al. Amyloid and tau PET-positive cognitively unimpaired individuals are at high risk for future cognitive decline. *Nat Med*. 2022;28(11):2381–7.
56. Jagust WJ, Landau SM. Temporal dynamics of β -amyloid accumulation in aging and Alzheimer disease. *Neurology*. 2021;96(9):e1347–57.
57. Matsuda H, et al. Quantitative evaluation of 18F-flutemetamol PET in patients with cognitive impairment and suspected Alzheimer's disease: a multicenter study. *Front Neurol*. 2021;11:578753.
58. Collij LE, et al. Regional amyloid accumulation predicts memory decline in initially cognitively unimpaired individuals. *Alzheimers Dement (Amst)*. 2021;13(1): e12216.
59. Bullich S, et al. Converging evidence for a “gray-zone” of amyloid burden and its relevance. *Alzheimers Dement*. 2020;16(S5): e044786.
60. Salvado G, et al. Centiloid cut-off values for optimal agreement between PET and CSF core AD biomarkers. *Alzheimers Res Ther*. 2019;11(1):27.
61. Dore V, et al. Comparison of (18)F-florbetaben quantification results using the standard Centiloid, MR-based, and MR-less CapAIBL((R)) approaches: validation against histopathology. *Alzheimers Dement*. 2019;15(6):807–16.
62. La Joie R, 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(2):205–16.
63. Su Y, et al. Utilizing the Centiloid scale in cross-sectional and longitudinal PiB PET studies. *NeuroImage: Clinical* 2018;19:406–416.
64. Shekari M, et al. Evaluating the sensitivity of Centiloid quantification to pipeline design and image resolution. *Alzheimers Dement*. 2022;18(S6): e067896.
65. Mattsson N, et al. Staging beta-amyloid pathology with amyloid positron emission tomography. *JAMA Neurol*. 2019;76(11):1319–29.
66. Guo T, et al. Rate of beta-amyloid accumulation varies with baseline amyloid burden: Implications for anti-amyloid drug trials. *Alzheimers Dement*. 2018;14(11):1387–96.
67. Insel PS, et al. Neuroanatomical spread of amyloid beta and tau in Alzheimer's disease: implications for primary prevention. *Brain Commun*. 2020;2(1):fcaa007.
68. Lopes Alves I, et al. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. *Alzheimers Dement*. 2020;16(5):750–8.
69. Pascoal TA, et al. Topographical distribution of Abeta predicts progression to dementia in Abeta positive mild cognitive impairment. *Alzheimers Dement (Amst)*. 2020;12(1): e12037.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.