Evaluation of novel data-driven metrics of amyloid β deposition for longitudinal PET studies

Ariane Bollack , Pawel J Markiewicz , Alle Meije Wink , Lloyd Prosser , Johan Lilja , Pierrick Bourgeat , Jonathan M Schott , William Coath , Lyduine E Collij , Hugh G Pemberton , Gill Farrar , Frederik Barkhof , David M Cash , on behalf on the AMYPAD consortium

PII: S1053-8119(23)00464-0
DOI: https://doi.org/10.1016/j.neuroimage.2023.120313
Reference: YNIMG 120313

To appear in: *NeuroImage*

Received date: 5 January 2023
Revised date: 29 May 2023
Accepted date: 4 August 2023

Please cite this article as: Ariane Bollack , Pawel J Markiewicz , Alle Meije Wink , Lloyd Prosser , Johan Lilja , Pierrick Bourgeat , Jonathan M Schott , William Coath , Lyduine E Collij , Hugh G Pemberton , Gill Farrar , Frederik Barkhof , David M Cash , on behalf on the AMYPAD consortium, Evaluation of novel data-driven metrics of amyloid β deposition for longitudinal PET studies, *NeuroImage* (2023), doi: https://doi.org/10.1016/j.neuroimage.2023.120313

Highlights

- Novel data-driven approaches to quantifying amyloid burden from in-vivo PET imaging have been developed to address common sources of variability such as the dependence on the definition of a reference region.

- We evaluated four novel data-driven amyloid metrics against the binding potential and Centiloid scale using consistent data and evaluation criteria.

- Three metrics (Aβ load, Aβ index, CL_{NMF}) showed strong association with the binding potential and Centiloid scale. The CL_{NMF} and Aβ load could offer a more precise alternative to the more established Centiloid.

- As the Aβ load, Aβ index and CL_{NMF} rely on the decomposition of PET images in specific and non-specific binding components, our results are consistent with a multivariate pattern of amyloid deposition over time.

Evaluation of novel data-driven metrics of amyloid β deposition for longitudinal PET studies

Ariane Bollack[1], Pawel J Markiewicz[1], Alle Meije Wink[2], Lloyd Prosser[3], Johan Lilja[4], Pierrick Bourgeat[5], Jonathan M Schott[3], William Coath[3], Lyduine E Collij[2,6], Hugh G Pemberton[1,7,8], Gill Farrar[7], Frederik Barkhof[1,2,8], David M Cash[2,9], *on behalf on the AMYPAD consortium*

[1] Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, UCL, London, UK
[2] Amsterdam UMC, location VUmc, Department of Radiology and Nuclear Medicine, Amsterdam, The Netherlands
[3] Dementia Research Centre, UCL Queen Square Institute of Neurology, London, UK
[4] Hermes Medical Solutions, Stockholm, Sweden
[5] The Australian e-Health Research Centre, CSIRO, Brisbane, Australia
[6] Amsterdam Neuroscience, Brain Imaging, Amsterdam, the Netherlands
[7] GE Healthcare, Amersham, UK
[8] Queen Square Institute of Neurology, University College London, UK
[9] UK Dementia Research Institute at University College London, London, UK

Corresponding author: Ariane Bollack ariane.bollack.19@ucl.ac.uk

ORCID iDs

AB – https://orcid.org/0000-0002-9169-7530
PM – https://orcid.org/0000-0002-3114-0773
AMW - https://orcid.org/0000-0002-8197-0118
LP - https://orcid.org/0000-0003-0299-1525
JL - https://orcid.org/0000-0002-9752-6142
PB - https://orcid.org/0000-0002-2605-4766
JS - https://orcid.org/0000-0003-2059-024X
WC - https://orcid.org/0000-0002-3976-9461
LC – https://orcid.org/0000-0001-6263-1762
HP – https://orcid.org/0000-0001-8419-6423
GF – https://orcid.org/0000-0002-0726-723X
FB – https://orcid.org/0000-0003-3543-3706
DC – https://orcid.org/0000-0001-7833-616X

Key words: quantification, longitudinal, machine learning, amyloid, PET, Alzheimer's

Abstract

Purpose

Positron emission tomography (PET) provides *in vivo* quantification of amyloid-β (Aβ) pathology. Established methods for assessing Aβ burden can be affected by physiological and technical factors. Novel, data-driven metrics have been developed to account for these sources of variability. We aimed to evaluate the performance of four data-driven amyloid PET metrics against conventional techniques, using a common set of criteria.

Methods

Three cohorts were used for evaluation: Insight 46 (*N*=464, [^{18}F]florbetapir), AIBL (*N*=277, [^{18}F]flutemetamol), and an independent test-retest data (*N*=10, [^{18}F]flutemetamol). Established metrics of amyloid tracer uptake included the Centiloid (CL) and where dynamic data was available, the non-displaceable binding potential (BP_{ND}). The four data driven metrics computed were the amyloid load (Aβ load), the Aβ PET pathology accumulation index (Aβ index), the Centiloid derived from non-negative matrix factorisation (CL_{NMF}), and the amyloid pattern similarity score (AMPSS). These metrics were evaluated using reliability and repeatability in test-retest data, associations with BP_{ND} and CL, and sample size estimates to detect a 25% slowing in Aβ accumulation.

Results

All metrics showed good reliability. Aβ load, Aβ index and CL_{NMF} were strong associated with the BP_{ND}. The associations with CL suggests that cross-sectional measures of CL_{NMF}, Aβ index and Aβ load are robust across studies. Sample size estimates for secondary prevention trial scenarios were the lowest for CL_{NMF} and Aβ load compared to the CL.

Conclusion

Among the novel data-driven metrics evaluated, the Aβ load, the Aβ index and the CL_{NMF} can provide comparable performance to more established quantification methods of Aβ PET tracer uptake. The

CL_{NMF} and Aβ load could offer a more precise alternative to CL, although further studies in larger cohorts should be conducted.

LIST OF ABBREVIATIONS

AD – Alzheimer's disease

ADNI – Alzheimer's Disease Neuroimaging Initiative

AIBL – Australian Imaging Biomarkers and Lifestyle Study of Ageing

AMPSS – amyloid pattern similarity score

Aβ index – Aβ-PET pathology accumulation index

ARC – annualised rate of change

BP_{ND} – non-displaceable binding potential

CL – Centiloid scale

CL_{NMF} – Centiloid scale based on non-negative matrix factorisation

CSF – cerebrospinal fluid

DARTEL – diffeomorphic anatomical registration through exponentiated lie algebra

DVR – distribution volume ratio

GAIIN – Global Alzheimer's Association Interactive Network

GIF – geodesic information flow

ICC – intraclass correlation coefficient

K – amyloid carrying capacity

LME – linear mixed effects models

MCI – mild cognitive impairment

MNI – Montreal Neurological Institute

MRI – magnetic resonance imaging

NifTI – Neuroimaging Informatics Technology Initiative

NS – non-specific binding

p.i. – post injection

PCA – principal component analysis

PET – positron emission tomography

ROI – region of interest

SMC – subjective memory complaint

SUVr – standard uptake value ratio

SVM – support vector machine

INTRODUCTION

Longitudinal Positron Emission Tomography (PET) studies measuring the rate of change in amyloid and tau burden can provide a better understanding of the earliest stages and subsequent spread of Alzheimer's disease (AD) pathology. Measuring small spatiotemporal changes in protein deposition with high accuracy and precision can facilitate early detection and tracking of the disease [1], increase the prognostic value by determining the risk of cognitive decline [2], and better inform clinical trial design [3]. PET-based imaging endpoints are particularly relevant for regulatory approval of an anti-amyloid therapy as a disease modifying treatment for individuals with AD [4,5], especially for preclinical trials like the AHEAD 3-45 study [6–9]. In clinical practice, quantification plays a supplemental role in assisting visual reads of scans in the earliest stages of pathological build-up, where there is lower confidence in image interpretation [10]. When disease-modifying therapies would become more widely available, quantification might also be used in response monitoring with more precision than a discrete visual rating scale [11], and help inform treatment efficacy decisions [12].

The accuracy and precision of PET data quantification is influenced by the measure used to determine the cortical amyloid β (Aβ) burden and rate of accumulation. Quantitative PET measurements based on kinetic modelling of fully dynamic acquisitions can best account for specific and non-specific contributions to tracer uptake. Although the non-displaceable binding potential (BP_{ND}) derived from those measurements is considered the most accurate in the absence of histopathology, it is often not feasible to acquire this data in most patients. Thus, the most common technique to measure amyloid burden is the standard uptake value ratio (SUVr) using shorter static acquisitions at a pre-determined period post-injection, where the tracer is assumed to have reached pseudo-equilibrium. The SUVr corresponds to the ratio between the average uptake of the

radiotracer in a target region of interest (ROI) and a reference region which is assumed to have no specific tracer uptake. To mitigate biases and variabilities attributed to various tracer effects, the Centiloid scale (CL) was introduced as a means of calibrating measures of amyloid burden to a tracer-independent, unbounded scale where 0 (PET signal characteristic of young healthy controls) and 100 (typical AD subjects) serve as anchor points. The calibration that the Centiloid scale offers may be advantageous for multi-centre and multi-tracer studies [13].

However, semi-quantitative measures obtained with SUVr and CL can be affected by a wide range of additional biological and technical sources of error and variability [14], such as the reliance on an MR for accurate quantification. In addition, the choice of reference region is a recurrent subject of debate as the tracer used, the stability and the definition of the reference region significantly affect the amyloid signal over time [15–18]. Accurate spatial definition of the ROIs relies on the quality of the segmentation, and in particular its ability to capture an accurate sampling of non-specific uptake. Finally, some reference regions have also shown evidence of amyloid pathology in specific forms of AD, such as cerebellar Aβ plaques in autosomal dominant forms of AD [19].

To alleviate some of these concerns for model-based analysis methods, novel data-driven methods to measure Aβ deposits have been proposed:

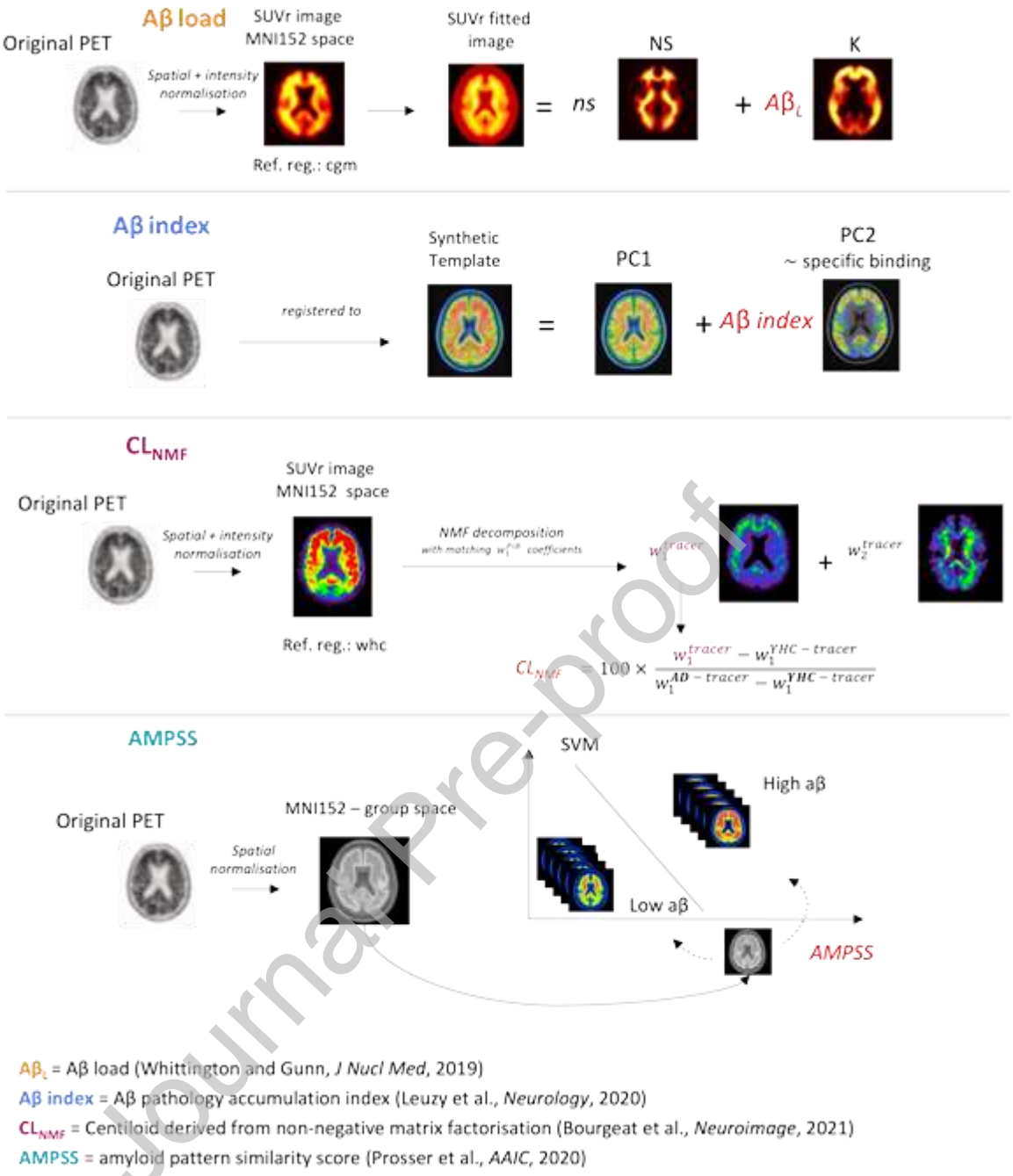
- The amyloid load (Aβ load) was developed by Whittington and Gunn and is based on image decomposition to extract the specific tracer binding. This decomposition is based upon characterizing the SUVr trajectory over the disease course with a logistic growth model, assuming unique maximum amyloid carrying capacity for each brain region [20].
- The Aβ-PET pathology accumulation index (Aβ index) by Leuzy *et al.* relies on a principal component analysis (PCA) approach where one component, representing the tracer specific binding to Aβ is estimated as part of a registration process between a new image and an adaptive group template [21].

- The improved Centiloid scale (CL_{NMF}) developed by Bourgeat *et al.*, uses non-negative matrix factorisation (NMF), another type of image decomposition, to generate a metric less susceptible to tracer-related changes in longitudinal studies compared to the original CL scale [22].
- The Amyloid Pattern Similarity Score (AMPSS) proposes a classifier-based approach that uses a support vector machine (SVM) to identify the multivariate patterns of amyloid burden, and produces a probabilistic score reflecting this burden relative to a reference population [23].

These four metrics have been developed using independent datasets and have been validated using different sets of evaluation criteria, such as agreement with the SUVr and histopathology, or precision in detecting changes over time. However, there has yet to be a systematic validation of these novel data-driven metrics using consistent data and evaluation criteria. This study aims to evaluate these data-driven metrics based on their performance against metrics established in the literature, in terms of repeatability, reliability, association with BP_{ND} and CL, and sample size requirements to power clinical trials. These assessments have been conducted using two longitudinal datasets and a test-retest dataset. In addition, we also considered the advantages and limitations in the implementations of those metrics to best assess how they could help capture small changes in protein deposition over time.

MATERIALS & METHODS

The four data-driven approaches that were considered are described below: the Aβ load [20], the Aβ index [21], the CL_{NMF} [22], and the AMPSS [23] (key concepts detailed in Supplementary material). To simplify the evaluation of these metrics, we summarised the key implementation steps and harmonised the nomenclature around the different aspects of these methods (see **Figure 1**). A comparative overview of the different methods can be found in **Table 1**.



$A\beta_L$ = Aβ load (Whittington and Gunn, *J Nucl Med*, 2019)
Aβ index = Aβ pathology accumulation index (Leuzy et al., *Neurology*, 2020)
CL_{NMF} = Centiloid derived from non-negative matrix factorisation (Bourgeat et al., *Neuroimage*, 2021)
AMPSS = amyloid pattern similarity score (Prosser et al., *AAIC*, 2020)

Figure 1 Key concepts behind four novel metrics of amyloid deposition

Abbreviations – SUVr: standard uptake value ratio; Ref. reg.: reference region used for SUVr computation; CGM: cerebellar grey matter reference region; WHC: whole cerebellum reference region; PC: principal component; NS: non-specific; K: amyloid carrying capacity; NMF: non-negative matrix factorisation.

Main characteristics	Aβ load *Whittington et al., 2019*	Aβ index *Leuzy et al., 2020*	CL_{NMF} *Bourgeat et al., 2021*	AMPSS *Prosser et al.*
Key idea	Image decomposition into 2 components: non-specific binding and Aβ carrying capacity determined *via* logistic growth model	Image decomposition *via* principal component analysis isolates specific binding	Image decomposition *via* non-negative matrix factorisation (NMF)	Support vector machine to produce probabilities score from little to no specific binding through to AD-like scans
Assumptions	Spatially synchronised accumulation according to the maximum amyloid carrying capacity of each region	Second principal component represents specific binding	First component represents specific binding	-
Range	% unbounded	-1 to 1 unbounded	0 to 100 0 and 100 are anchor points	% bounded (or unbounded using a logit transform)
Reference region independent	No	No	No	Yes
MR independent	No	Yes (but No for training)	No	No
Specificity when processing scans from different tracers	Needs to match NS_K image scale [24]	Principal components specific to each tracer	NMF components specific to each tracer	Training on tracer specific datasets
Possible application for tau	Implemented [25]	Not implemented, comparable approach by Cho *et al.* [26]	N/A	Not implemented
Validation	Against SUVr on ADNI [20] and GAAIN data [24]	Against SUVr, CSF, visual read and neuropathology on BioFINDER and ADNI data [21,27]	Against standard CL on GAAIN and AIBL data [22]	Against SUVr and CSF Aβ42 using ADNI data
Availability	Available *via* Invicros's IQ Analytics Platform	Software freely available for research upon request	Open source https://doi.org/10.25919/5f8400a0b6a1e.	Plans to make it open source
Implementation in studies	Zammit et al., 2019, 2021 [28,29]	Haller et al., 2021 [30]	-	-
Main strengths	- Increased sensitivity for longitudinal change in amyloid load - Implemented for all amyloid tracers (PiB [29]; ^{18}F tracers [24]) - Implemented for tau [25]	- MR independent - Reference region independent - Software includes pre-processing - Fully automatic process (~20 seconds)	- Robustness to change in tracer in a longitudinal setting - Improve longitudinal consistency compared to CL - Implemented for all amyloid tracers	Reference region independent
Main limitations	Relies on a reference region	Relies on a reference region for training	- Relies on a reference region - Sub-optimal decomposition for ^{18}F tracers	Sensitivity to training set
Possible improvements	-	Allow for more principal components	Independence from MR using CapAIBL [31,32]	Independent from MR

Table 1 Overview of four data-driven metrics of Aβ deposition

Abbreviations - NS: non-specific; K: maximum amyloid carrying capacity; ADNI: Alzheimer's Disease Neuroimaging Initiative; AIBL: Australian Imaging Biomarkers and Lifestyle Study of Ageing; GAIIN: Global Alzheimer's Association Interactive Network; AMPSS: amyloid pattern similarity score; SUVr: standard uptake value ratio; CL: Centiloid scale; CSF: cerebrospinal fluid.

Data acquisition

Data from three separate cohorts was used (see **Table 2**).

Insight46

First, 350 subjects from Insight46, a prospective neuroscience sub-study of the MRC National Survey of Health and Development [33], with baseline and follow-up dynamic PET-MR scans (follow-up time: 2.4 ± 0.2 years) acquired with [^{18}F]florbetapir were included. All study members were born in the same week of 1946. The majority were cognitively normal.

Images were acquired on the same Biograph mMR 3T PET/MRI scanner (Siemens Healthcare, Erlangen). The full study protocol is described elsewhere [33]. In short, 370 MBq of [^{18}F]florbetapir (Amyvid) was injected intravenously. PET data were acquired continuously for ~60 minutes from the time of injection to allow both full kinetic modelling and static analysis during the last ~10 min of scanning (from 50 to 60 min). Both static and dynamic data were acquired in list-mode and reconstructed offline via a 3D OSEM algorithm with three iterations and 21 subsets, using the software Niftypet [34]. Attenuation correction was based on a pseudo-CT synthesis from T1 and T2 weighted MR images [35,36]. The dynamic acquisition consisted of 31 frames (4x15 seconds, 8x30 seconds, 9x60 seconds, 2x180 seconds, and 8x300 seconds). Images were smoothed with a 4mm Gaussian kernel with no partial volume correction was applied and were then used as inputs for the image processing and analysis pipelines.

AIBL

We included 277 subjects from the Australian Imaging, Biomarkers and Lifestyle (AIBL) flagship study of aging with PET scans acquired at up to three timepoints using [^{18}F]flutemetamol [37]. The AIBL study was designed in part to determine the individual risk of developing AD, the time from symptom onset, as well as specific risk factors, for which healthy controls (HC; *N*=69), and subjects from the AD spectrum were recruited, including 118 subjects with subjective memory complaint (SMC), 64 with mild cognitive impairment (MCI), 17 with AD, and 9 with another diagnosis.

PET images were acquired for 20 minutes, starting at 90min after intravenous injection of approximately 185 MBq of [^{18}F]flutemetamol. The AIBL data is a multi-centre study using several PET scanners: Allegro Body (Philips Medical Systems), Biograph128 (Siemens Healthcare, Erlangen), Discovery (GE Medical Systems). Images were reconstructed iteratively via 3D RAMLA, 3D OSEM or VPHD algorithms respectively of their scanners, then smoothed with a 6mm Gaussian kernel. No partial volume correction was applied. The full study protocol is described elsewhere [37]. Images were then used as inputs for the image processing and analysis pipelines.

Test-retest data

[^{18}F]florbetapir test-retest data (Insight46 cohort)

Because there were no dedicated test-retest studies for [^{18}F]florbetapir available, we identified individuals from Insight46 who are unlikely to accumulate Aβ over time as a proxy for test-rest [^{18}F]florbetapir data, as in Landau *et al.* [38]. Subjects were included based on the following criteria:

(1) both baseline and follow-up PET available (follow-up time: 2.4 ± 0.1 years), (2) CSF Aβ42/40 and ptau181 available at follow-up, (3) at follow-up, CSF Aβ42/40 value in the top quartile, (4) at follow-up, CSF ptau181 <=57 pg/ml (cut-off from the manufacturer and further validated [39]), (5) no mild cognitive impairment or major brain disorder at baseline (based clinical consensus criteria [40]), yielding 16 individuals.

[^{18}F]flutemetamol test-retest data (independent cohorts)

Finally, 10 test-retest scans from subjects with an AD diagnosis acquired using [18F]flutemetamol were also included [41,42].

The test-retest [18F]flutemetamol data comprised of two [18F]flutemetamol studies: 5 subjects from a Japanese Phase II study (GE067-017) with scans acquired 1-4 weeks apart [42] and 5 subjects from a European Phase II study (ALZ201) with scans acquired 7 days apart [41]. All subjects had a confirmed AD diagnosis. The 10 subjects were injected ~120 MBq of the tracer at each timepoint. PET images were acquired from ~90min to 120min post injection (p.i.) [42] or ~85min to 115min p.i. [41]. PET scans were reconstructed into 6x5min frames either iteratively or *via* filtered back projection depending on the scanner used. Images were smoothed with a 6mm Gaussian kernel. No partial volume correction was applied.

	N Subjects (1st timepoint)	Age (years)	Gender (% female)	ApoE4 ε4 (% carriers)	*N* Timepoints	Follow-up time	Radiotracer	Dynamic acquisition
Insight 46	350	70.6 ± 0.7	48.9	29.3	2	2.4 ± 0.2 years	^{18}F-florbetapir	yes
AIBL	277	73.4 ± 5.9	51.6	30.4	3	fu1: 1.6 ± 0.6 years fu2: 3.2 ± 0.5 years	^{18}F-flutemetamol	no
Test-retest								
^{18}F-florbetapir test-retest (Insight46 cohort)	16	70.5 ± 0.6	25	-	2	2.4 ± 0.1 years	^{18}F-florbetapir	yes
^{18}F-flutemetamol test-retest (independent cohort)	10	72.5 ± 6.5	40.0	-	2	7 days or 1-4 weeks	^{18}F-flutemetamol	no

Table 2 Demographic characteristics for each dataset

Where applicable, values are described as mean ± standard deviation.

Abbreviations - ApoE ε4 carriers: carriers of at least one copy of the ε4 allele of the APOE gene; fu= follow-up

Image processing

Harmonised pre-processing

Except for the Aβ index which does not require any pre-processing, each of the amyloid metrics uses a slightly different pipeline. For a fairer comparison, we tried to harmonise these steps as much as possible. When required by the methodologies (see **Table S1**), the PET scans were registered into Montreal Neurologic Institute (MNI) 152 space [43]. First, $^{[18F]}$florbetapir or $^{[18F]}$flutemetamol scans were co-registered to the corresponding T1w MR using a rigid body transformation with the SPM 12 toolbox (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London, UK) [44]. Using diffeomorphic nonlinear registration tools (DARTEL) [45], the bias-corrected T1w scans were used to create a group average template that was then registered into MNI152 space. Finally, the

DARTEL flow fields were applied (without modulation) to the PET data to register the scans into MNI152 space.

Conventional metrics

T1w images were parcellated with a multi-atlas based segmentation using geodesic information flow (GIF) [46]. The labels were propagated to the PET scan through the transformation obtained from the PET-to-MR registration. Centiloid values were computed with the whole cerebellum as reference region, as in the recommended processing guidelines [13]. SUVr was transformed into Centiloid using the previously published calibration equations for test-retest [^{18}F]flutemetamol [47] and Insight46 [^{18}F]florbetapir data [48]. We focused on Centiloid values though SUVr results can be found in Supplementary material. The dynamic analysis was performed with tissue activity curves fitted using the reference Logan method [49] with a start time of 30 minutes, from which the distribution volume ratio (DVR) was derived, and BP_{ND} is defined as DVR - 1. The mean time activity curve was obtained from the same composite ROI as used in the static analysis and with the cerebellar grey matter as the reference region.

Data-driven metrics

All data-driven metrics except for AMPSS and Aβ load used the training sets as laid out in the original papers. Specific processing requirements and implementation for the data-driven metrics can be found in Supplementary material.

Evaluation of the performance and statistical analysis

To evaluate the performance of these novel data-driven metrics against established metrics in the literature, we compared these methodologies based on the following criteria (1) repeatability and

reliability; (2) strength of the association with the BP_{ND}, which is considered to be closer to ground truth than other available metrics, (3) strength of the association with the CL scale; (4) variability of the annualised rates of change and (5) sample size requirements to power a clinical trial aimed at detecting a 25% decrease in amyloid accumulation over 2 years. Statistical analyses were performed using R, version 4.0.4 (R Program for Statistical Computing) [50].

Repeatability and Reliability

Repeatability of both the data-driven metrics and conventional measures of amyloid burden were assessed according to the standard deviation between test and retest scans, decomposed into between subject and within-subject components. These metrics have different ranges and various offsets, while being either bounded or not. Discrepancies in their mathematical definition hence restricts our ability to perform head-to-head comparisons.

The test-retest reliability was assessed using the intraclass correlation coefficient (ICC) [51]. ICC estimates and their 95% confidence intervals were calculated using R *irr* package based on a mean-rating (k = 2), absolute-agreement and a 2-way mixed-effects model, which assumes the same measurement error at each timepoint.

Two of the cohorts in this study provided some capability for assessing reliability of the data driven metrics. The first was a [^{18}F]florbetapir test-retest dataset build selection of 16 stable subjects from Insight46, and the second was the [^{18}F]flutemetamol test-retest dataset (from independent cohorts) with 10 individuals with high amyloid burden and short interval between scans.

Correlation with BP_{ND}

No ground truth, such as histopathology, exists for *in vivo* quantification of amyloid burden, however, BP_{ND} obtained from dynamic acquisitions using kinetic modelling accounts for variations in physiological factors such as blood flow alterations and radiotracer clearance [14,52]. These advantages provide a more accurate representation of the true measurements of specific amyloid

binding compared to the SUVr. We therefore computed the repeated measurement correlation between each data driven metric and the BP_{ND}. The repeated measurement correlation (implemented in the R package *rmcorr*) was chosen to take into consideration the non-independence of the measures obtained for the same individual over time [53]. It does so by shifting the common use of analysis of variance (ANOVA) and using it to determine the relationship between two continuous variables (amyloid metrics at baseline and follow-up) while controlling for between-subject variance. From this atypical ANOVA implementation can be derived a repeated measurement correlation coefficient (r_{rm}), which here represents the within-person change in one metric versus the BP_{ND}, and is defined as follows:

$$r_{rm} = \sqrt{\frac{SS_{measure}}{SS_{measure} + SS_{error}}}$$

with $SS_{measure}$ and SS_{error} representing the sum of squares of the measures and errors respectively. 95% confidence intervals for r_{rm} were built using 2,000 bootstrap replicates. This was only evaluated in the Insight46 dataset, as the AIBL data acquisitions included only late static frames.

Correlation with CL

The repeated measurement correlation of each metric with CL values was also evaluated, as it allows for the comparison of [^{18}F]flutemetamol (AIBL) and [^{18}F]florbetapir (Insight46) data. In each case, the strength of the association was determined by the repeated measurement correlation r_{rm} and 95% confidence intervals for r_{rm} were built using 2,000 bootstrap replicates. The CL scale being developed to harmonise amyloid measurements across tracers, we would not expect to observe a significant difference in longitudinal change across metrics.

Annualised rates of change

For Insight46, the annualised rates of changes (ARC) were computed as the difference between follow-up and baseline metric over the time interval between scans. As the AIBL data comprises a subset of subject with three timepoints (*N*=119), the amyloid accumulation rates were estimated with linear mixed effects models (LME; using the *lme4* package in R). We included both random slopes and random intercepts to account for inter-individual variability, with an unstructured covariance matrix allowing for a covariance between the slope and intercept. With dt representing the time since the baseline scans in years, the LME model was as follows:

$$y_{i,j} = \beta_{0,i} + \beta_{1,i} t_{i,j} + e_{i,j}$$

$$\text{with, } e_{i,j} \sim N(0, \sigma_e^2), \; for\ i \neq k, cov(u_i, u_k) = 0, \; \text{cov}(e_{i,j}, u_j) = 0$$

Here $y_{i,j}$ is the value of the outcome variable (i.e., amyloid metric) for the i^{th} subject, attending their j^{th} PET acquisition (j = 1 to 3). $\beta_{0,i} = \beta_0 + u_{0,i}$ and $\beta_{1,i} = \beta_1 + u_{1,i}$ with β_0 and β_1 the overall fixed effects for the intercept and the slope, and $u_{0,i}$ and $u_{1,i}$ the random effects for the intercept and the slope at the group level such that $\begin{pmatrix} u_{0,i} \\ u_{1,i} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}\right)$. $e_{i,j}$ is the random effect at the individual level and $t_{i,j}$ the time of the j^{th} PET acquisition relative to the time of the baseline acquisition (in years). The corresponding R syntax is : 'Metric ~ dt + (1 + dt|subject)'.

In some cases, this model would not converge due to the inability of estimating the random slope effect in the amyloid negative population. In these instances, a random intercept only model was used for the amyloid negative group. The previous equation can be simplified with $\beta_{1,i} = \beta_1$, $u_{0,i} \sim N(0, \sigma_0^2)$ and the R syntax is then: 'Metric ~ dt + (1 |subject)'.

Additionally, coefficients of variation for the ARC were computed (standard deviation/mean), with the sampling uncertainty assessed using bootstrap resampling. Bias-corrected and accelerated confidence intervals were obtained using 5,000 bootstrap replicates.

Sample size estimates

Finally, the ARC and standard deviations were used to compute the samples size estimates ($\alpha = 0.05$; $1-\beta = 80\%$) required to detect 25% decrease in annualised amyloid accumulation. The sample size estimates represent the required sample size per arm, in the context of a trial with one measure of change and equal number of participants in each arm. As in Lopes Alves et al. [3], we designed two clinical trial scenario, based on CL cut-off values. The first is a secondary prevention trial with individuals in early amyloid accumulation or preclinical AD ($20 \leq CL \leq 50$ at baseline) [7]. The second is a secondary prevention trial with individuals with at least intermediate amyloid load ($CL > 20$ at baseline). Non-parametric, bias-corrected, and accelerated confidence intervals were estimated using 5,000 bootstrap replicates.

RESULTS

Repeatability and Reliability

A summary of our results on the repeatability and reliability of each metric can be found in **Table 3**.

	^{18}F-florbetapir (Insight46 cohort)						^{18}F-flutemetamol (independent cohort)				
	DVR (BP_{ND} + 1)	**CL**	**Aβ load**	**Aβ index**	**CL_{NMF}**	**AMPSS**	**CL**	**Aβ load**	**Aβ index**	**CL_{NMF}**	**AMPSS**
Sample Mean	1.06	-6.74	0.17	-0.53	-2.39	0.25	97.96	0.94	0.21	106.32	0.84
Sample SD	0.03	7.95	0.05	0.09	8.99	0.11	21.55	0.18	0.14	22.39	0.06
RMS between subjects	0.05	11.25	0.08	0.13	12.72	0.16	30.48	0.26	0.20	31.66	0.09
RMS within subjects	0.02	3.69	0.01	0.04	2.78	0.04	2.63	0.02	0.01	2.10	0.01
ICC	0.81 [0.36, 0.94]	0.89 [0.69, 0.96]	0.97 [0.91, 0.99]	0.92 [0.76, 0.97]	0.95 [0.87, 0.98]	0.93 [0.80, 0.97]	0.99 [0.97, 1.00]	0.99 [0.96, 1.00]	1.00 [0.98, 1.00]	1.00 [0.98, 1.00]	0.99 [0.95, 1.00]

Table 3 Repeatability and reliability of amyloid metrics evaluated on test-retest [^{18}F]florbetapir data from a subset of Insight46 scans assumed to be stable over time (*N*=16, except for DVR: *N*=13), and on test-retest [^{18}F]flutemetamol data from independent cohorts (*N*=10). Selection criteria for Insight46 stable group: (1) baseline and follow-up PET available, (2) CSF Aβ42/40 and ptau181 available at follow-up, (3) at follow-up, CSF Aβ42/40 value in the top quartile, (4) at follow-up, CSF ptau181 <=57 pg/ml, (5) no mild cognitive impairment or major brain disorder at baseline

ICC and 95% confidence intervals (in brackets) estimates were calculated based on a mean-rating (k = 2), absolute-agreement and a 2-way mixed-effects model.

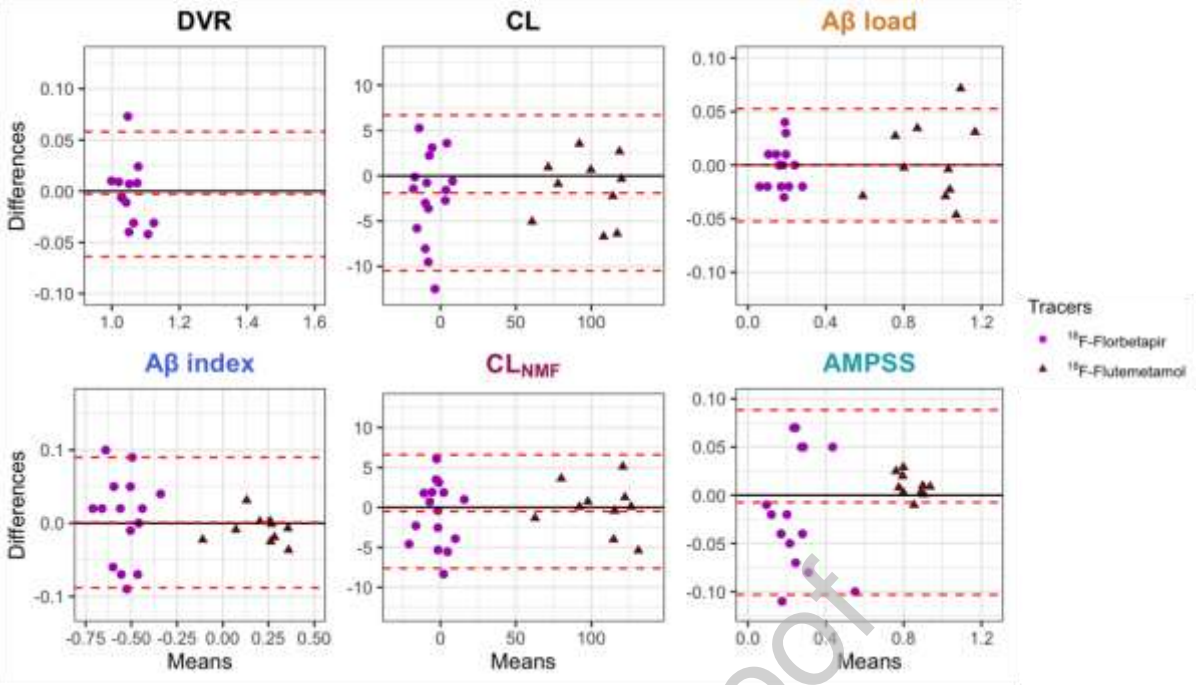


Figure 2 Bland-Altman plots indicating bias between test and retest measurements. Dashed lines indicate the mean, lower and upper limit of agreement (+/- 1.96 standard deviation from the mean)

The [^{18}F]flutemetamol test-retest data represents short-interval acquisitions in 10 subjects with an AD diagnosis, while the [^{18}F]florbetapir data was a much longer scan interval in a subset of individuals that represented a stable group of 16 healthy controls. Test and retest scans acquired with [^{18}F]flutemetamol show higher agreement with each other than the ones in [^{18}F]florbetapir subset (**Figure 2**).

Repeatability can only be compared for the CL and CL_{NMF} as they are on a similar scale. Both in the [^{18}F]flutemetamol and [^{18}F]florbetapir data, the within-subject RMS of these two metrics is within 1 unit. Furthermore, we can assess the reliability of all metrics against each other as the ICC is not dependant on the range of the metric. For the [^{18}F]flutemetamol test-retest scans, we observed excellent reliability of all the metrics (ICC ≥ 0.98), which indicates that the variability observed between the two scans is mostly attributable to an underlying difference in protein deposition across individuals rather than measurement error. The overlapping confidence intervals suggest that there is no evidence of difference in this across methods. For [^{18}F]florbetapir, our results suggest

that all metrics have acceptable reliability. We verified using an *F* test that the high ICC values could not be explained by high between-subject variability.

Correlation with BP_{ND}

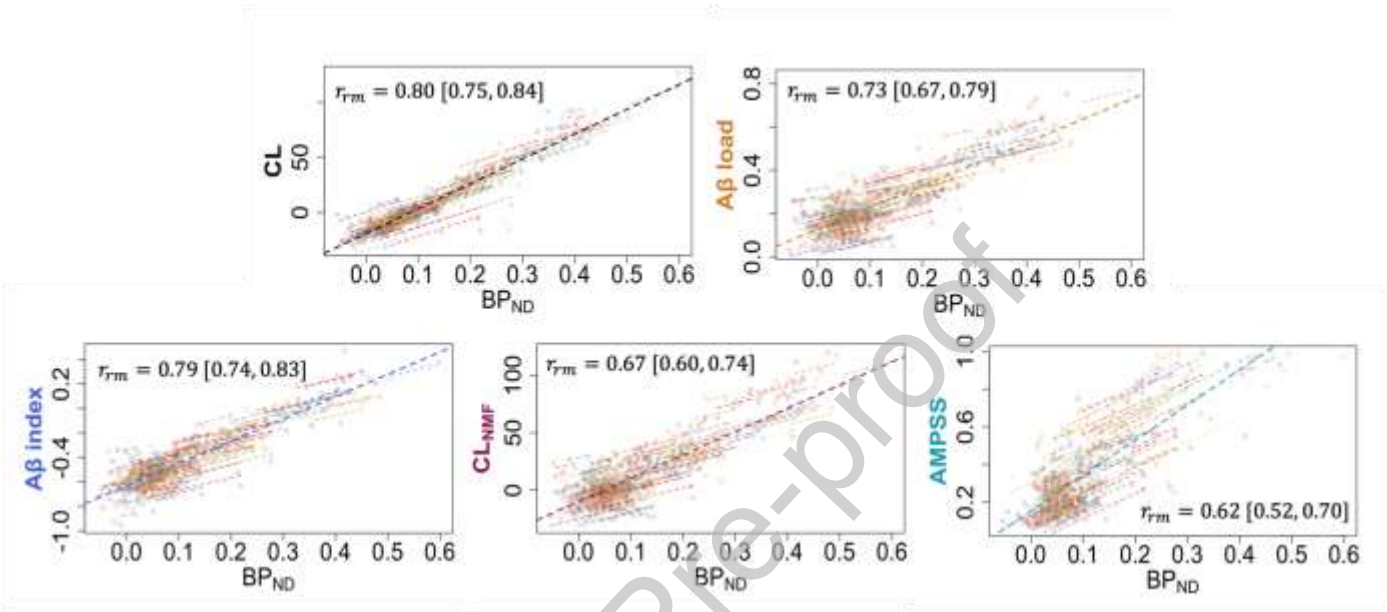


Figure 3 Correlation between amyloid metrics and non-displaceable binding potential BP_{ND}, evaluated with dynamic acquisition scans from Insight46 data. The 95% confidence interval for r_{rm} was built using 2,000 bootstrap replicates. The dotted lines represent the regression for the metrics averaged per subject. Abbreviation - r_{rm}: repeated measurement correlation coefficient

Repeated measure correlations between BP_{ND} and the other metrics were computed for the Insight46 dataset (only dataset with dynamic acquisitions) and are shown in **Figure 3**. There is a positive association between the longitudinal change in all metrics and BP_{ND}, with highest r_{rm} values for the CL, Aβ load and Aβ index. A correlation of the novel metrics and the SUVr can be found in Supplementary material.

Correlation with CL

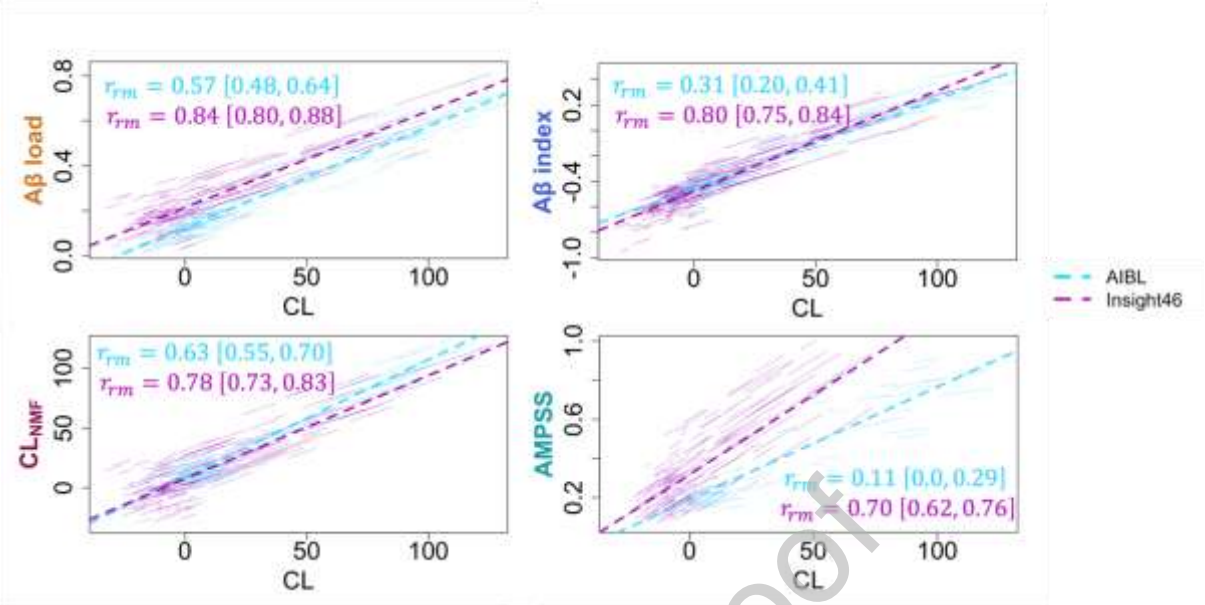


Figure 4 Correlation between amyloid metrics and the Centiloid scale (CL). The dotted lines represent the regression for the metrics averaged per subject. The 95% confidence interval for r_{rm} was built using 2,000 bootstrap replicates.

Comparison of the data driven metrics with the CL scale over both datasets is shown in **Figure 4**. The CL_{NMF}, Aβ load and Aβ index show close regression slopes for both AIBL and Insight46. The r_{rm} is highest for these three metrics indicating a high association with longitudinal change in CL especially in Insight46, although a lower r_{rm} for the Aβ index in AIBL. All data-driven metrics systematically show higher correlation with CL in Insight46 compared to AIBL.

Annual rate of change & Coefficients of variation

	Insight46					AIBL				
	ARC				CV	ARC				CV
	All	CL ≤ 15	20 ≤ CL ≤ 50	CL > 20	All	All	CL ≤ 15	20 ≤ CL ≤ 50	CL > 20	All
N	438	331	39	67	438	185	100	11	52	185
BP_{ND}	0.01 ± 0.02	0.01 ± 0.02	0.03 ± 0.02	0.03 ± 0.02	1.6 [1.3, 2.2]	N/A				
CL	2.4 ± 3.5	1.5 ± 2.8	6.1 ± 3.8	5.5 ± 3.5	1.5 [1.3, 1.7]	0.5 ± 13.8	-0.1 ± 12.5	3.0 ± 13.4	2.4 ± 15.7	26.8 [8.7, >100]
Aβ load	0.01 ± 0.02	0.01 ± 0.01	0.03 ± 0.02	0.03 ± 0.01	1.7 [1.4, 2.0]	0.01 ± 0.04	0.01 ± 0.03	0.02 ± 0.02	0.02 ± 0.05	2.9 [2.3, 3.7]
Aβ index	0.01 ± 0.03	0.01 ± 0.03	0.04 ± 0.03	0.04 ± 0.03	2.2 [1.8, 3.2]	0.02 ± 0.08	0.01 ± 0.04	0.04 ± 0.14	0.03 ± 0.09	3.6 [2.7, 6.3]
CL_{NMF}	2.1 ± 3.0	1.2 ± 2.3	5.9 ± 3.1	5.2 ± 2.8	1.5 [1.3, 1.8]	2.4 ± 5.6	1.6 ± 5.6	3.6 ± 4.2	3.5 ± 5.5	2.4 [2.0, 3.0]
AMPSS	0.02 ± 0.04	0.02 ± 0.04	0.04 ± 0.05	0.04 ± 0.05	1.9 [1.6, 2.4]	0.01 ± 0.11	0.01 ± 0.11	0.03 ± 0.07	0.02 ± 0.11	12.9 [4.6, >100]

Table 4 Annualised rates of change in amyloid deposition and coefficients of variation in AIBL and Insight46 datasets. Values are described as mean ± standard deviation. Bias corrected and accelerated confidence intervals for the coefficients of variation were built via bootstrap resampling using 5,000 replicates. Abbreviations – ARC: annualised rates of change, CV: coefficients of variation

The annualised amyloid accumulation rates were computed for all metrics in Insight46 and AIBL, for three subsets of the datasets defined using the baseline CL value of each subject (**Table 4**). The ability to compare these ARC in between metrics is restricted by the various ranges of these methodologies as well as the different cognitive profiles of the subjects in the Insight46 and AIBL studies. Nonetheless, we can observe that the ARC variability is from two and up to eight times higher in the AIBL dataset compared to the Insight46 dataset. In Insight46, the coefficients of variation of the ARC for the CL_{NMF} and the Aβ load were in line with the ones from the conventional metrics while in AIBL, these two metrics had a lower CV compared to the CL. The mean CL ARC was found to be negative in AIBL in the subset of scans with CL≤15).

Sample size estimates

	Insight46	
	20 ≤ CL ≤ 50	CL > 20
N	39	67
BP_{ND}	93 [39, 340]	105 [52, 284]
CL	96 [56, 183]	117 [69, 232]
Aβ load	81 [52, 139]	111 [69,208]
Aβ index	131 [69, 288]	147 [84,348]
CL_{NMF}	71 [45, 113]	74 [52, 109]
AMPSS	558 [220, 3123]	553 [217, 3069]

Table 5 Samples size estimates per arm (α = 0.05; 1-β = 80%) required to detect a 25% decrease in annualized amyloid accumulation. This assumes two arms, one treatment and one placebo with equal size. Two scenarios were assessed: a secondary prevention trial focusing on early accumulators (20<CL≤ 50), and a secondary prevention trial for individuals with at least moderate amyloid burden (CL≥20). Bias-corrected and accelerated confidence intervals were obtained using 5,000

The sample size estimates to detect a 25% reduction in accumulation rates are summarised in **Table 5**. For secondary prevention trials scenario, targeting early accumulators and subjects with at least moderate amyloid burden, the CL_{NMF} consistently displayed lower sample size requirements in comparison to all other metrics, though the confidence intervals do overlap between measures. The Aβ load performed in line with the BP_{ND} and the CL approaches.

DISCUSSION

In this study, we evaluated four novel data-driven metrics of amyloid deposition on three datasets, acquired with [^{18}F]florbetapir and [^{18}F]-flutemetamol, and compared them the more established model-based values including CL and BP_{ND}.

We first evaluated the repeatability and reliability of the data-driven methodologies. The variability of repeated measurements should be considered together with the mean, standard deviation, and range of each metric. Indeed, a direct comparison between these metrics is not always possible because of the different scales and ranges of these novel measures. No difference in ICC between metrics was observable in the test-retest [^{18}F]flutemetamol and [^{18}F]florbetapir data, except for the DVR which had a slightly lower ICC and greater confidence interval than all other metrics. This might be explained by small between-subject variability with the DVR (rather than high within-subject variability), as a result of DVR values across amyloid negative individuals being very similar (around 1.0-1.1). This results in the within-subject variation being almost as high as the between subject variation, which in turn decreases the ICC.

Repeated measure correlations between the amyloid metrics and the BP_{ND} (DVR - 1) suggest that the longitudinal trajectories of the CL, Aβ index, Aβ load and CL_{NMF} are close to those of the BP_{ND}.

When examining the association between these data-driven metrics with the well-established CL scale, we found strong cross-sectional relationships between the CL_{NMF}, Aβ load and Aβ index across both of our main datasets. AMPSS, on the other hand, showed a more divergent relationship between the two datasets, suggesting that it could be dependent on the training set used. In our case, we withheld part of the original dataset to form the training data, hence differences in the original dataset might be reflected on the final AMPSS. While there are similar relationships between datasets, all metrics show a noticeably higher association to longitudinal/intra-individual change in CL in the Insight46 dataset compared to AIBL. This could be explained by the definition of the r_{rm}

coefficient as it may be unstable in cases where the expected slope across subjects varies significantly, and in cases where the number of observations per subject differs. These two factors would be more prominent in AIBL compared to Insight46, which is mostly comprised of cognitively unimpaired subjects who underwent two PET scans. Additionally, differences in the type of scanner used between these datasets could result in slightly different observed uptake in the reference region. The Insight 46 uses a single Siemens Biograph combined PET-MR scanner, which tends to be more modern than many of the PET-CT scanners used in AIBL, as well as having a larger axial field of view. Another explanation could be that the Insight 46 cohort comprises mostly healthy controls while the linear regression for AIBL has more individuals, given the inclusion of MCI and AD patients, having higher amyloid values.

Finally, our results show than the BP_{ND}, CL_{NMF} and Aβ load are the metrics that are the most sensitive to changes in amyloid accumulation rate, leading to lower coefficients of variation and lower sample size estimates required to power hypothetical secondary prevention trials. We chose not to include the sample size estimates for AIBL (where the sample sizes were also lowest for the CL_{NMF} and Aβ load) because there was no evidence for many of the metrics, data-driven or established, of the ARC being significantly different from zero, which appears to be driven more by a high level of longitudinal variability. This variability might stem from differences in uptake values in the cerebellum at each timepoint, that are dependent on the patient positioning within the scanner field of view. Some approaches rely on an accurate definition and registration of the cerebellum or cerebellar grey matter to build their image component templates from which the data-driven metrics are derived. Under or over-estimation of the amyloid burden in the cerebellum might therefore increase the longitudinal variability across all metrics. Overall, the CL_{NMF} and Aβ load seem to offer a more precise alternative to the established CL, though further studies on larger cohorts that also include [^{18}F]florbetaben are required for confirmation.

Beyond quantitative performance criteria, several of the data driven metrics provide other potential benefits that should be taken into consideration. As mentioned previously, the AMPSS and the Aβ index do not require a pre-defined reference region which is regularly subject of debate and a well-documented source of uncertainty. For the AMPSS, as the original training consisted of F-18 florbetapir PET data from ADNI, we first attempted to use this training data for the F-18 flutametamol AIBL dataset as well, but the poor performance of the SVM suggests that the voxel-wise multivariate pattern captured by the training is tracer-specific (data not shown). For unseen data, the optimisation of a well-calibrated unbiased, cross-tracer training set could potentially increase the accuracy and precision of this metric. There may be instances, such as individuals with contraindications to MR, where PET-only methods are desirable. The Aβ index and an updated version of the CL_{NMF} do not require an MR, and can be alternatives to PET-only pipelines that were developed for the SUVr and CL [54,55]. The Aβ load can also be computed without MR, provided that the NS and K templates are available and that SUVr maps are obtained with a PET-only pipeline. Moreover, the availability of the software is important to consider towards the aims of reproducibility and open science. If the software to perform the data-driven metric includes all pre-processing steps, or alternatively documentation that is suitably comprehensive such that these steps can be accurately reproduced in other user's environments, then the variability caused by using various pipelines (and in particular, the registration software) could be limited. Indeed, we found that the Aβ index was the metric which performs most similarly across both our datasets in terms of correlation with CL (cross-sectional data only). Importantly, the $CL_{NMF,}$ like the CL, is tracer independent and was shown to be more robust than CL to change in tracer in the original paper [22]. The improved precision of the CL_{NMF} could further encourage the use of Centiloid and standardised scales by establishing a harmonised outcome metric for clinical practice, and by being key to several stages of drug development [11]. Lastly, it is worth noting that the Aβ index, the AMPSS and the CL_{NMF} are either open-source or available to researchers upon request.

These novel metrics, being data-driven, may help us gain insight into the spatiotemporal pattern of amyloid accumulation. The Aβ index, Aβ load and CL_{NMF} that isolate the specific binding pattern, and the AMPSS that dichotomises Aβ accumulation between HC and AD, overall show relatively close performance compared to the SUVr. This would suggest that for a large proportion of the population, there is a multivariate pattern of amyloid uptake rather than a regional spreading pattern. This supports a relatively generalisable global multivariate pattern of uptake, where some regions accumulate faster than others. This is line with the work of Whittington and Gunn [56] and Cho that showed that amyloid accumulation occurs simultaneously across many regions in the brain, usually by the time tau deposition in the entorhinal cortex emerges [26].

Data-driven metrics also present with some limitations. They are all surrogate measures of amyloid deposition and usually compress spatial information into a single summary metric. It is nonetheless worth noting that two methodologies can also help image amyloid patterns: the AMPSS which can output a feature weighted map from the SVM and the Aβ load which computation requires a SUVr fitted image from the NS and K maps. We could however take advantage of specific spatial patterns to target specific populations. For instance, subjects with cerebral amyloid angiopathy (CAA) have shown significant amyloid uptake in the occipital cortex [57]. A data-driven approach could potentially isolate a component specific to CAA, allowing the discrimination of two types of amyloid deposition. Finally, as these methods are data-driven and more susceptible to bias, demographic characteristics and technical specificities should be considered before implementing them on novel datasets.

There are potential limitations with regards to our implementation of each method. We chose to process the scans with little or no modification to the methods, including using the training and calibration data provided for each method where available. Although these methods are all data-driven, the training and calibration datasets differed in the number of scans required, amyloid burden distribution in the dataset and spatial and intensity normalization parameters. The use of

different training and calibration sets unique to each data driven metric is another potential source of variability. On the other hand, our approach avoids overfitting and allowed us to evaluate the methods how they will most likely be implemented by other researchers and clinicians. Furthermore, the training set for the Aβ load from which the NS and K maps are derived was made of [^{18}F]florbetapir scans and is therefore not best suited for the AIBL study acquired with [^{18}F]flutemetamol. Despite the data-driven metrics being less reliant on formal anatomical boundaries to define a specific reference or target region, most implementations of these metrics still involve some reference region definition as part of the process, and the choice of reference region is different between methods.

Apart from the four metrics evaluated in this study, other data-driven methods have been developed. The AMYQ method by Pegueroles et al. [58] is similar to the Aβ index in that it relies on a PCA approach and does not require the definition of a reference region. The idea of the adaptive template created via PCA in the Aβ index has also been achieved using MR independent deep learning approach by Kang et al. [59] Another Aβ load implemented by Tanaka et al. [60], which mainly differs from the Aβ load by Whittington and Gunn in the creation of K and NS templates, for which they use a PCA. Defining a metric without a formally defined reference region was also the goal of Chincanari et al., who developed ELBA, a metric of amyloid deposition relying on the intensity distribution patterns rather than specific ROIs. [61] Finally, Liu et al. proposed a method to estimate the specific amyloid burden (SAβ load) relying on the estimate of the non-specific binding using deep learning [62].

Future directions

The methods examined seem to be able to capture population-level, longitudinal, patterns of amyloid deposition. Further investigation assessing their capacity to accurately measure the change in protein burden at the subject level would be clinically relevant. To implement these metrics in a

longitudinal setting, their robustness to different tracers and scanners with different spatial resolution, as often is the case in multi-centre studies, should be investigated further. Future evaluation could also include [^{18}F]-florbetaben data which was not tested our study, as well as novel tau metrics which were beyond the scope of this study.

Overall, the recent drive to develop innovative, alternative methods to measure protein deposition via PET might benefit from a formal challenge in the future where these methods are evaluated against an unseen reference dataset that each group processes and provides results on.

CONCLUSION

We evaluated and validated four types of data-driven metrics of amyloid deposition on three separate datasets using a joint set of criteria against the more established CL and BP_{ND}. We found Aβ index, Aβ load and CL_{NMF} all have a close longitudinal association with the BP_{ND}, while the CL_{NMF} and Aβ load could offer a more precise alternative to the CL. Moreover, these novel data driven metrics can provide benefits such as being robust across tracers and pre-processing pipelines while offering a precision comparable or superior to the CL. Reducing the variability of PET longitudinal measurements may be achieved by using data-driven metrics of amyloid uptake.

REFERENCES

[1] Ortner M, Drost R, Heddderich D, Goldhardt O, Müller-Sarnowski F, Diehl-Schmid J, et al. Amyloid PET, FDG-PET or MRI? - the power of different imaging biomarkers to detect progression of early Alzheimer's disease. BMC Neurol 2019;19:264. https://doi.org/10.1186/s12883-019-1498-9.

[2] Hanseeuw BJ, Betensky RA, Jacobs HIL, Schultz AP, Sepulcre J, Becker JA, et al. Association of Amyloid and Tau With Cognition in Preclinical Alzheimer Disease: A Longitudinal Study. JAMA Neurol 2019;76:915–24. https://doi.org/10.1001/jamaneurol.2019.1424.

[3] Lopes Alves I, Heeman F, Collij LE, Salvadó G, Tolboom N, Vilor-Tejedor N, et al. Strategies to reduce sample sizes in Alzheimer's disease primary and secondary prevention trials using longitudinal amyloid PET imaging. Alzheimers Res Ther 2021;13:82. https://doi.org/10.1186/s13195-021-00819-2.

[4] Cummings J, Salloway S. Aducanumab: Appropriate use recommendations. Alzheimers Dement 2021:alz.12444. https://doi.org/10.1002/alz.12444.

[5] Budd Haeberlein S, Aisen PS, Barkhof F, Chalkias S, Chen T, Cohen S, et al. Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. J Prev Alzheimers Dis 2022. https://doi.org/10.14283/jpad.2022.30.

[6] Klein G, Delmar P, Voyle N, Rehal S, Hofmann C, Abi-Saab D, et al. Gantenerumab reduces amyloid-β plaques in patients with prodromal to moderate Alzheimer's disease: a PET substudy interim analysis. Alzheimers Res Ther 2019;11:101. https://doi.org/10.1186/s13195-019-0559-z.

[7] Aisen PS, Zhou J, Irizarry MC, Kramer LD, Swanson CJ, Dhadda S, et al. AHEAD 3-45 study design: A global study to evaluate the efficacy and safety of treatment with BAN2401 for 216 weeks in preclinical Alzheimer's disease with intermediate amyloid (A3 trial) and elevated amyloid (A45 trial). Alzheimers Dement 2020;16:e044511. https://doi.org/10.1002/alz.044511.

[8] Mintun MA, Lo AC, Duggan Evans C, Wessels AM, Ardayfio PA, Andersen SW, et al. Donanemab in Early Alzheimer's Disease. N Engl J Med 2021;384:1691–704. https://doi.org/10.1056/NEJMoa2100708.

[9] Rafii MS, Sperling RA, Donohue MC, Zhou J, Roberts C, Irizarry MC, et al. The AHEAD 3-45 Study: Design of a prevention trial for Alzheimer's disease. Alzheimers Dement 2022:alz.12748. https://doi.org/10.1002/alz.12748.

[10] Bucci M, Savitcheva I, Farrar G, Salvadó G, Collij L, Doré V, et al. A multisite analysis of the concordance between visual image interpretation and quantitative analysis of [18F]flutemetamol amyloid PET images. Eur J Nucl Med Mol Imaging 2021;48:2183–99. https://doi.org/10.1007/s00259-021-05311-5.

[11] Pemberton HG, Collij LE, Heeman F, Bollack A, Shekari M, Salvadó G, et al. Quantification of amyloid PET for future clinical use: a state-of-the-art review. Eur J Nucl Med Mol Imaging 2022. https://doi.org/10.1007/s00259-022-05784-y.

[12] Lopes Alves I, Collij LE, Altomare D, Frisoni GB, Saint-Aubert L, Payoux P, et al. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. Alzheimers Dement 2020;16:750–8. https://doi.org/10.1002/alz.12069.

[13] Klunk WE, Koeppe RA, Price JC, Benzinger T, Devous MD, Jagust W, et al. The Centiloid Project: Standardizing Quantitative Amyloid Plaque Estimation by PET. Alzheimers

Dement J Alzheimers Assoc 2015;11:1-15.e4. https://doi.org/10.1016/j.jalz.2014.07.003.

[14] Schmidt ME, Chiao P, Klein G, Matthews D, Thurfjell L, Cole PE, 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;11:1050–68. https://doi.org/10.1016/j.jalz.2014.09.004.

[15] Barthel H, Bullich S, Sabri O, Seibyl J, Villemagne V, Rowe C, et al. 18F-Florbetaben (FBB) PET SUVR quantification: Which reference region? J Nucl Med 2015;56:1563–1563.

[16] Bullich S, Villemagne VL, Catafau AM, Jovalekic A, Koglin N, Rowe CC, et al. Optimal Reference Region to Measure Longitudinal Amyloid-β Change with ^{18} F-Florbetaben PET. J Nucl Med 2017;58:1300–6. https://doi.org/10.2967/jnumed.116.187351.

[17] Cho SH, Choe YS, Park S, Kim YJ, Kim HJ, Jang H, et al. Appropriate reference region selection of 18F-florbetaben and 18F-flutemetamol beta-amyloid PET expressed in Centiloid. Sci Rep 2020;10:14950. https://doi.org/10.1038/s41598-020-70978-z.

[18] Heeman F, Hendriks J, Lopes Alves I, Ossenkoppele R, Tolboom N, van Berckel BNM, et al. [11C]PIB amyloid quantification: effect of reference region selection. EJNMMI Res 2020;10:123. https://doi.org/10.1186/s13550-020-00714-1.

[19] Ghisays V, Lopera F, Goradia DD, Protas HD, Malek-Ahmadi MH, Chen Y, et al. PET evidence of preclinical cerebellar amyloid plaque deposition in autosomal dominant Alzheimer's disease-causing Presenilin-1 E280A mutation carriers. NeuroImage Clin 2021;31:102749. https://doi.org/10.1016/j.nicl.2021.102749.

[20] Whittington A, Gunn RN. Amyloid Load: A More Sensitive Biomarker for Amyloid Imaging. J Nucl Med 2019;60:536–40. https://doi.org/10.2967/jnumed.118.210518.

[21] Leuzy A, Lilja J, Buckley CJ, Ossenkoppele R, Palmqvist S, Battle M, et al. Derivation and utility of an Aβ-PET pathology accumulation index to estimate Aβ load. Neurology 2020. https://doi.org/10.1212/WNL.0000000000011031.

[22] Bourgeat P, Doré V, Doecke J, Ames D, Masters CL, Rowe CC, et al. Non-negative matrix factorisation improves Centiloid robustness in longitudinal studies. NeuroImage 2021;226:117593. https://doi.org/10.1016/j.neuroimage.2020.117593.

[23] Prosser L, Veale T, Malone IB, Coath W, Fox NC, Cash DM. Amyloid Pattern Similarity Score (AMPSS): A reference region free measure of amyloid PET deposition in Alzheimer's disease: Neuroimaging / New imaging methods. Alzheimers Dement 2020;16. https://doi.org/10.1002/alz.042673.

[24] Rizzo G, Whittington A, Hesterman J, Gunn RN. AmyloidIQ: An advanced analytical algorithm to quantify amyloid-PET [18F]NAV4694 scans. Alzheimers Dement 2020;16:e043823. https://doi.org/10.1002/alz.043823.

[25] Whittington A, Gunn R. TauIQ - A canonical image based algorithm to quantify tau PET scans. J Nucl Med 2021:jnumed.120.258962. https://doi.org/10.2967/jnumed.120.258962.

[26] Cho H, Baek MS, Lee HS, Lee JH, Ryu YH, Lyoo CH. Principal components of tau positron emission tomography and longitudinal tau accumulation in Alzheimer's disease. Alzheimers Res Ther 2020;12. https://doi.org/10.1186/s13195-020-00685-4.

[27] Lilja J, Leuzy A, Chiotis K, Savitcheva I, Sörensen J, Nordberg A. Spatial Normalization of 18F-Flutemetamol PET Images Using an Adaptive Principal-Component Template. J Nucl Med 2019;60:285–91. https://doi.org/10.2967/jnumed.118.207811.

[28] Zammit MD, Tudorascu DL, Laymon CM, Hartley SL, Zaman SH, Ances BM, et al. PET measurement of longitudinal amyloid load identifies the earliest stages of amyloid-

beta accumulation during Alzheimer's disease progression in Down syndrome. NeuroImage 2021;228:117728. https://doi.org/10.1016/j.neuroimage.2021.117728.
[29] Zammit MD, Laymon CM, Betthauser TJ, Cody KA, Tudorascu DL, Minhas DS, et al. Amyloid accumulation in Down syndrome measured with amyloid load. Alzheimers Dement Diagn Assess Dis Monit 2020;12. https://doi.org/10.1002/dad2.12020.
[30] Haller S, Montandon M-L, Lilja J, Rodriguez C, Garibotto V, Herrmann FR, et al. PET amyloid in normal aging: direct comparison of visual and automatic processing methods. Sci Rep 2020;10:16665. https://doi.org/10.1038/s41598-020-73673-1.
[31] Bourgeat P, Doré V, Fripp J, Ames D, Masters CL, Salvado O, et al. Implementing the centiloid transformation for 11C-PiB and β-amyloid 18F-PET tracers using CapAIBL. NeuroImage 2018;183:387–93. https://doi.org/10.1016/j.neuroimage.2018.08.044.
[32] Bourgeat P, Doré V, Burnham SC, Benzinger T, Tosun D, Li S, et al. β-amyloid PET harmonisation across longitudinal studies: Application to AIBL, ADNI and OASIS3. NeuroImage 2022;262:119527. https://doi.org/10.1016/j.neuroimage.2022.119527.
[33] Lane CA, Parker TD, Cash DM, Macpherson K, Donnachie E, Murray-Smith H, et al. Study protocol: Insight 46 – a neuroscience sub-study of the MRC National Survey of Health and Development. BMC Neurol 2017;17:75. https://doi.org/10.1186/s12883-017-0846-x.
[34] Markiewicz PJ, Ehrhardt MJ, Erlandsson K, Noonan PJ, Barnes A, Schott JM, et al. NiftyPET: a High-throughput Software Platform for High Quantitative Accuracy and Precision PET Imaging and Analysis. Neuroinformatics 2018;16:95–115. https://doi.org/10.1007/s12021-017-9352-y.
[35] Burgos N, Cardoso MJ, Thielemans K, Modat M, Pedemonte S, Dickson J, et al. Attenuation Correction Synthesis for Hybrid PET-MR Scanners: Application to Brain Studies. IEEE Trans Med Imaging 2014;33:2332–41. https://doi.org/10.1109/TMI.2014.2340135.
[36] Ladefoged CN, Law I, Anazodo U, St. Lawrence K, Izquierdo-Garcia D, Catana C, et al. A multi-centre evaluation of eleven clinically feasible brain PET/MRI attenuation correction techniques using a large cohort of patients. NeuroImage 2017;147:346–59. https://doi.org/10.1016/j.neuroimage.2016.12.010.
[37] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 2009;21:672–87. https://doi.org/10.1017/S1041610209009405.
[38] Landau SM, Fero A, Baker SL, Koeppe R, Mintun M, Chen K, et al. Measurement of Longitudinal -Amyloid Change with 18F-Florbetapir PET and Standardized Uptake Value Ratios. J Nucl Med 2015;56:567–74. https://doi.org/10.2967/jnumed.114.148981.
[39] Keshavan A, Wellington H, Chen Z, Khatun A, Chapman M, Hart M, et al. Concordance of CSF measures of Alzheimer's pathology with amyloid PET status in a preclinical cohort: A comparison of Lumipulse and established immunoassays. Alzheimers Dement Diagn Assess Dis Monit 2021;13. https://doi.org/10.1002/dad2.12131.
[40] Lu K, Nicholas JM, Collins JD, James S-N, Parker TD, Lane CA, et al. Cognition at age 70: Life course predictors and associations with brain pathologies. Neurology 2019;93:e2144–56. https://doi.org/10.1212/WNL.0000000000008534.
[41] Vandenberghe R, Van Laere K, Ivanoiu A, Salmon E, Bastin C, Triau E, et al. 18F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: A

phase 2 trial: 18F-Flutemetamol Phase 2 Trial. Ann Neurol 2010;68:319–29. https://doi.org/10.1002/ana.22068.
[42] Miki T, Shimada H, Kim J-S, Yamamoto Y, Sugino M, Kowa H, et al. Brain uptake and safety of Flutemetamol F 18 injection in Japanese subjects with probable Alzheimer's disease, subjects with amnestic mild cognitive impairment and healthy volunteers. Ann Nucl Med 2017;31:260–72. https://doi.org/10.1007/s12149-017-1154-7.
[43] Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond Ser B 2001;356:1293–322. https://doi.org/10.1098/rstb.2001.0915.
[44] SPM12 n.d.
[45] Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage 2007;38:95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007.
[46] Cardoso MJ, Modat M, Wolz R, Melbourne A, Cash D, Rueckert D, et al. Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. IEEE Trans Med Imaging 2015;34:1976–88. https://doi.org/10.1109/TMI.2015.2418298.
[47] Battle MR, Pillay LC, Lowe VJ, Knopman D, Kemp B, Rowe CC, et al. Centiloid scaling for quantification of brain amyloid with [18F]flutemetamol using multiple processing methods. EJNMMI Res 2018;8:107. https://doi.org/10.1186/s13550-018-0456-7.
[48] Coath W, Modat M, Cardoso MJ, Markiewicz P, Lane CA, Parker TD, et al. Operationalising the Centiloid Scale for [^{18}F]florbetapir PET Studies on PET/MR. Radiology and Imaging; 2022. https://doi.org/10.1101/2022.02.11.22270590.
[49] Logan J, Fowler JS, Volkow ND, Wang G-J, Ding Y-S, Alexoff DL. Distribution Volume Ratios without Blood Sampling from Graphical Analysis of PET Data. J Cereb Blood Flow Metab 1996;16:834–40. https://doi.org/10.1097/00004647-199609000-00008.
[50] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021.
[51] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
[52] van Berckel BNM, Ossenkoppele R, Tolboom N, Yaqub M, Foster-Dingley JC, Windhorst AD, et al. Longitudinal Amyloid Imaging Using ^{11}C-PiB: Methodologic Considerations. J Nucl Med 2013;54:1570–6. https://doi.org/10.2967/jnumed.112.113654.
[53] Bakdash JZ, Marusich LR. Repeated Measures Correlation. Front Psychol 2017;8:456. https://doi.org/10.3389/fpsyg.2017.00456.
[54] Bourgeat P, Doré V, Fripp J, Villemagne V, Rowe C, Salvado O. Computational Analysis of PET by AIBL (CapAIBL): A cloud-based processing pipeline for the quantification of PET images. J Nucl Med 2015;56:149–149.
[55] Iaccarino L, La Joie R, Koeppe R, Siegel BA, Hillner BE, Gatsonis C, et al. rPOP: Robust PET-only processing of community acquired heterogeneous amyloid-PET data. NeuroImage 2022;246:118775. https://doi.org/10.1016/j.neuroimage.2021.118775.
[56] Whittington A, Sharp DJ, Gunn RN. Spatiotemporal Distribution of β-Amyloid in Alzheimer Disease Is the Result of Heterogeneous Regional Carrying Capacities. J Nucl Med 2018;59:822–7. https://doi.org/10.2967/jnumed.117.194720.
[57] Sperling RA, Jack CR, Black SE, Frosch MP, Greenberg SM, Hyman BT, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials:

Recommendations from the Alzheimer's Association Research Roundtable Workgroup. Alzheimers Dement 2011;7:367–85. https://doi.org/10.1016/j.jalz.2011.05.2351.
[58] Pegueroles J, Montal V, Bejanin A, Vilaplana E, Aranha M, Santos-Santos MA, et al. AMYQ: An index to standardize quantitative amyloid load across PET tracers. Alzheimers Dement 2021:alz.12317. https://doi.org/10.1002/alz.12317.
[59] Kang SK, Seo S, Shin SA, Byun MS, Lee DY, Kim YK, et al. Adaptive template generation for amyloid PET using a deep learning approach. Hum Brain Mapp 2018;39:3769–78. https://doi.org/10.1002/hbm.24210.
[60] Tanaka T, Stephenson MC, Nai Y-H, Khor D, Saridin FN, Hilal S, et al. Improved quantification of amyloid burden and associated biomarker cut-off points: results from the first amyloid Singaporean cohort with overlapping cerebrovascular disease. Eur J Nucl Med Mol Imaging 2020;47:319–31. https://doi.org/10.1007/s00259-019-04642-8.
[61] Chincarini A, Sensi F, Rei L, Bossert I, Morbelli S, Guerra UP, et al. Standardized Uptake Value Ratio-Independent Evaluation of Brain Amyloidosis. J Alzheimers Dis 2016;54:1437–57. https://doi.org/10.3233/JAD-160232.
[62] Liu H, Nai Y-H, Saridin F, Tanaka T, O' Doherty J, Hilal S, et al. Improved amyloid burden quantification with nonspecific estimates using deep learning. Eur J Nucl Med Mol Imaging 2021. https://doi.org/10.1007/s00259-020-05131-z.

Statements and Declarations

Funding

This work is supported by the EPSRC-funded UCL Centre for Doctoral Training in Intelligent, Integrated Imaging in Healthcare (i4health) (**EP/S021930/1**), the Department of Health's NIHR-funded Biomedical Research Centre at University College London, and GE Healthcare.

FB is supported by the NIHR biomedical research centre at UCLH.

Insight 46 funding

Insight 46 is funded by Alzheimer's Research UK, the Medical Research Council Dementia Platforms UK, Selfridges Group Foundation, Wolfson Foundation, Wellcome Trust, Brain Research UK, Alzheimer's Association. Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly, kindly provided the ^{18}F-florbetapir tracer (Amyvid^{TM}) free of cost but had no role in the design, conduct,

analysis or reporting of Insight 46 study findings. We are grateful to the Insight 46 study team and indebted to the study participants

AMYPAD funding

The project leading to this paper has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115952. This Joint Undertaking receives the support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This communication reflects the views of the authors and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein.

Data-code availability statement

The Centiloid calibration datasets can be downloaded at http://www.gaain.org/centiloid-project.
The AIBL data can be downloaded through LONI after registration at http://adni.loni.usc.edu/category/aibl-study-data/.
The Insight46 data is a sub-study of the National Survey of Health and Development (NSHD). The data sharing statement for NSHD can be found at https://nshd.mrc.ac.uk/data-sharing/.
The Aβ index can be made available to researchers.
The python code used to build the NMF models as well as the NMF models can be downloaded from https://doi.org/10.25919/5f8400a0b6a1e.

CRediT authorship contribution statement

Ariane Bollack: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - original draft.
Pawel J Markiewicz: Supervision, Conceptualization, Methodology, Writing - review & editing
Alle Meije Wink: Software, Validation, Writing - review & editing
Lloyd Prosser: Software, Validation, Writing - review & editing
Johan Lilja: Software, Validation, Writing - review & editing
Pierrick Bourgeat: Software, Validation, Writing - review & editing
Jonathan M Schott: Funding, Resources, Writing - review & editing
William Coath: Data curation, Methodology, Writing - review & editing
Lyduine E Collij: Methodology, Writing - review & editing
Hugh G Pemberton: Methodology, Writing - review & editing
Gill Farrar: Supervision, Funding, Resources, Writing - review & editing
Frederik Barkhof: Supervision, Funding, Conceptualization, Methodology, Writing - review & editing

David M Cash: Supervision, Conceptualization, Data curation, Methodology, Formal analysis, Writing - review & editing

Declaration of interest statement

Ariane Bollack, **Pawel J Markiewicz, Alle Meije Wink**, **Lloyd Prosser**, **William Coath, and David M Cash** have nothing to disclose.
Johan Lilja is an employee of Hermes Medical Solutions.
Pierrick Bourgeat is an employee of CSIRO Health and Biosecurity.
Jonathan M Schott has received research funding and PET tracer from AVID Radiopharmaceuticals (a wholly owned subsidiary of Eli Lilly) and Alliance Medical; has consulted for Roche, Eli Lilly, Biogen, AVID, Merck and GE; and received royalties from Oxford University Press and Henry Stewart Talks. He is Chief Medical Officer for Alzheimer's Research UK, and Medical Advisor to UK Dementia Research Institute.
Lyduine E Collij has received research support from GE HealthCare.
Hugh G Pemberton is a former employee of GE HealthCare.
Gill Farrar is an employee of GE HealthCare.
Frederik Barkhof is a steering committee and iDMC member of studies by Biogen, Merck, Roche, and EISAI. He is a consultant to Roche, Biogen, Merck, IXICO, Jansen, and Combinostics. He has research agreements with Novartis, Merck, Biogen, GE, and Roche and is co-founder of Queen Square Analytics Ltd. His research is sponsored by the NIHR-UCLH Biomedical Research Centre, UK MS Society, MAGNIMS-ECTRIMS, EC-H2020, EC-JU (IMI), and EPSRC.