European Journal of Nuclear Medicine and Molecular Imaging
https://doi.org/10.1007/s00259-020-04689-y

ORIGINAL ARTICLE

A kinetics-based approach to amyloid PET semi-quantification

A. Chincarini[1] · E. Peira[1,2] · M. Corosu[1] · S. Morbelli[3,4] · M. Bauckneht[3,4] · S. Capitanio[3,4] · M. Pardini[2,4] · D. Arnaldi[2,4] · C. Vellani[5] · D. D'Ambrosio[6] · V. Garibotto[7,8] · F. Assal[7,9] · B. Paghera[10] · G. Savelli[11] · A. Stefanelli[11] · U. P. Guerra[11] · F. Nobili[2,4]

Received: 30 July 2019 / Accepted: 7 January 2020

Abstract

Purpose To develop and validate a semi-quantification method (time-delayed ratio, TDr) applied to amyloid PET scans, based on tracer kinetics information.

Methods The TDr method requires two static scans per subject: one early (~ 0–10 min after the injection) and one late (typically 50–70 min or 90–100 min after the injection, depending on the tracer). High perfusion regions are delineated on the early scan and applied onto the late scan. A SUVr-like ratio is calculated between the average intensities in the high perfusion regions and the late scan hotspot. TDr was applied to a naturalistic multicenter dataset of 143 subjects acquired with [^{18}F]florbetapir. TDr values are compared to visual evaluation, cortical–cerebellar SUVr, and to the geometrical semi-quantification method ELBA. All three methods are gauged versus the heterogeneity of the dataset.

Results TDr shows excellent agreement with respect to the binary visual assessment (AUC = 0.99) and significantly correlates with both validated semi-quantification methods, reaching a Pearson correlation coefficient of 0.86 with respect to ELBA.

Conclusions TDr is an alternative approach to previously validated ones (SUVr and ELBA). It requires minimal image processing; it is independent on predefined regions of interest and does not require MR registration. Besides, it takes advantage on the availability of early scans which are becoming common practice while imposing a negligible added patient discomfort.

Keywords Amyloid PET · Semi-quantification · Tracer kinetics · Fluorinated tracer

This article is part of the Topical Collection on Technology.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00259-020-04689-y) contains supplementary material, which is available to authorized users.

✉ A. Chincarini
andrea.chincarini@ge.infn.it

[1] Istituto Nazionale di Fisica Nucleare (INFN), via Dodecaneso 33, I-16146 Genova, Italy

[2] Dept. of Neuroscience, Rehabilitation ,Ophthalmology, Genetics, Child and Maternal Health (DINOGMI), University of Genoa, Genova, Italy

[3] Nuclear Medicine Unit (DISSAL), University of Genoa, Genova, Italy

[4] Neurology Clinic, IRCCS Ospedale Policlinico San Martino, Genova, Italy

[5] Nuclear Medicine Unit of Pavia Institute, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy

[6] Medical Physics Unit of Pavia Institute, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy

[7] Faculty of Medicine, University of Geneva, Geneva, Switzerland

[8] Division of Nuclear Medicine and Molecular Imaging, University Hospitals of Geneva, Geneva, Switzerland

[9] Memory Clinic, University Hospitals of Geneva, Geneva, Switzerland

[10] Nuclear Medicine Unit, ASST Spedali Civili, Brescia, Italy

[11] Nuclear Medicine Unit, Fondazione Poliambulanza, Brescia, Italy

Published online: 25 January 2020

Introduction

The recent NIA-AA diagnostic framework [1] considers the direct (i.e., through PET imaging) or indirect (i.e., through cerebrospinal fluid, CSF, assay) evidence of abnormal brain beta-amyloid deposition as the characteristic signature to define the Alzheimer's spectrum in vivo. This position is shared nowadays by most researchers and clinicians and had been previously proposed by the International Working Group-II research criteria in 2014 [2].

In the past few years, thanks to a number of commercial fluorinated radiotracers, amyloid PET has indeed emerged as a validated and effective proxy of brain amyloidosis [3–6], becoming a major tool in the diagnostic process in order to confirm or, even more robustly, rule out Alzheimer's disease (AD) [1, 2], and for improving patient selection for clinical trials [7].

Currently, amyloid PET assessment consists of a visual evaluation by trained nuclear medicine physicians complemented by the (still optional to date) support of a semi-quantification technique. The outcome is usually a binary evaluation (negative/positive) of the scan, at least in the clinical setting.

In the set of semi-quantification techniques, the standardized uptake value ratio (SUVr) is the most widely used [8]. SUVr values represent the ratio of PET counts between one or more target regions of interest (ROI) versus a reference region.

SUVr-based measures of amyloid tracers uptake show good agreement both with histopathological measures of the density of neuritic plaques [4, 9] and with the binary classification obtained by CSF analysis [10] and visual assessment [9].

However, there are several possible choices both in drawing target ROIs and in identifying reference ROIs for the purpose of amyloid status evaluation, and especially on the latter an agreement is still lacking [11]. In addition, SUVr values can be impaired by several factors, such as number of ROI, ROI drawing criteria [12, 13], partial volume effect (PVE) [14, 15], and regional cerebral blood flow (rCBF) fluctuations (particularly for longitudinal studies) [15].

In recent years, such variabilities have paved the way to explore alternative, SUVr-independent approaches, such as ELBA [16]—a ROI-independent method designed to capture intensity distribution patterns rather than actual counts—and the more sophisticated combined method SLOPE [17] which corrects for atrophy, spillover, and blood flow using a PET/MRI acquisition and sophisticated post-processing steps.

Many of the SUVr nuisances are due to methodological issues, such as differences in ROI selection and placement [18]. In summary, the main issues that researchers are trying to overcome derive from (i) instability or inconsistency of cortical ROIs (MRI-guided versus atlas-guided, hand-drawn versus automatic); (ii) effect of rCBF changes between patients and intra-patient, possibly affecting tracer delivery to the cortex; (iii) PVE; and (iv) lack of a standard in the choice of a reference region.

As a possible solution to address these issues, we propose an approach based on a dual time-point acquisition, whose ROI are adapted to the individual patient anatomic and pathophysiological characteristics. This approach does not need any template or MRI registration, and may be suitable for longitudinal evaluations.

We take into account the surrogate rCBF information extracted by an *early* ("E") *PET static acquisition*: high rCBF regions are concentrated in the cortical gray matter (GM) and are used to delineate the uptake ROIs. The reference ROI, instead, is defined on the standard static acquisition—*late* ("L") *PET*.

The ratio between the average intensities in the uptake ROIs (from the early scan E) and the reference ROI (from the late scan L) is named "time-delayed ratio" (TDr). The name underlines the fact that the uptake ROI is defined at an early time point in the acquisition (E) but it is used at a later time point (it is mapped onto the late scan L).

TDr is therefore based on tracer kinetics, and the ROI are uniquely adapted to each individual by using two static acquisitions at the opposite ends of the time frame for a lipophilic tracer.

Material and methods

PET scans and subject selection

A set of 143 subjects consisting of 107 subjects from naturalistic populations (aged 54–87) and 36 elderly normal aging (NA) subjects—undergoing a comprehensive evaluation in the frame of previous studies—were acquired in four clinical centers with [^{18}F]florbetapir PET.

The participating institutions were Nuclear Medicine Unit, Department of Health Sciences (DISSAL), University and IRCCS Ospedale Policlinico S. Martino, Genoa, Italy (HSM-GE, 29 subjects acquired with a Siemens BioGraph HiRes 1080); Department of Nuclear Medicine, Fondazione Poliambulanza, Brescia, Italy (FPA-BS, 55 subjects acquired with a Siemens BioGraph 40 mCT); University Hospitals and University of Geneva, Geneva, Switzerland (HUG-GE, 40 subjects acquired with a Siemens BioGraph 128 mCT); and Nuclear Medicine Unit, IRCCS Istituti Clinici Scientifici Maugeri, Pavia, Italy (ICSM-PV, 19 subjects acquired with GE Discovery 690 VCT). Demographics and final diagnostic hypotheses are provided in Tables 1 and 2. A further refinement taking visual assessment into account is shown in the supplementary materials (table 7).

Symptomatic outpatients underwent amyloid PET because of a clinical suspicion of AD which could not be confirmed (or ruled out) on the basis of standard clinical-neuropsychological assessment and morphological neuroimaging (usually MRI); a number of patients also had previously undergone brain F-18

Table 1 Demographics

Center	# subjects	Gender M/F	Age (95% CI)	Visual assessment, neg/pos
FPA-BS	55	24/31	71 (69, 72)	29/26
HUG-GE	40	19/21	72 (69, 74)	24/16
ICSM-PV	19	9/10	76 (73, 79)	7/12
HSM-GE	29	15/14	75 (72, 78)	10/19

fluorodeoxyglucose PET. Patients were enrolled according to the local clinical practice and in keeping with the international appropriate use criteria for amyloid PET [19]. No other selection criteria was set.

Injected dose for HSM-GE, FPA-BS, and ICSM-PV had an average of ~442 MBq (min 370 MBq, max 536 MBq), in compliance to the minimum dosage suggested by the tracer manufacturer. Due to local regulations, HUG-GE injected dose ranged between 180 and 220 MBq with an average of ~200 MBq.

Acquisition protocol was as follows: the patient was positioned into the scanner, injected with the tracer and acquired for 5–7 min (*E*). Then, he/she was let out of the scanner and left resting for ~40 min (i.e., 50 min after the injection). Lastly, the patient was positioned again into the scanner for the L static acquisition (10 min) according to the tracer manufacturer indications. *E* and *L* were therefore acquired with patient repositioning and hence each scan needed its own CT for attenuation correction.

Kinetics

The TDr is a semi-quantification method that is inspired by the absolute quantification approach.

Initially, the tracer flows into the blood vessels and then migrates to the tissues. Thus, in the first 5–8 min, the signal is dominated by the uptake in brain structures as a function of rCBF, so that *E* can be taken as a proxy of brain perfusion [20–23]. Then, from *E*, it is possible to select highly perfused tissues, and this selection is restricted to the broad cortical areas.

Several studies compared rCBF in both gray matter (GM) and white matter (WM): MRI-based arterial spin labeling (ASL) studies reported that WM blood flow is 1.6 to 4.6 times lower then GM flow [24–27], while PET perfusion studies showed 0.15 to 0.18 ml/g/min for WM [28–30] and 1.05 ml/g/min for GM [28] resulting in a GM/WM ratio of 5.8/1 to 7.0/1.

A high rCFB in a tissue ensures that the tracer is properly delivered. If Aβ is present in this tissue, the tracer will bind to it; otherwise, it will be quickly washed out (good clearance). Therefore, the selection of the relatively higher rCBF tissue in the brain—which largely coincides with the GM—is the best place to look for good contrast in the tracer uptake.

TDr computation

We denote by D_E the highest perfusion domain delineated on *E* and $D_L$is the highest uptake domain identified on *L*. These two domains are a collection of voxels defined on *E* and *L* as follows:

$$D_E = \left(v \in E \mid I_v \geq I_0^E\right) \tag{1}$$

$$D_L = \left(v \in L \mid I_v \geq I_0^L\right) \tag{2}$$

where *v* represents the voxels, I_v is the intensity (counts) in *v*. I_0^E and I_0^L are thresholds applied on *E* and *L* respectively (typical values are $I_0^E = 0.85$ percentile and $I_0^L = 0.99$ percentile on the respective intensity distributions). In words, D_E and D_L

Table 2 Post-test diagnosis

Center	NA	Diagnosis	Subtype	
		MCI/ Dem.	MCI/MCI-AD/AD dem.	FTD/VCI/CBS/ MSA
FPA-BS	9	12/34	12/0/19	6/6/2/1
HUG-GE	13	19/8	13/6/6	2/0/0/0
ICSM-PV	5	4/10	4/0/9	1/0/0/0
HSM-GE	9	6/14	4/2/14	0/0/0/0

NA, normal aging; *MCI*, mild cognitive impairment; *MCI-AD*, MCI due to Alzheimer's disease; *AD dem*, dementia due to Alzheimer's disease; *FTD*, frontotemporal dementia; *VCI*, vascular cognitive impairment; *CBS*, cortico-basal syndrome; *MSA*, multisystem atrophy; *Dem*, dementia

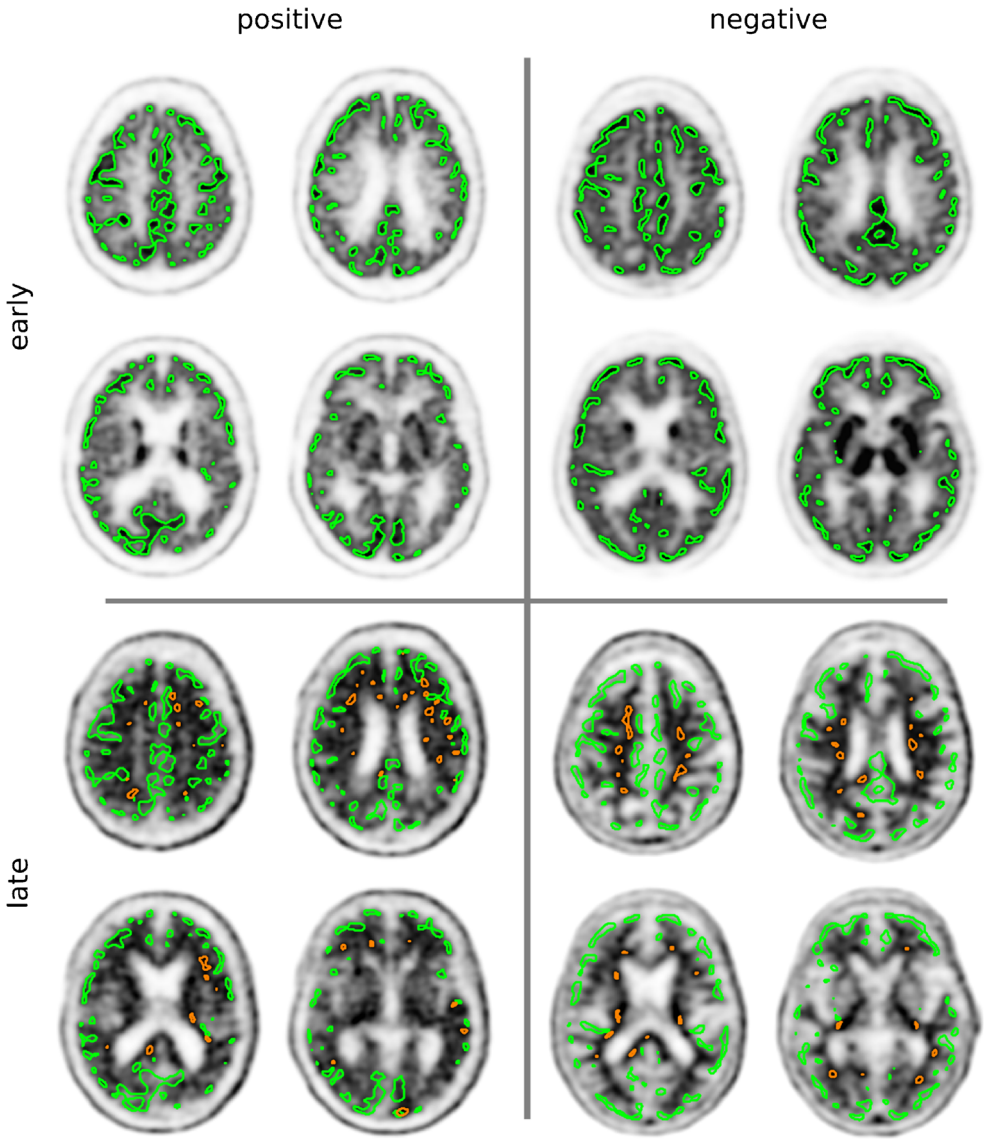

Fig. 1 The early (E) and the late (L) scans of a positive and negative subject. On the E scans the highest flow domain (D_E) are outlined in green. On the L scans, the highest uptake domains (D_L) are also outlined (in orange)

are collections of voxels whose intensities are greater than a specified value (see Fig. 1).

The TDr is then defined as follows:

$$\mathrm{TDr} = \frac{< I^L >|_{D_E}}{< I^L >|_{D_L}} \tag{3}$$

The numerator of the TDr formula is the mean intensity in L ($<I^L>$), averaged on the voxels with high perfusion defined in E (D_E, Eq. 1). Clearly, a high value of $I^L|_{D_E}$ corresponds to a high uptake and vice versa. As the absolute value of this intensity can vary among subjects for a number of reasons (injected activity, rCBF, physiological characteristics), this intensity must be normalized.

There are several choices for the reference ROIs according to the semi-quantification approaches that require normalization: the pons [31], the cerebellar GM, or the whole cerebellum [4, 12, 32, 33]. In recent years, concern has been raised about these reference regions: the small size of the pons or the peripheral position of the cerebellum in the field of view may introduce artifacts, and thus noise and variability in longitudinal measurements [34, 35]. These works, in agreement with other longitudinal studies [36–38], highlighted the potential advantages of a WM-based ROI as reference: reduction of the variability in longitudinal measurement, improvement of semi-quantification to track the increases of Aβ deposition, and stability of SUVr at single subject level.

Since the delineation of the most appropriate WM region has not yet been standardized, it was decided to use the highest uptake domain (D_L), otherwise known as the "hot spot,"

which is typically located in the WM (a specific uptake). The denominator of the TDr formula is therefore the mean intensity in L ($<I^L>$), averaged on the voxels with highest uptake (D_L).

Threshold definition

MRI-based measurements of the cerebral cortical GM showed its thickness ranging from 2.5 ± 0.7 mm up to 3.74 ± 0.32 mm [39–43]. As we look for amyloid presence in the cortex, it is reasonable to select an intensity threshold I_0^E that outlines a domain D_E whose thickness is comparable to the cortical GM thickness taking into account the spatial resolution and the point spread function. The threshold $I_0^E = 0.85$ is defined as a quantile on the E intensity distribution: it is practically independent on the single subject scan and it can be defined once and for all. To show this, we computed the threshold statistics on each patient finding the quantile corresponding to an average thickness $\mu = 3.0 \pm 0.16$ mm. It turns out that the quantiles cluster along the mean ($\mu = 0.847 \pm 0.036$). Moreover, small variations within these limits do not affect the TDr overall results.

Similarly, if D_L indicates the hot spot in the image, we choose the threshold $I_0^L = 0.99$ percentile on the L scan intensity distribution. This value was chosen so that D_L is a non negligible volume ($\mu = 13.47 \pm 1.2$ ml) and therefore less susceptible to intensity fluctuations.

Further information on the threshold computation are in Appendix A, supplementary materials.

Image processing

TDr implementation is rather straightforward. It only needs a spatial registration of E onto L, so that D_E voxels can be mapped onto the late scan. This is done with a 6-parameter linear transformation, with the L scan as fixed image.

TDr can be implemented as an automatic procedure which does not require any supervision (save an image registration check after mapping the early scan onto the late scan).

In order to compare TDr with other semi-quantification methods (SUVr and ELBA), we spatially normalized all scans to the MNI space as additional step. This procedure is identical to the one used in [16] and allowed us to take advantage of predefined MNI segmentations to provide the confidential volume C. The confidential volume was used to compute SUVr and ELBA values and therefore we now use it to constrain the domains D_E and D_L, so that TDr is compared to other semi-quantification methods using information coming from the same brain volume. The confidential volume C comes from the MNI lobe atlas and is defined as the whole brain parenchyma minus some specific regions: cerebellum, ventricles, brainstem, and basal ganglia (figure 4 in supplementary materials).

The average cortico-cerebellar SUVr was computed following the approach described in [12], whereas the ELBA score was computed according to [16].

Validation

TDr values were compared to (a) the binary visual assessment and (b) two validated semi-quantification methods (SUVr and ELBA).

The 143 PET scans were assessed by two expert readers (one nuclear medicine physician—MB—and one neurologist with certified reader training for [^{18}F]florbetapir and with 5-year experience in evaluation—FN) who reached consensus on the visual evaluation (i.e., negative or positive) on all scans. The discriminating power of the TDr was measured by the area under the receiver operating characteristic curve (AUC) for negative versus positive labeled scans.

To get an estimate of the generalized performances of the TDr versus visual assessment, we used a bootstrap procedure combined with a leave-10-out cross-validation. For each bootstrap step, 10 subjects were omitted (testing group), a cut-off value maximizing the accuracy was calculated with the remaining subjects (training group) and applied to the testing group. The whole process was repeated until all subjects were labeled: then, the accuracy, sensitivity, and specificity were calculated. The cut-off value was calculated as the mean of the cut-offs applied to the testing groups. This procedure was iterated 500 times.

TDr was compared to SUVr and ELBA, and performance measured by Pearson correlation. Finally, we looked at possible TDr-specific center effect with respect to SUVr and ELBA.

As last analysis, we looked at residuals in the linear models TDr-SUVr and TDr-ELBA in order to rule out non-linear relationships and to verify that TDr is indeed an independent measure of amyloid load. Residual analysis involved the Lilliefors normality test and the estimation of the Pearson correlation coefficient between the model residuals and the predictors.

Table 3 Performance (AUC) of TDr, SUVr, and ELBA versus visual assessment

Site	TDr	SUVr	ELBA
FPA-BS	1.00	0.99	1.00
HUG-GE	1.00	0.95	1.00
ICSM-PV	1.00	0.94	0.98
HSM-GE	0.99	0.92	0.99
Whole set	0.99	0.95	0.99

Results

TDr performance versus the consensus binary visual assessment is summarized in Table 3, together with SUVr and ELBA values. TDr shows excellent results both on the whole dataset and on single center cohorts.

The bootstrapped, cross-validation results are summarized in Table 4. Within the same procedure, we computed the cut-off values: $c_{TDr} = 0.611$ (0.610–0.620) (95% CI). Similarly, $c_{ELBA} = 0.956$ (0.944–0.958) and $c_{SUVr} = 1.133$ (1.113–1.149).

Figure 2 shows the direct comparison of the quantifier values together with the visual evaluation (negative/positive label). c_{SUVr}, c_{ELBA}, and c_{TDr} are also reported in figure.

Correlation results are summarized in Table 5. TDr significantly correlates with both SUVr and ELBA although it relates better with ELBA, as evidenced by the correlation coefficients on the negative and positive classes separately.

The impact of scan provenance is shown in Fig. 3. Negative and positive scans are grouped by center and plotted on the same scale for all quantifier methods; the common scale being the z-score values computed on the whole dataset (Fig. 3a). For comparison, Fig. 3b shows the distribution of the whole dataset grouped by binary evaluation.

For each method and for each visual class, we computed t test statistics among centers to look for possible bias in centers and/or methods. Significant differences are shown in Fig. 3a and summarized in Table 6. There is no clear pattern related to a specific center and/or method; rather, each method shows significant differences between cohorts that seem unrelated to other methods.

Finally, we assessed whether TDr is an independent measure. We took a linear model in the form $TDr = k_1\,(ELBA) + k_2$ and found $k_1 = 0.849$ $(0.765 - 0.934)$ and $k_2 = -0.149$ $(-0.226 - -0.073]$. Similarly for TDr vs. SUVr, $k_1 = 0.464$ $(0.370 - 0.559)$ and $k_2 = 0.068$ $(-0.042 - 0.179]$. Residual analysis showed no significant deviations from normality (Lilliefors test, p value $< 10^{-3}$) in both models. Moreover, the residuals and the predictor were found to be uncorrelated and the linear regression of the residuals versus ELBA was found to be compatible with the null model. The same analysis was carried on by switching the dependent variable, and with TDr vs. SUVr; all yielded no significant relations. Graphical representations for this latter analysis are shown in figure 5a, figure 5b, figure 6a, and figure 6b (supplementary materials).

Discussion

This work describes a semi-quantification method to be applied to amyloid PET scans. It is based on the typical properties of a non-receptorial tracer kinetics, that is, the radioligand exhibits a perfusion-like trait if acquired for a short time after the injection.

As [^{18}F]florbetapir and the other radiopharmaceuticals to image brain amyloidosis are lipophilic, they cross the blood–brain barrier and distribute to the brain as a function of rCBF, with a ratio of about 0.8 (at least for C-11 PiB) [21]. It may be assumed that these early scans are a rCBF surrogate since a very good correlation has been demonstrated with ^{18}F-FDG PET data [44] according to the well-known rCBF metabolism coupling.

Clearly, a highly perfused region has the capability of good imaging contrast because it holds the highest probability of ligand binding (if amyloid is present) and high wash-out rate (when there is no amyloid). The drawback to exploit this information is that we need an ad-hoc acquisition consisting of two separate scans per subject.

We should underline though that the term "high perfusion regions" does not signify a selection of the most perfused GM volumes. We identify the cortex areas by exploiting the differential flux between the WM and GM so that the "high perfusion regions" are indeed the "only" perfused regions. That is, the GM is in general more perfused than the WM.

When we select the 85% percentile on the global WM + GM brain ROI, we actually select the only (relatively) more perfused part: the GM (plus some spillover onto the WM). Therefore, even when we consider a lesser perfusion of the GM due to aging, pharmacological treatment, or atrophy, the GM will still be selected as the relatively more perfused part in the GM + WM volume. In other words, we do not select the highest perfusion ROIs *within* the GM, but rather we select the whole GM (be it normal or relatively preserved), which is identified in the same space and with the same resolution as the late PET image.

Some notable exceptions exist however, such as focal loss of perfusion due, for instance, to seizures, strokes, or advanced atrophy. In these cases, the D_E domain (i.e., the selected GM in the early scan) would not include these areas but we would not expect to evaluate them either.

Table 4 Bootstrap estimation of accuracy, specificity, and sensitivity versus visual assessment calculated on the entire dataset

	Accuracy (95% CI)	Specificity	Sensitivity
TDr	0.945 (0.937, 0.951)	0.933 (0.931, 0.934)	0.957 (0.928, 0.970)
SUVr	0.863 (0.846, 0.881)	0.842 (0.831, 0.859)	0.888 (0.854, 0.908)
ELBA	0.955 (0.944, 0.958)	0.958 (0.958, 0.959)	0.953 (0.930, 0.957)

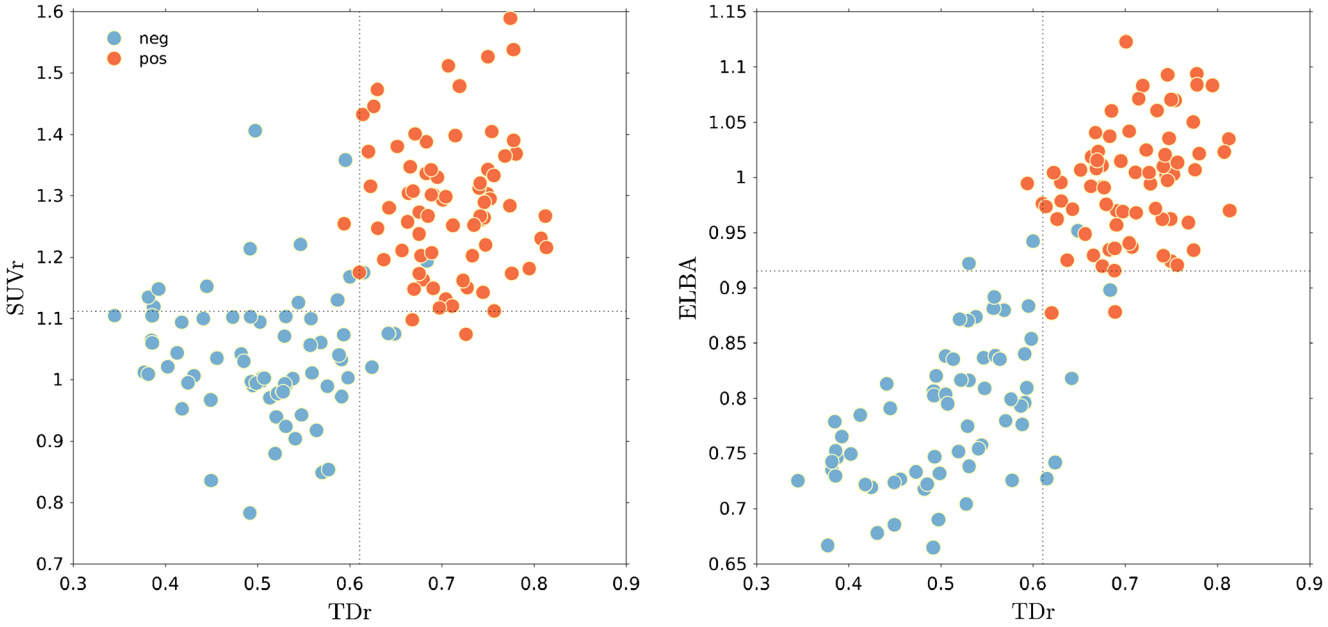

Fig. 2 Scatter plot of all three semi-quantification methods versus the visual assessment. Dotted horizontal and vertical lines show the cut-off (c_{SUVr} = 1.133, c_{ELBA} = 0.956, c_{TDr} = 0.611)

Early image acquisition can provide insights regarding rCBF (which is coupled with metabolism and thus indirectly, with synaptic function/dysfunction), and at the same time can be exploited to help in the quantification of the late scan. Indeed, the introduction of early scans is becoming more and more popular for it is investigated to be a proxy even of ^{18}F-FDG PET [44–47]. Besides, the acquisition of an early scan poses little inconvenience on the patient and on the scanner management.

We have therefore hoped that more data will be available with two scans per subject, that we shall use to consolidate the validation procedure.

Methodologically, TDr relies on a (crude) estimation of high perfusion volumes, which are closely related to the cortical thickness. Works in literature show a rather large spectrum of values, depending also on the used techniques. These however should not be taken at face value as we must take into consideration the peculiarity of the estimation technique and its meaning and equivalent for PET. For instance, ASL-based perfusion values may be affected by partial volume effect and blurring that can lead to underestimation of GM and overestimation of WM perfusion amplitude. Anyway, we showed that the estimation of perfusion volume—defined by imposing a threshold on the intensity statistics of the early scan——is not overly critical, so that a mean value can be effectively used for the sake of robustness.

A particular, potential useful characteristic of TDr is that its computation depends on the definition of regions with the high rCBF and thus should be less affected by the issue of reduced tracer delivery because of focal hypoperfusion. This characteristic may be especially relevant not only in cross-sectional but even more in longitudinal studies because it is known that in AD as well as in other neurodegenerative diseases both atrophy increases and rCBF decreases with time as a function of hypometabolism. Using TDr should ensure that the cortical ROI are always positioned in regions with sufficient rCBF and not in severely atrophic or hypoperfused ones. Whether TDr could improve semi-quantification in longitudinal studies remains to be tested although preliminary analyses (to be published) seem to confirm this hypothesis.

Concerning overall performance, the TDr is an independent measure with very good accuracy and on par—if not better—with previously validated methods.

As TDr is mathematically similar to SUVr (a ratio between two uptakes, averaged over some ROIs), one would expect a closer relationship with it. Instead, we find TDr to be more related to ELBA, both as correlation on the whole population and on the separate negative and positive subjects. This could be ascribed to the inherent SUVr variability due to the fact that uptake and reference ROIs are predefined in size and positioning, which can lead to larger errors on the single subject. This variability underlines the need to use the subject's MRI when

Table 5 Between methods correlation coefficients

Methods	Pearson r (p value)		
	All	Negative	Positive
TDr/SUVr	0.63 ($< 10^{-3}$)	0.02 (0.86)	0.02 (0.86)
TDr/ELBA	0.86 ($< 10^{-3}$)	0.57 ($< 10^{-3}$)	0.31 (0.006)
SUVr/ELBA	0.69 ($< 10^{-3}$)	0.23 (0.05)	− 0.006 (0.96)

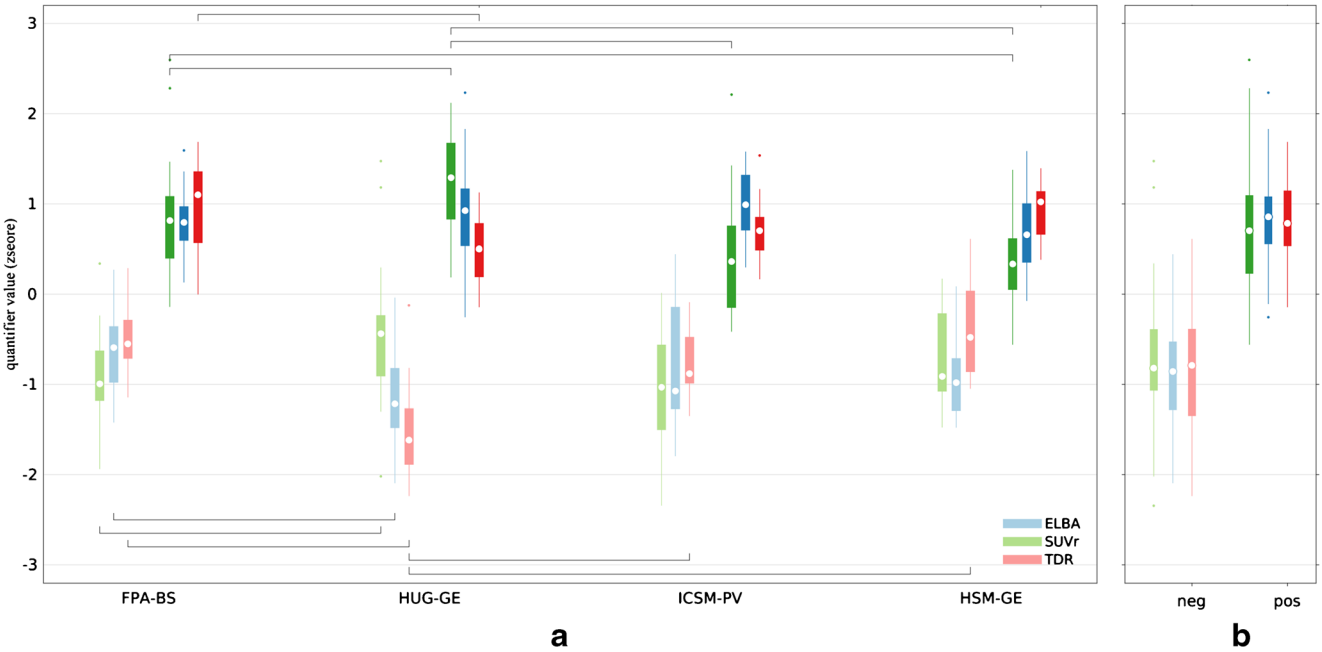

Fig. 3 Boxplot of ELBA, SUVr, and TDr (*z*-score) versus center and visual assessment (**a**). The whole dataset distribution (*z*-score) in (**b**). Between cohorts (same method) significant differences are marked with brackets (*t* test p value < 0.05); a summary is also shown in Table 6. Negative scans are represented with lighter hues, positive with darker hues. White dots inside the colored boxes are the medians

positioning ROIs, a step that was lacking in the SUVr computation of this study.

The intra-class correlation between TDr and ELBA impacts not only the TDr validation but it also has pathophysiological implications: it confirms that the transition between the negative and positive state is neither abrupt nor untraceable and that the concept of borderline scans is actually not a mere technical nuisance.

Results grouped by center shows some uneven behavior. The discrepancies can depend on several factors: on the one hand, there is the heterogeneity in patient selection, scanner, and the image reconstruction protocol; on the other hand, each semi-quantification method may respond differently to scan type, quality, and reconstruction parameters.

We argue that the latter hypothesis is indeed the most likely, as we do not see a consistent response of all methods on a particular center; rather, differences are scattered in methods, centers, and visual assessment. It is therefore likely that significant cohort differences be attributed to the sensitivity each quantifier has with respect to the hardware and reconstruction parameters, that is, due to the specific methodological approach. For instance, the peculiar response for HUG-GE could be explained by the larger number of NA subjects and by the difference in acquisition protocol (lowest injected dose paired with the newest scanner model).

Among the technical heterogeneities, we could mention the variability on the acquisition time of the early scan, which might be of concern since it captures a significant transient

Table 6 Significant differences among cohorts in each visual class and by quantifier method (*t* test). Upper triangular slots refer to negative subjects (light hues); lower triangular slots refer to positive subjects (darker hues)

	FPA-BS	HUG-GE	ICSM-PV	HSM-GE
FPA-BS		XXX		
HUG-GE	XX		X	X
ICSM-PV		X		
HSM-GE	X	X		

X = ELBA; X = SUVr; X = TDr

in the kinetics. The early scan timings defined in this work (5–8 min) has been agreed upon by our Nuclear Medicine physicians to be a reasonable range that takes into consideration the variability due to patient handling and possible nuisances in the positioning in the scanner. In practice though, all participating centers followed the stricter acquisition interval (0 min–1 min) as the starting time, to (5 min–7 min) as the end time.

As concluding remark, one might observe from Table 6 that ELBA is different among centers in just one case while SUVr is different in 4 cases and TDr in 5 cases. Indeed, the lack of uptake and reference regions that characterizes the ELBA method is a strong point in favor of this approach and might explain the better robustness with respect to data provenance. However, ELBA is not immune to the nuisances typical of multicenter studies, such as the signal-to-noise ratio and severe differences in the image reconstruction methods (as also discussed in [48]). Hence, we believe there is no single winner in the semi-quantification race, but all methods must be considered and—possibly—integrated in order to deliver more reliable results. On this note, and taking into consideration the residuals independency, we argue that a suitable weighted average of all three methods would be a more robust estimate of the brain amyloid burden.

Study limitations

The main limitation of this study is the lack of histopathological validation. While this is true for most semi-quantification works, this study also lack the cerebrospinal fluid (CSF) assays. In our dataset, we only have a handful of patients with CSF-Aβ42 and tau assays, and these were collected from different centers (3 from HSM-GE and 6 from HUG-GE); a plot summarizing these data is provided in the supplementary materials (figure 7). While we believe that a further validation of TDr should entail the comparison with Aβ42 CSF levels, we can point out that ELBA was successfully compared with CSF on a much larger number of subjects and that the strong correlation of TDr with ELBA suggest that TDr would perform equally well.

Another limitation of this study is the lack of data with other fluorinated tracers. This is an unfortunate consequence caused by the need of the supplementary early acquisition, which is usually not available in larger, public dataset such as the ADNI. However, we refer to Chincarini et al. [48] for a thorough validation of ELBA and SUVr on a naturalistic population using all three fluorinated tracers. The substantial equivalent of all tracers both for negative/positive contrast and between quantification methods, together with the similar non-receptorial characteristics, suggest that the TDr should perform similarly on the other tracers (florbetaben and flutemetamol).

Funding information E.P. was supported by Airalzh Onlus—COOP Italia (grant no. 138812/Rep n° 2459). V.G. was supported by the Swiss National Science Foundation (grant no. 320030_169876) and by the Velux foundation (grant no. 1123).

Compliance with ethical standards

The scans were acquired in the clinical setting for diagnostic purposes. All subjects (or their legal representative, if demented) were informed that their scans would have been used for research purposes and gave their written consent. All procedures performed were in accordance with the ethical standards of each local institutional Ethics Committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

The supervising ethics committee for this study is the CER (Comitato Etico della Regione Liguria), based in Genoa, Italy. Ethics Committees approvals included the transfer of imaging data, all anonymized brain amyloid PET were collected from the centers in DICOM format.

Quality of images was checked by an experienced Nuclear Medicine Physician (S.M.).

Conflict of interest In the past years, Dr. Nobili received fees from Eli-Lilly & Co for giving teaching course on visual reading of [^{18}F]Florbetapir, and from Bayer Pharma for participation in an advisory board on [^{18}F]Florbetaben.

Dr. Pardini receives research support from Novartis and Nutricia and received personal fees from Novartis, Merck, Roche. Dr. Pardini is partly supported by a Curiosity-driven grant from the University of Genoa.

All other authors disclose no conflict of interest.

References

1. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14:535–62. https://doi.org/10.1016/j.jalz.2018.02.018.
2. Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol. 2014;13:614–29.
3. Ikonomovic MD, Klunk WE, Abrahamson EE, Mathis CA, Price JC, Tsopelas ND, et al. Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. Brain. 2008;131:1630–45. https://doi.org/10.1093/brain/awn016.
4. Clark CM, Pontecorvo MJ, Beach TG, Bedell BJ, Coleman RE, Doraiswamy PM, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-β plaques: a prospective cohort study. Lancet Neurol. 2012;11:669–78.
5. Thal DR, Beach TG, Zanette M, Heurling K, Chakrabarty A, Ismail A, et al. [18 F]flutemetamol amyloid positron emission tomography in preclinical and symptomatic Alzheimer's disease: Specific detection of advanced phases of amyloid-β pathology. Alzheimers Dement. 2015;11:975–85. https://doi.org/10.1016/j.jalz.2015.05.018.
6. Villemagne VL, Ong K, Mulligan RS, Holl G, Pejoska S, Jones G, et al. Amyloid imaging with (18)F-florbetaben in Alzheimer disease and other dementias. J Nucl Med. 2011;52:1210–7.
7. Wolz R, Schwarz AJ, Gray KR, Yu P, Hill DLG, Alzheimer's Disease Neuroimaging Initiative. Enrichment of clinical trials in MCI due to AD using markers of amyloid and neurodegeneration. Neurology. 2016;87:1235–41.
8. Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical

practice and assessing response to therapy. Semin Ultrasound CT MR. 2010;31:496–505.
9. Thurfjell L, Lilja J, Lundqvist R, Buckley C, Smith A, Vandenberghe R, et al. Automated quantification of 18F-flutemetamol PET activity for categorizing scans as negative or positive for brain amyloid: concordance with visual image reads. J Nucl Med. 2014;55:1623–8.
10. Mattsson N, Insel PS, Landau S, Jagust W, Donohue M, Shaw LM, et al. Diagnostic accuracy of CSF Ab42 and florbetapir PET for Alzheimer's disease. Ann Clin Transl Neurol. 2014;1:534–43. https://doi.org/10.1002/acn3.81.
11. Jagust WJ, Landau SM, Koeppe RA, Reiman EM, Chen K, Mathis CA, et al. The Alzheimer's disease neuroimaging initiative 2 PET core: 2015. Alzheimers Dement. 2015;11:757–71. https://doi.org/10.1016/j.jalz.2015.05.001.
12. Klunk WE, Koeppe RA, Price JC, Benzinger TL, Devous MD Sr, Jagust WJ, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement. 2015;11:1–15.e1–4.
13. Tryputsen V, DiBernardo A, Samtani M, Novak GP, Narayan VA, Raghavan N, et al. Optimizing regions-of-interest composites for capturing treatment effects on brain amyloid in clinical trials. J Alzheimers Dis. 2015;43:809–21.
14. Su Y, Blazey TM, Snyder AZ, Raichle ME, Marcus DS, Ances BM, et al. Partial volume correction in quantitative amyloid imaging. Neuroimage. 2015;107:55–64.
15. 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.
16. 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.
17. Cecchin D, Barthel H, Poggiali D, Cagnin A, Tiepolt S, Zucchetta P, et al. A new integrated dual time-point amyloid PET/MRI data analysis method. Eur J Nucl Med Mol Imaging. 2017;44:2060–72.
18. Landau SM, Thomas BA, Thurfjell L, Schmidt M, Margolin R, Mintun M, et al. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. Eur J Nucl Med Mol Imaging. 2014;41:1398–407.
19. Johnson KA, Minoshima S, Bohnen NI, Donohoe KJ, Foster NL, Herscovitch P, et al. Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. J Nucl Med. 2013;54:476–90. https://doi.org/10.2967/jnumed.113.120618.
20. Wong DF, Rosenberg PB, Zhou Y, Kumar A, Raymont V, Ravert HT, et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18). J Nucl Med. 2010;51:913–20. https://doi.org/10.2967/jnumed.109.069088.
21. Price JC, Klunk WE, Lopresti BJ, Lu X, Hoge JA, Ziolko SK, et al. Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh Compound-B. J Cereb Blood Flow Metab. 2005;25:1528–47.
22. Contractor KB, Kenny LM, Coombes CR, Turkheimer FE, Aboagye EO, Rosso L. Evaluation of limited blood sampling population input approaches for kinetic quantification of [18F]fluorothymidine PET data. EJNMMI Res. 2012;2:11. https://doi.org/10.1186/2191-219x-2-11.
23. Meyer PT, Hellwig S, Amtage F, Rottenburger C, Sahm U, Reuland P, et al. Dual-biomarker imaging of regional cerebral amyloid load and neuronal activity in dementia with PET and 11C-labeled Pittsburgh compound B. J Nucl Med. 2011;52:393–400.
24. van Osch MJP, Teeuwisse WM, van Walderveen MAA, Hendrikse J, Kies DA, van Buchem MA. Can arterial spin labeling detect white matter perfusion signal? Magn Reson Med. 2009;62:165–73.
25. Roberts DA, Detre JA, Bolinger L, Insko EK, Leigh JS. Quantitative magnetic resonance imaging of human brain perfusion at 1.5 T using steady-state inversion of arterial water. Proc Natl Acad Sci. 1994;91:33–7. https://doi.org/10.1073/pnas.91.1.33.
26. Asllani I, Borogovac A, Brown TR. Regression algorithm correcting for partial volume effects in arterial spin labeling MRI. Magn Reson Med. 2008;60:1362–71.
27. Ye FQ, Berman KF, Ellmore T, Esposito G, van Horn JD, Yang Y, et al. H215O PET validation of steady-state arterial spin tagging cerebral blood flow measurements in humans. Magn Reson Med. 2000;44:450–6. https://doi.org/10.1002/1522-2594(200009)44:3<450::aid-mrm16>3.0.co;2-0.
28. Law I, Iida H, Holm S, Nour S, Rostrup E, Svarer C, et al. Quantitation of regional cerebral blood flow corrected for partial volume effect using O-15 water and PET: II. Normal values and gray matter blood flow response to visual activation. J Cereb Blood Flow Metab. 2000;20:1252–63.
29. Momjian S, Owler BK, Czosnyka Z, Czosnyka M, Pena A, Pickard JD. Pattern of white matter regional cerebral blood flow and autoregulation in normal pressure hydrocephalus. Brain. 2004;127:965–72.
30. Owler BK, Momjian S, Czosnyka Z, Czosnyka M, Péna A, Harris NG, et al. Normal pressure hydrocephalus and cerebral blood flow: a PET study of baseline values. J Cereb Blood Flow Metab. 2004;24:17–23.
31. Leinonen V, Rinne JO, Wong DF, Wolk DA, Trojanowski JQ, Sherwin PF, et al. Diagnostic effectiveness of quantitative [18F]flutemetamol PET imaging for detection of fibrillar amyloid β using cortical biopsy histopathology as the standard of truth in subjects with idiopathic normal pressure hydrocephalus. Acta Neuropathol Commun. 2014;2. https://doi.org/10.1186/2051-5960-2-46.
32. Landau SM, Breault C, Joshi AD, Pontecorvo M, Mathis CA, Jagust WJ, et al. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. J Nucl Med. 2013;54:70–7. https://doi.org/10.2967/jnumed.112.109009.
33. Doraiswamy PM, Sperling RA, Coleman RE, Johnson KA, Reiman EM, Davis MD, et al. Amyloid-β assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study. Neurology. 2012;79:1636–44.
34. Chen K, Roontiva A, Thiyyagura P, Lee W, Liu X, Ayutyanont N, et al. Improved power for characterizing longitudinal amyloid- PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. J Nucl Med. 2015;56:560–6. https://doi.org/10.2967/jnumed.114.149732.
35. 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.
36. Brendel M, Högenauer M, Delker A, Sauerbeck J, Bartenstein P, Seibyl J, et al. Improved longitudinal [(18)F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction. Neuroimage. 2015;108:450–9.
37. Fleisher AS, Joshi AD, Sundell KL, Chen Y-F, Kollack-Walker S, Lu M, et al. Use of white matter reference regions for detection of change in florbetapir positron emission tomography from completed phase 3 solanezumab trials. Alzheimers Dement. 2017;13:1117–24.
38. Su Y, Blazey TM, Owen CJ, Christensen JJ, Friedrichsen K, Joseph-Mathurin N, et al. Quantitative amyloid imaging in autosomal dominant Alzheimer's disease: results from the DIAN study group. PLoS One. 2016;11:e0152082.

39. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A. 2000;97:11050–5.
40. Sisodiya S, Free S, Fish D, Shorvon S. MRI-based surface area estimates in the normal adult human brain: evidence for structural organisation. J Anat. 1996;188(Pt 2):425–38.
41. Jones SE, Buchbinder BR, Aharon I. Three-dimensional mapping of cortical thickness using Laplace's equation. Hum Brain Mapp. 2000;11:12–32. https://doi.org/10.1002/1097-0193(200009)11:1<12::aid-hbm20>3.0.co;2-k.
42. Hutton C, De Vita E, Ashburner J, Deichmann R, Turner R. Voxel-based cortical thickness measurements in MRI. NeuroImage. 2008;40:1701–10. https://doi.org/10.1016/j.neuroimage.2008.01.027.
43. Lerch JP, Pruessner JC, Zijdenbos A, Hampel H, Teipel SJ, Evans AC. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. Cereb Cortex. 2005;15:995–1001.
44. Daerr S, Brendel M, Zach C, Mille E, Schilling D, Zacherl MJ, et al. Evaluation of early-phase [18 F]-florbetaben PET acquisition in clinical routine cases. NeuroImage: Clinical. 2017;14:77–86. https://doi.org/10.1016/j.nicl.2016.10.005.
45. Ottoy J, Miedema M, De Puydt C, Verhaeghe J, Deleye S, Engelborghs S, et al. Early frame 18F-AV45 and 18F-FDG-PET as proxies of CBF: comparison to 15O-H2O PET data. Alzheimers Dement. 2017;13:P763–4. https://doi.org/10.1016/j.jalz.2017.06.1018.
46. Rostomian AH, Madison C, Rabinovici GD, Jagust WJ. Early 11C-PIB frames and 18F-FDG PET measures are comparable: a study validated in a cohort of AD and FTLD patients. J Nucl Med. 2011. https://doi.org/10.2967/jnumed.110.082057.
47. Chen YJ, Rosario BL, Mowrey W, Laymon CM, Lu X, Lopez OL, et al. Relative 11C-PiB delivery as a proxy of relative CBF: quantitative evaluation using single-session 15O-water and 11C-PiB PET. J Nucl Med. 2015;56:1199–205.
48. Chincarini A, Peira E, Morbelli S, Pardini M, Bauckneht M, Arbizu J, et al. Semi-quantification and grading of amyloid PET: a project of the European Alzheimer's Disease Consortium (EADC). NeuroImage: Clinical. 2019;23:101846. https://doi.org/10.1016/j.nicl.2019.101846.