



Parametric imaging of dual-time window [^{18}F]flutemetamol and [^{18}F]florbetaben studies

Fiona Heeman^{a,*}, Maqsood Yaqub^a, Janine Hendriks^a, Ilona Bader^a, Frederik Barkhof^{a,b}, Juan Domingo Gispert^{c,d,e,f}, Bart N.M. van Berckel^a, Isadora Lopes Alves^a, Adriaan A. Lammertsma^a

^a Amsterdam UMC, Vrije Universiteit Amsterdam, Radiology and Nuclear Medicine, Amsterdam Neuroscience, De Boelelaan 1117, Amsterdam, the Netherlands

^b UCL, Institutes of Neurology and Healthcare Engineering, London, United Kingdom

^c Barcelonaβeta Brain Research Centre, Pasqual Maragall Foundation, Barcelona, Spain

^d Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain

^e Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain

^f IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

ARTICLE INFO

Keywords:

Amyloid PET
[^{18}F]flutemetamol
[^{18}F]florbetaben
Parametric imaging
PET quantification

ABSTRACT

Optimal pharmacokinetic models for quantifying amyloid beta ($A\beta$) burden using both [^{18}F]flutemetamol and [^{18}F]florbetaben scans have previously been identified at a region of interest (ROI) level. The purpose of this study was to determine optimal quantitative methods for parametric analyses of [^{18}F]flutemetamol and [^{18}F]florbetaben scans. Forty-six participants were scanned on a PET/MR scanner using a dual-time window protocol and either [^{18}F]flutemetamol ($N=24$) or [^{18}F]florbetaben ($N=22$). The following parametric approaches were used to derive DVR estimates: reference Logan (RLogan), receptor parametric mapping (RPM), two-step simplified reference tissue model (SRTM2) and multilinear reference tissue models (MRTM0, MRTM1, MRTM2), all with cerebellar grey matter as reference tissue. In addition, a standardized uptake value ratio (SUVR) was calculated for the 90–110 min post injection interval. All parametric images were assessed visually. Regional outcome measures were compared with those from a validated ROI method, i.e. DVR derived using RLogan. Visually, RPM, and SRTM2 performed best across tracers and, in addition to SUVR, provided highest AUC values for differentiating between $A\beta$ -positive vs $A\beta$ -negative scans ([^{18}F]flutemetamol: range AUC=0.96–0.97 [^{18}F]florbetaben: range AUC=0.83–0.85). Outcome parameters of most methods were highly correlated with the reference method ($R^2 \geq 0.87$), while lowest correlation were observed for MRTM2 ($R^2=0.71$ –0.80). Furthermore, bias was low ($\leq 5\%$) and independent of underlying amyloid burden for MRTM0 and MRTM1. The optimal parametric method differed per evaluated aspect; however, the best compromise across aspects was found for MRTM0 followed by SRTM2, for both tracers. SRTM2 is the preferred method for parametric imaging because, in addition to its good performance, it has the advantage of providing a measure of relative perfusion (R_1), which is useful for measuring disease progression.

Introduction

Early detection of amyloid-beta ($A\beta$) plaques has become increasingly relevant, in particular with respect to identifying individuals for secondary prevention trials and monitoring disease progression in Alzheimer's disease (AD) (Farrar et al., 2017; Collij et al., 2020). These $A\beta$ plaques can be identified in vivo with positron emission tomography (PET) scans, using either a dichotomous classification of $A\beta$ burden (Mallik et al., 2017) or a more fine-grained quantitative measure of $A\beta$, which is required for measuring the extent of pathology or quantitative tracking disease progression (Lammertsma, 2017).

For the latter quantitative applications, it has been suggested that dynamic or dual-time window scanning protocols should be used for obtaining the most accurate $A\beta$ estimates (Berckel BNM van et al., 2013; Bullich et al., 2018; Heeman et al., 2019) and, since amyloid pathology might not follow anatomical boundaries, voxel-wise quantitative analyses might be preferred as they provide more detailed spatial information, independent of predefined regions. In addition, quantitative parametric images allow for voxel-by-voxel analysis between groups or as a function of time within the same subjects. Thus, comprehensive evaluation of parametric methods used for voxel-wise analysis is warranted. However, as different tracers have different kinetics, the optimal parametric method for quantification needs to be established for each $A\beta$ tracer separately. In addition, the ability to obtain a voxel-wise measure of the relative tracer influx rate (R_1) could play a role

* Corresponding author.

E-mail addresses: f.heeman@amsterdamumc.nl, f.heeman@vumc.nl (F. Heeman).

<https://doi.org/10.1016/j.neuroimage.2021.117953>

Received 28 October 2020; Received in revised form 12 January 2021; Accepted 5 March 2021

Available online 21 March 2021

1053-8119/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

when deciding which method is most suitable for parametric analysis. This additional parameter can be used as a proxy for (relative) cerebral blood flow (Heeman et al., 2021) and may therefore be considered as a measure of disease severity or progression, as first shown by single photon emission computed tomography (SPECT) and [^{15}O]H $_2$ O-PET studies, and more recently replicated with arterial spin labelling (ASL) (Collij et al., 2016; Binnewijzend et al., 2016; Jagust et al., 1997).

For most amyloid tracers, voxel-wise quantitative approaches have been evaluated to a certain extent (Bullich et al., 2018; Yaqub et al., 2008; Verfaillie et al., 2021; Heurling et al., 2015). However, [^{18}F]flutemetamol and [^{18}F]florbetaben studies were performed with full dynamic acquisitions that did not include the recommended 90–110 min post injection (p.i.) scanning window used for visual assessment and to derive standard uptake value ratios (SUVR) (Heurling et al., 2015; Becker et al., 2013). Further, only Heurling and colleagues evaluated parametric methods for full dynamic [^{18}F]flutemetamol studies by comparing results with those obtained using the gold standard, i.e. full kinetic analysis using an arterial plasma input function (Heurling et al., 2015). They showed that best correlations with the gold standard were obtained using reference Logan (RLogan), the 70–90 min standardized uptake value ratio (SUVR $_{70-90}$) and receptor parametric mapping (RPM) ($R^2 > 0.94$). With respect to full dynamic [^{18}F]florbetaben studies, the performance of two parametric methods has been evaluated only visually, while regional reference tissue models have been quantitatively compared with the gold standard (Becker et al., 2013). It remains unclear whether the results from full dynamic acquisitions remain valid for a dual-time window protocol, given that the performance of parametric methods could be compromised by the gap in the data. Previous work has shown that in some cases, a gap in the data may lead to sub-optimal curve fitting, which in turn may result in a small bias in the results (Heeman et al., 2019). This impact is mediated by the effect of noise and may potentially affect parametric methods in different ways due to differences in their sensitivity to noise. In addition, small interpolation or coregistration errors could also affect performance of parametric methods. These shortened acquisition protocols are especially relevant for fluorine-18 tracers such as [^{18}F]flutemetamol and [^{18}F]florbetaben, considering that routine acquisition of lengthy dynamic scans (110 min duration) is not feasible, especially for older participants (Bullich et al., 2018; Heeman et al., 2019). In addition to increased patient comfort, these protocols can increase patient throughput and result in more cost-efficient use of tracer batch productions compared with full dynamic acquisitions. While still less efficient than standard static acquisitions, a dual-time window protocol can provide an additional parameter (R_1), which can be advantageous for studies focusing on disease progression (Tiepol et al., 2016; Son et al., 2020–, (Daerr et al., 2016). These characteristics seem to have increased the popularity of dual-time window acquisitions, as can be seen by the adoption of these protocols by large-scale studies such as AMYPAD (Farrar et al., 2017).

The present study aimed to evaluate the performance of the most widely used parametric methods for the specific case of dual-time window scans using [^{18}F]flutemetamol and [^{18}F]florbetaben. Performance of these parametric methods was evaluated with respect to three different aspects: 1) visual assessment of parametric images 2) ability to differentiate between A β -positive and A β -negative scans, and 3) quantitative accuracy and precision with respect to a reference method.

Materials and methods

Participants

Forty-six cognitively unimpaired subjects were selected from the AMYPAD PNHS (Lopes Alves et al., 2020). In short, all subjects underwent neurological and neuropsychological assessment and had at least one PET and one MR scan available. All PET scans were visually assessed by a trained nuclear medicine physician (BvB) following the

formal guidelines for reading [^{18}F]flutemetamol and [^{18}F]florbetaben scans, as defined by the respective manufacturers. Approximately half of the participants ($N=24$) was scanned using [^{18}F]flutemetamol and the other half ($N=22$) using [^{18}F]florbetaben. For both tracers, 50% of the participants was A β -positive. All participants provided written informed consent before participating in the study in accordance with the Declaration of Helsinki. The Medical Ethics Review Committee of Amsterdam UMC, location VUmc, approved the study protocol (EudraCT Number: 2018-002,277-22).

Image acquisition

All subjects underwent a dynamic PET scan on a Philips Ingenuity TF PET/MR scanner (Philips Medical Systems, Best, The Netherlands), according to a dual-time window protocol (Heeman et al., 2019). This scanning protocol consisted of an initial dynamic scan from 0 to 30 min p.i. followed by a break of 60 min, and then a second scan from 90 to 110 min p.i. Prior to each of the scans, a T1-weighted gradient echo pulse MR sequence was acquired for attenuation correction purposes. At the start of the early PET scan, participants received a bolus injection of either [^{18}F]flutemetamol ($N=24$, 184 ± 11 MBq) or [^{18}F]florbetaben ($N=22$, 280 ± 18 MBq), following manufacturer dosage guidelines. Next, scans were reconstructed into 22 frames (6×5 , 3×10 , 4×60 , 2×150 , 2×300 , and 1×600 s for the first part of the scan, and 4×300 s for the second part) with a matrix size of $128 \times 128 \times 90$ and a voxel size of $2 \times 2 \times 2$ mm with a standard line-of-response-based row-action maximum-likelihood algorithm (LOR-RAMBLA) (Hu et al., 2007) for the brain. All usual corrections, e.g. for MR-based attenuation (MRAC) (Hu et al., 2010), decay, scatter, randoms and dead time were performed. In addition, structural T1-weighted MR images were acquired within, on average, 4.6 ± 3.0 months from the PET scan for [^{18}F]flutemetamol and 4.0 ± 2.5 months for [^{18}F]florbetaben (maximum difference between PET and MR scan was one year), respectively, also on the same Philips Ingenuity TF PET/MR scanner.

Image processing

First, all PET scans were visually checked for between-frame movement. As none showed motion, all were included for further analysis. Next, both structural T1-weighted MR images and early PET images (0–30 min p.i.) were co-registered to their corresponding late PET images (90–110 min p.i.) and visually checked. This was carried out in two steps: first the T1-weighted MR and early MRAC image were co-registered with the late MRAC image and then, the transformation matrix corresponding to the early MRAC was applied to each frame of the early PET image sequence. Following co-registration, the early and late parts of the dual-time window PET scan were combined into a single file using an in-house developed, MATLAB-based software tool. Subsequently, a PVE-lab-based implementation of SPM8 (Rask et al., 2004) was used to segment the MR image into grey matter, white matter and cerebrospinal fluid (CSF). The reference tissue (cerebellar grey matter) region of interest (ROI) was delineated based on the Hammers atlas (Hammers et al., 2003) and time-activity curves (TACs) were extracted.

Parametric analysis

Given that within the AMYPAD PNHS study no arterial input data were acquired, a regional reference tissue method was used as standard to evaluate the performance of all parametric methods. RLogan was chosen as the “regional standard” based on previous studies in which good results in terms of correlation and bias compared with the gold standard (arterial input) had been observed for this method, and stability against variation in confounding factors had also been demonstrated (Heeman et al., 2021; Nelissen et al., 2009).

First, missing data points from the reference tissue TACs were interpolated using the reversible two tissue compartment model (4 rate

Table 1
Subject demographics split by visual A β status.

	^[18F] flutemetamol			^[18F] florbetaben		
	All (N=24)	A β -negative (N=12)	A β -positive (N=12)	All (N=22)	A β -negative (N=11)	A β -positive (N=11)
Age	71.7 \pm 6.0	69.9 \pm 5.1	73.5 \pm 6.5	69.4 \pm 7.4	68.1 \pm 7.4	70.6 \pm 7.5
Females (%)	58.8	75.0**	41.7	63.6	72.7	54.5
MMSE	28.3 \pm 1.7	29.2 \pm 1.1*	27.5 \pm 1.9	28.9 \pm 1.2	29.5 \pm 0.7**	28.4 \pm 1.4

Values depicted as mean \pm SD.

* $p < 0.05$.

** $p < 0.10$, compared with the A β -positive group.

constants) with additional blood volume fraction parameter (2T4k_{V_b}) and a typical, tracer-specific plasma input function, based on the interpolation procedure described previously (Heeman et al., 2019). Next, the PPET software tool (Boellaard et al., 2006) was used to compute either distribution volume ratio (DVR), non-displaceable binding potential (BP_{ND}) or standardized uptake value ratio (SUVR) parametric images based on the following methods: RLogan, receptor parametric mapping (RPM), the voxel-based implementation of the two step simplified reference tissue model (SRTM2), multilinear reference tissue models (MRTM0, MRTM1, MRTM2), and SUVR (calculated for the 90–110 min p.i. interval), all with cerebellar grey matter as reference tissue (Logan et al., 1996; Gunn et al., 1997; Lammertsma and Hume, 1996; Wu and Carson, 2002; Ichise et al., 2003). In addition, parametric relative delivery (R_1) images were generated using RPM and SRTM2. For SRTM2, k_2^* was determined across all voxels with a BP_{ND} higher than 0.05 from a first RPM run, while this parameter was omitted in PPET's implementation of RLogan. In line with previous studies, the RLogan linearization start time (t^*) was set to 50 min p.i. for both tracers (Heeman et al., 2021; Heurling et al., 2015; Logan et al., 1996). For both tracers, RPM's basis function (BF) settings were first optimized. The optimization procedure consisted of evaluating a range of BF settings (see supplementary materials, Table S1) and defining the settings that resulted in the highest correlation and the lowest bias (as assessed by R^2 and slope of the regression line) between RPM and regional standard derived DVR. To allow for comparability across methods, $BP_{ND} + 1$ values (here referred to as DVR) were computed for methods with BP_{ND} as outcome parameter. Finally, the following cortical grey matter volumes of interest (VOIs) were superimposed on the parametric images in order to extract regional values for comparison with those estimated using the regional standard: anterior and posterior cingulate gyrus, middle and orbitofrontal gyrus, inferior and superior frontal gyrus, gyrus rectus, pre- and post-central gyrus, superior parietal gyrus, (infero)lateral remainder of the parietal lobe, insula, cuneus, lateral remainder of the occipital lobe, medial and lateral anterior temporal lobe, posterior temporal lobe, superior, middle and inferior temporal gyrus, lingual gyrus, parahippocampal and ambient gyrus, and the fusiform gyrus. Finally, a volume-weighted global cortical average was calculated from all regional data.

Statistical analysis

All statistical analyses were performed in IBM SPSS Statistics for Windows Version 26.0 (IBM Corp. Armonk New York U.S.A) or GraphPad Prism for Windows Version 7.04, (La Jolla California U.S.A.). These analyses can be divided into three groups: 1) assessment of population equivalence between ^[18F]flutemetamol and ^[18F]florbetaben scans, 2) visual assessment of parametric images 3) quantitative assessment of the performance of parametric methods compared with the regional standard.

Population equivalence

First, a between tracer comparison of potential differences in age and mini-mental state examination (MMSE) scores of the participants was

performed using nonparametric Mann-Whitney U tests, while potential differences in the proportion of males and females were compared using a chi-square test. Subsequently, a within tracer sub-group comparison of age and MMSE scores between A β -positive and A β -negative subjects was performed using the same tests as described above.

Visual assessment of parametric images

Image artifacts were identified first visually and subsequently by calculating the percentage of (extreme) outliers within total grey matter using a threshold that was defined based on expected clinical values: 2.00 for BP_{ND} and 3.00 for DVR and SUVR images (Heurling et al., 2015; Becker et al., 2013). Values larger than these thresholds were considered to be outliers. In addition, a more lenient and more stringent threshold were applied to demonstrate the robustness of the analyses, i.e. 1.80, 2.20 for BP_{ND} images and 2.80, 3.20 for DVR and SUVR images.

Quantitative performance of parametric methods

Performance of each parametric method was first determined by assessing possible differences in global cortical DVR, SUVR and R_1 derived from A β -positive and A β -negative scans. This was assessed using one-way analysis of variance (ANOVA) and the area under the curve (AUC) from a Receiver Operating Characteristic (ROC) curve which is indicative of how well a parameter is able to differentiate between A β -positive and A β -negative scans.

Next, linear regression analyses, corrected for age, sex and visual status, were used to assess the correlation (R^2) between binding estimates obtained using each parametric method and the regional standard. In addition, Bland-Altman analyses were used to assess bias compared to the regional standard, the variability of the data based on the 95% Limits of Agreement, and whether proportional bias was present. Any proportional bias was further determined by fitting a regression line through the Bland-Altman plot.

Results

Population equivalence

There were no significant between-tracer differences with respect to the participant's age, proportion of males and females or average MMSE scores (Table 1). Within each tracer group, MMSE scores were higher for A β -negative compared with A β -positive participants (^[18F]flutemetamol: $p = 0.014$, ^[18F]florbetaben: $p = 0.061$), while there were no differences in age or proportion males and females (Table 1).

Visual assessment of parametric images

Representative parametric BP_{ND} , DVR, SUVR and R_1 images of both A β -positive and A β -negative scans are presented in Fig. 1 for each tracer. Visual inspection of parametric BP_{ND} , DVR and SUVR images showed that, for both tracers, most image artifacts caused by outliers (i.e. speckles) were present for MRTM2 and MRTM1 (Fig. 1). Visually, no artifacts were observed in parametric R_1 images, and differences in R_1 derived from A β -positive and A β -negative scans were small. Furthermore, the

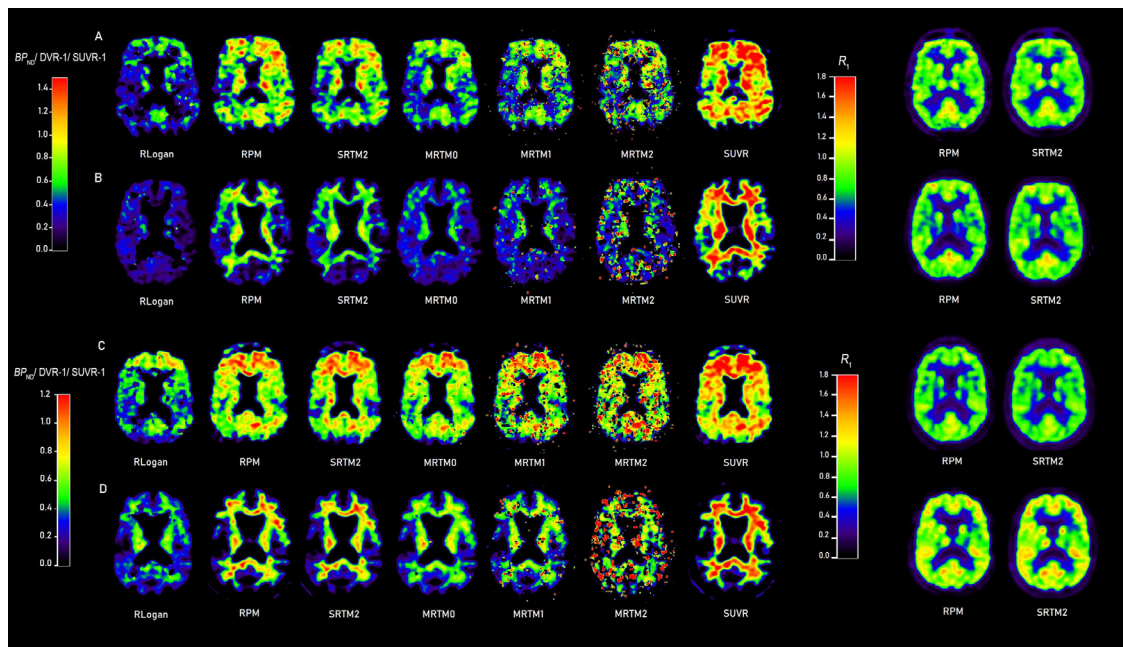


Fig. 1. Parametric amyloid and relative perfusion images

Parametric images of each of the methods for A) an A β -positive and B) an A β -negative [^{18}F]flutemetamol scan, and for C) an A β -positive and D) an A β -negative [^{18}F]florbetaben scan.

Note: [^{18}F]florbetaben images are shown in colour for comparison, while grey scale is recommended for visual assessment.

Table 2

AUC: Models' ability to correctly classify A β -positive and A β -negative scans.

	[^{18}F]flutemetamol	[^{18}F]florbetaben
RRLogan DVR	0.94	0.79
RLogan DVR	0.88	0.77
RPM DVR	0.97	0.85
SRTM2 DVR	0.96	0.84
MRTM0 DVR	0.94	0.80
MRTM1 DVR	0.95	0.80
MRTM2 DVR	0.81	0.79
SUVR	0.96	0.83
RPM R_1	0.54	0.45
SRTM2 R_1	0.49	0.45

RRLogan corresponds to the regional implementation of reference Logan (regional standard).

Values correspond to the Area Under the Curve (AUC) from ROC analyses.

quantitative outlier analysis showed that for [^{18}F]flutemetamol, most outliers were present in MRTM2, followed by MRTM1 and SUVR images (range: 0.3–2.8%, supplementary Table S2a). For [^{18}F]florbetaben, most outliers (range: 0.6–4.7%) were present in MRTM2 followed by MRTM1 images (supplementary Table S2b). For both tracers, results showed a very similar pattern using the alternative thresholds (supplementary Tables S2a and S2b).

Quantitative performance of parametric methods

For both tracers, all parametric methods showed a higher A β burden (as shown by DVR or SUVR values) for the A β -positive group ($p < 0.05$). Nevertheless, RPM and SRTM2 derived DVR and SUVR yielded highest area under the curve (AUC) values, as shown in Table 2. As expected, for both tracers there were no significant differences in R_1 between A β groups and AUC values of RPM and SRTM2-derived R_1 were low (Table 2).

With respect to [^{18}F]flutemetamol, linear regression analyses yielded high correlations ($R^2 \geq 0.87$) between outcome measures from all methods and regional standard derived DVR, except for MRTM2 where the

correlation was $R^2 = 0.71$ (Table 3 and Fig. 2A). Furthermore, compared with the regional standard, MRTM2 derived DVR showed a constant, average overestimation of 7%, as shown by the Bland-Altman analyses (Fig. 3A). On the other hand, SUVR and parametric RLogan derived DVR showed a bias of 10 and –5% compared with the regional standard, which was proportional to the underlying amyloid burden ($\text{SUVR}_{\text{slope}} = 0.49$, $p < 0.001$, $\text{parametric RLogan}_{\text{slope}} = -0.20$, $p < 0.001$) (Fig. 3A). In addition, proportional bias was also observed for RPM and SRTM2 ($\text{RPM}_{\text{slope}} = 0.23$, $p < 0.001$ ($\text{SRTM2}_{\text{slope}} = 0.15$, $p < 0.001$). Furthermore, Bland-Altman analyses showed most variability for MRTM2 and RPM derived DVR and SUVR (Fig. 3A).

With respect to [^{18}F]florbetaben, linear regression analyses yielded high correlations ($R^2 \geq 0.94$) between outcome measures of all parametric methods and regional standard derived DVR, except for MRTM2 where the correlation was $R^2 = 0.80$ (Table 3 and Fig. 2B). Furthermore, compared with the regional standard, MRTM2 derived DVR showed a constant, average overestimation of 11%, while SRTM2 showed a constant, average underestimation of 7%, as shown by the Bland-Altman analyses (Fig. 3B). On the other hand, SUVR, parametric RLogan and RPM derived DVR showed a small average bias (<5%), however, this bias was proportional to the underlying amyloid burden ($\text{SUVR}_{\text{slope}} = 0.31$, $p < 0.001$, $\text{parametric RLogan}_{\text{slope}} = -0.15$, $p < 0.001$, $\text{RPM}_{\text{slope}} = 0.15$, $p < 0.001$) (Fig. 3B). Finally, Bland-Altman analyses showed most variability for MRTM2 derived DVR and SUVR (Fig. 3B).

Discussion

In this study, various parametric methods for voxel-wise analysis of dual-time window [^{18}F]flutemetamol and [^{18}F]florbetaben studies were evaluated, considering the following three different aspects: 1) visual assessment of parametric images 2) ability to differentiate between A β -positive and A β -negative scans 3) quantitative accuracy and precision. All evaluated methods could differentiate between A β -positive and A β -negative scans based on DVR or SUVR measures, and they all showed high correlations with regional standard derived DVR, except for MRTM2. However, the quality of parametric images and the level

Table 3
Linear regression analysis.

[¹⁸ F] flutemetamol	Parametric RLogan	RPM	SRTM2	MRTM0	MRTM1	MRTM2	SUVR
R²	0.95	0.95	0.95	0.99	0.99	0.71	0.87
Slope	0.85	1.14	1.06	0.97	0.98	0.95	1.34
Intercept	0.20	−0.26	−0.18	0.05	0.06	0.15	−0.59
[¹⁸ F] florbetaben	Parametric RLogan	RPM	SRTM2	MRTM0	MRTM1	MRTM2	SUVR
R²	0.98	0.97	0.95	0.99	0.98	0.80	0.94
Slope	0.87	1.07	0.94	1.02	1.02	0.93	1.22
Intercept	0.11	−0.20	−0.03	−0.09	−0.07	0.20	−0.39

Regional RLogan was used as independent variable, and linear regression analyses were corrected for age, sex and visual status.

of bias compared with the regional standard varied substantially across methods. Hence, these findings demonstrate that the most appropriate parametric method depends on the research or clinical question to be addressed.

For both tracers, high correlations ($R^2 \geq 0.87$) were observed between DVR derived using the regional standard and the outcome measure from all parametric methods, except for MRTM2, which had relatively lower correlations ([¹⁸F]flutemetamol: $R^2=0.71$ and [¹⁸F]florbetaben: $R^2=0.80$) and showed a constant, average DVR overestimation of 7% for [¹⁸F]flutemetamol and 11% for [¹⁸F]florbetaben compared with the regional standard. It should be noted, that for interpreting the bias or accuracy of the results, it is important to take into account whether the bias is proportional to the underlying A β burden. In contrast to a constant bias, proportional bias cannot easily be accounted for in longitudinal studies and thereby pose challenges, in particular when aiming to measure small changes over time. In that regard, while small (max. 5%) on average (except for [¹⁸F]flutemetamol SUVR), the bias observed with the other methods was proportional to the underlying A β burden, especially in the case of SUVR, RPM and RLogan. For SRTM2 specifically, a small proportional bias was observed for [¹⁸F]flutemetamol, while a constant, average underestimation of 7% was seen for [¹⁸F]florbetaben.

The amyloid dependent bias, as reported above, could be explained by a variety of factors. One possible explanation could be the assumption of single-tissue compartment kinetics in both the target and reference tissues, which is assumed for RPM and SRTM2, but not for MRTM implementations (Gunn et al., 1997; Lammertsma and Hume, 1996; Wu and Carson, 2002–, Ichise et al., 2003). In fact, it has been reported that for these tracers the reference tissues' kinetics are better described by a 2TC model (Becker et al., 2013; Nelissen et al., 2009), possibly explaining the amyloid dependent bias for these methods, as demonstrated by Salinas and colleagues (Salinas et al., 2015). This assumption may be particularly problematic for regions with high A β burden, explaining the increase in bias as illustrated by Fig. 3. With respect to RLogan, the voxel-wise method showed marked differences with its regional implementation counterpart. Since all other aspects (e.g. linearization start time, method of interpolation for missing data-points and omission of k_2') were identical between both implementations, these differences should primarily be related to a general impact of noise on parametric imaging methods. In particular, it has been shown that underestimation of the amyloid burden with RLogan-derived DVR is more pronounced at higher noise levels and for regions with higher binding (Slifstein and Laruelle, 2000). Therefore, the parametric RLogan bias dependency on underlying A β burden is plausibly a result of the interaction between higher noise levels and a higher A β burden (Slifstein and Laruelle, 2000).

In general, the present results are well in line with previous amyloid PET studies, although for some methods, such as RPM and SUVR in case of [¹⁸F]flutemetamol, the bias was larger than previously reported (Heurling et al., 2015). For example, previous work has demonstrated an overestimation of SUVR compared with BP_{ND} or DVR obtained using plasma input modelling (the gold standard) (Berckel BNM van et al., 2013; Verfaillie et al., 2021; Heurling et al., 2015; Becker et al., 2013; Ottoy et al., 2017; Heeman et al., 2020). Since the reference method

used in the present study (regional RLogan) is known to underestimate true binding (Berckel BNM van et al., 2013; Verfaillie et al., 2021; Heurling et al., 2015; Heeman et al., 2020), this could explain the larger levels of bias observed in SUVR and RPM, compared with those reported by Heurling and colleagues (Heurling et al., 2015). However, given that the regional implementation of RLogan is less sensitive to noise compared with its parametric counterpart, and the A β burden covered by the CU participants is expected to be slightly lower than what is typically observed in AD patients, limited impact on the results was expected. With respect to both quantitative and visual performance of the various MRTM implementations, considerable between-tracer differences have been reported previously (Yaqub et al., 2008; Verfaillie et al., 2021; Heurling et al., 2015; (Becker et al., 2013). These differences in performance may be explained by the application (or lack) of pre- or post-reconstruction smoothing filters and differences in scanner resolution (Yaqub et al., 2008; Verfaillie et al., 2021; Heurling et al., 2015; Becker et al., 2013). Finally, small differences in the performance of parametric methods compared with previous reports can also be attributed to specific choices in processing and analysis pipelines, such as differences in SUVR uptake times, both starting and boundary values of fitting parameters, reference tissue selection and/or the inclusion of subcortical regions (Heurling et al., 2015; Becker et al., 2013). Of note, for a few data points, large differences were observed between MRTM2 and the regional standard in case of [¹⁸F]flutemetamol (Fig. 3A). All these data points belonged to a single subject and no large differences were observed for any of the other methods. Unfortunately, there is no obvious reason why data for this subject were different only for MRTM2.

Overall, the aim of this study was to evaluate different parametric methods on three different aspects. Regarding the first: *visual assessment of parametric images*, few image artefacts were observed for RLogan, RPM, SRTM2, SUVR and MRTM0, facilitating visual assessment. Furthermore, with respect to the tracer distribution, RPM and SRTM2, followed by MRTM0 images, appear most suitable. On the other hand, SUVR images showed high uptake in grey as well as white matter regions, increasing the risk of false positives, while RLogan images showed clearest underestimation of the A β burden, increasing the risk of false negatives, especially for [¹⁸F]flutemetamol (Fig. 1). Therefore, RPM and SRTM2 appear to be the methods of choice for visual assessment, followed by MRTM0. The second aspect was *the ability to differentiate between A β -positive and A β -negative scans*. For both tracers, all methods could detect significant differences between A β -positive and A β -negative scans based on DVR or SUVR measures and were therefore considered suitable for this purpose. However, RPM, SRTM2 and SUVR provided highest AUC values, making these the preferred methods for detecting more subtle differences in A β burden, followed by MRTM0 and MRTM1. Regarding the third aspect, *quantitative accuracy and precision*, methods should show a high correlation and low bias (independent of underlying A β burden) compared with the regional standard and low measurement variability. With respect to the first characteristic, all methods, except from MRTM2, would be considered appropriate, as all of them showed excellent correlations with the regional standard ($R^2 \geq 0.87$). Regarding the second, only MRTM0 and MRTM1 showed both minimal and constant bias (<5%), while all other methods showed

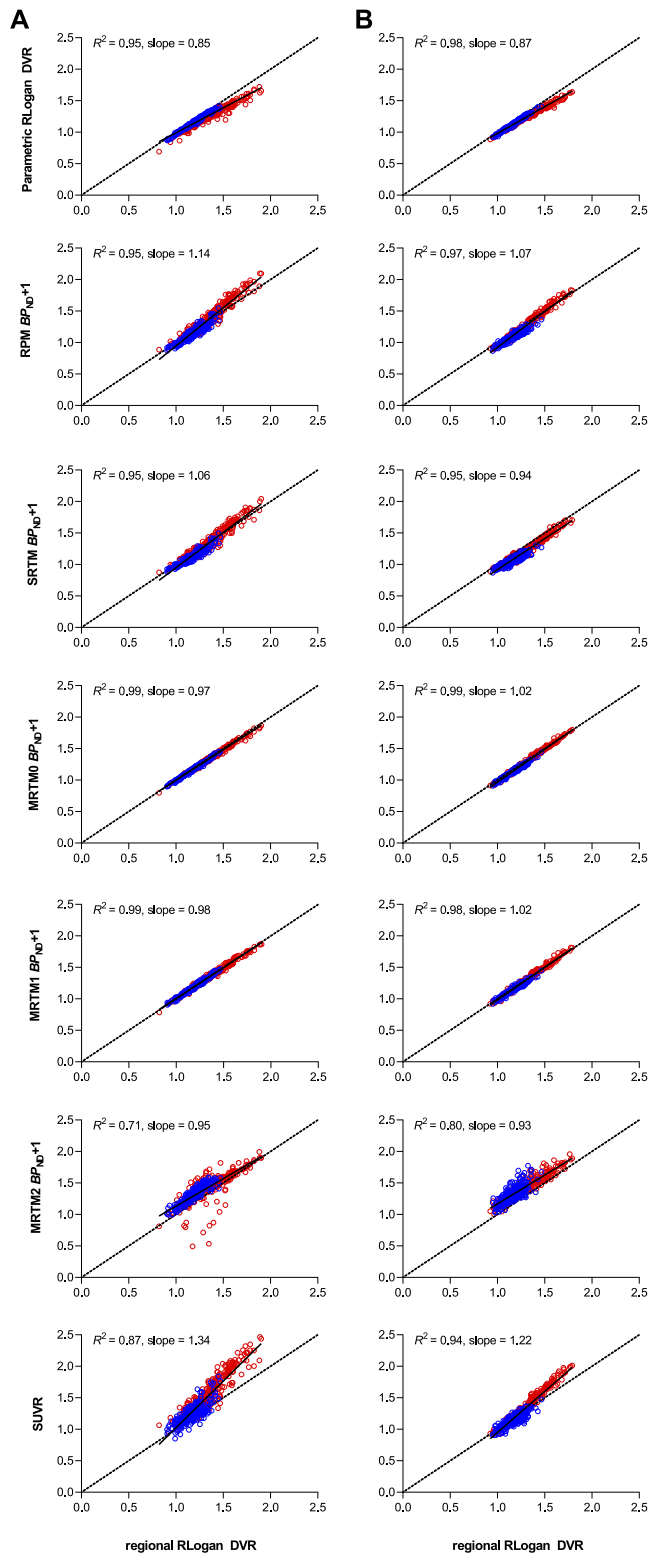


Fig. 2. Correlations between outcome measures of all parametric methods and the regional standard for [18 F]flutemetamol and [18 F]florbetaben. Correlations shown for A) [18 F]flutemetamol and B) [18 F]florbetaben, with regional RLogan DVR as regional standard for all regions of interest (ROIs). ROIs are colour coded based on the A β status of the scan they belong to, i.e. A β -positive = red, A β -negative = blue. The dashed line corresponds to the line of identity and the solid line corresponds to the regression line.

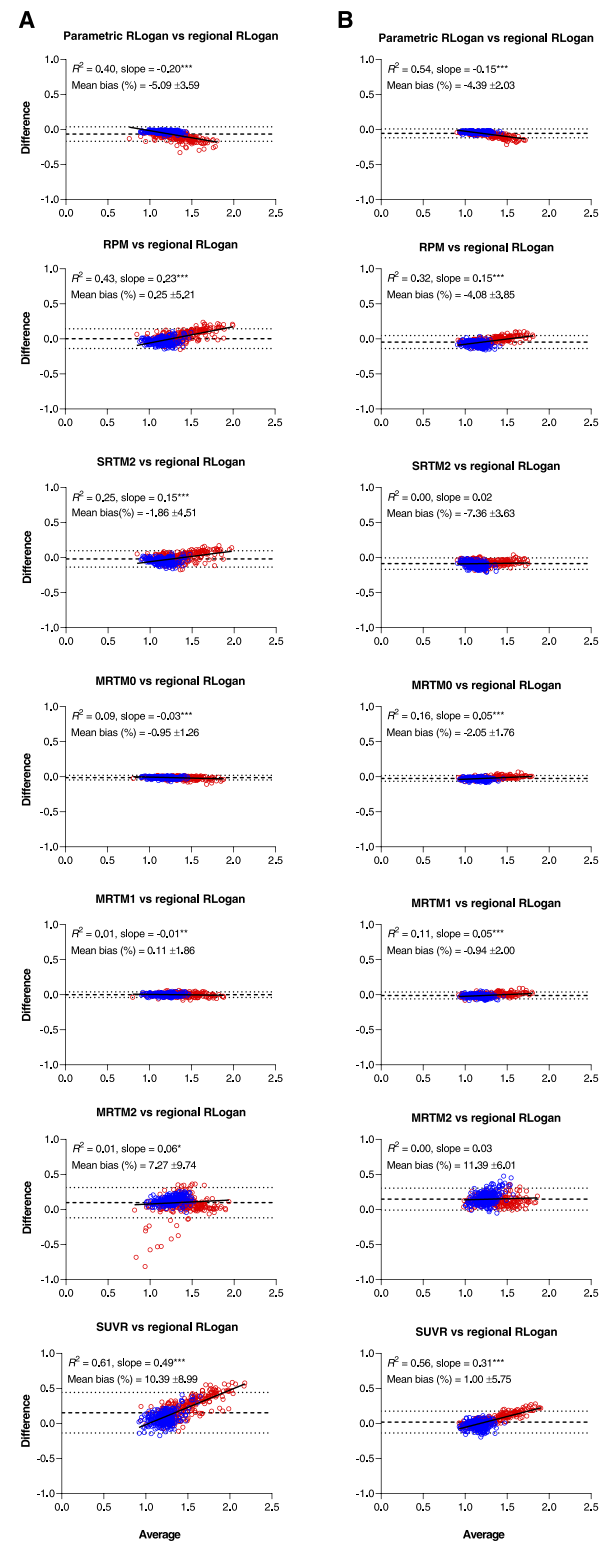


Fig. 3. Bland-Altman plots showing the agreement between outcome measures of all parametric methods and the regional standard for [18 F]flutemetamol and [18 F]florbetaben.

Bland-Altman plots for A) [18 F]flutemetamol and B) [18 F]florbetaben, showing the absolute difference between outcome measure of each parametric method and regional RLogan DVR for all regions of interest (ROIs). ROIs are colour-coded based on the A β status of the scan they belong to, i.e. A β -positive = red, A β -negative = blue. The dashed horizontal line corresponds to the average bias, the dotted horizontal lines correspond to the upper and lower limit of the 95% Limits of Agreement for all scans and the solid line to the linear regression of the BA data-points. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

a larger and/or proportional bias to a certain degree. Finally, lowest variability, as shown by the Limits of Agreement of the BA analyses (Fig. 3), was observed for MRTM0 and MRTM1, followed by parametric RLogan, SRTM2 and RPM, in case of [^{18}F]florbetaben. Therefore, MRTM0 and MRTM1 appear to be the most suitable methods with respect to accuracy and precision.

Of the preferred methods from the first and second aspect, RPM and SRTM2 have the additional benefit of providing a measure of relative perfusion (R_1) compared with the MRTM implementations. In the present study, R_1 yielded only low AUC values for differentiating between $A\beta$ -positive and $A\beta$ -negative scans. However, this finding was expected, as a cognitively unimpaired population was studied and changes in neuronal function and CBF tend to manifest at a later stage in the disease than $A\beta$ accumulation (Jack et al., 2010).

A limitation of this work is the lack of a direct comparison with the gold standard, plasma input modelling, to determine quantitative accuracy and precision, which could be done by future studies. No large differences would be expected, as several studies have shown that reference tissue approaches are adequate for quantification compared with the (plasma input) gold standard (Verfaillie et al., 2021; Heurling et al., 2015; Price et al., 2005). In addition, good results have been reported previously for RLogan in terms of correlation and bias compared with the gold standard and its robustness against confounding factors (Heeman et al., 2021; Nelissen et al., 2009). Furthermore, the performance of the parametric methods was evaluated using cognitively unimpaired subjects only. To cover a range of $A\beta$ burden, 50% $A\beta$ -positive and 50% $A\beta$ -negative subjects were selected. Nonetheless, the upper end of this range may be slightly lower than what is typically observed in AD dementia patients. It should be noted that, as the subjects scanned according to a dual-time window protocol will unlikely include patients with severe cognitive impairment (given that this is not the target population for clinical trials or studies measuring disease progression), the present range of $A\beta$ burden was considered appropriate for the goal of this study.

Conclusions

The preferred parametric methods for voxel-based amyloid quantification of [^{18}F]flutemetamol and [^{18}F]florbetaben dual-time window studies differed per evaluated aspect, but were relatively comparable between tracers. Compared with the reference standard method, regional RLogan, the best compromise across aspects was found for MRTM0, followed closely by SRTM2 which has the advantage of also providing R_1 . Given the current interest in R_1 , SRTM2 would be the preferred parametric method.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this project received funding from the EU/EFPIA Innovative Medicines Initiative (IMI) Joint Undertaking (EMIF grant 115372) and the EU-EFPIA IMI-2 Joint Undertaking (grant 115952). This joint undertaking receives support from the European Union's Horizon 2020 research and innovation program 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. FB is supported by the NIHR biomedical research centre at UCLH. Juan D. Gispert holds a 'Ramón y Cajal' fellowship (RYC-2013-13054) from the Spanish Ministry of Science, Innovation and Universities. The funding sources were not involved in any of the following aspects of the present study: design, data collection, analyses or interpretation, writing of the report, decision to submit the article for publication.

Author contribution statement

Fiona Heeman: Conceptualization, Methodology, Formal analysis, Writing – Original Draft, Writing – Review & Editing, Visualisation. **Maqsood Yaqub:** Conceptualization, Methodology, Software, Investigation, Writing – Review & Editing, Supervision. **Janine Hendriks:** Methodology, Formal analysis, Writing – Review & Editing. **Ilona Bader:** Investigation, Resources, Writing – Review & Editing. **Frederik Barkhof:** Methodology, Writing – Review & Editing, Funding acquisition, Supervision. **Juan Domingo Gispert:** Conceptualization, Writing – Review & Editing. **Bart N.M. van Berckel:** Investigation, Writing – Review & Editing. **Isadora Lopes Alves:** Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project administration. **Adriaan A. Lammertsma:** Conceptualization, Methodology, Writing – Review & Editing, Supervision.

Availability of software

In-house software that is currently not openly available was used in this study, however, similar analyses can be done using publicly available kinetic modelling software packages. Expressions of interest for collaboration and specific projects will be considered on a case-by-case basis.

Availability of data and materials

Data used in this study are not openly available at this stage due to ongoing data collection in the study. However, these will be made available to the research community upon end of the project (AMYPAD Prognostic and Natural History Study).

Declaration of Competing Interest

None.

Acknowledgements

The authors would like to thank Ronald Boellaard for his technical assistance, Lyduine Collij for coordinating collection of the scans and the staff of the department of Radiology and Nuclear Medicine of the Amsterdam UMC, location VUmc for skilful acquisition of the scans.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2021.117953](https://doi.org/10.1016/j.neuroimage.2021.117953).

References

- Becker, G.A., Ichise, M., Barthel, H., et al., 2013. PET quantification of 18F-florbetaben binding to β -amyloid deposits in human brains. *J. Nucl. Med.* 54, 723–731.
- Berckel BNM van, Ossenkoppele, R., Tolboom, N., et al., 2013. Longitudinal amyloid imaging using 11C-PiB: methodologic considerations. *J. Nucl. Med.* 54, 1570–1576.
- Binnewijzend, M.A.A., Benedictus, M.R., Kuijter, J.P.A., et al., 2016. Cerebral perfusion in the predementia stages of Alzheimer's disease. *Eur. Radiol.* 26, 506–514.
- Boellaard, R., Yaqub, M., Lubberink, M., et al., 2006. PPET: a software tool for kinetic and parametric analyses of dynamic PET studies. *Neuroimage* 31 (Supplement 2), T62.
- Bullich, S., Barthel, H., Koglin, N., et al., 2018. Validation of noninvasive tracer kinetic analysis of 18F-Florbetaben PET using a dual-time-window acquisition protocol. *J. Nucl. Med.* 59, 1104–1110.
- Collij, L.E., Heeman, F., Kuijter, J.P.A., et al., 2016. Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. *Radiology* 281, 865–875.
- Collij, L.E., Heeman, F., Salvadó, G., et al., 2020. Multi-tracer model for staging cortical amyloid deposition using PET imaging. *Neurology* doi:[10.1212/WNL.0000000000010256](https://doi.org/10.1212/WNL.0000000000010256).
- Daerr, S., Brendel, M., Zach, C., et al., 2016. Evaluation of early-phase [18F]-florbetaben PET acquisition in clinical routine cases. *Neuroimage Clin* 14, 77–86.
- Farrar, G., Barkhof, F., Pla, S., et al., 2017. AMYPAD: a European public-private partnership to investigate the value of β -amyloid brain scans as a diagnostic and therapeutic marker for Alzheimer's disease. *Alzheimer's Dement.: J. Alzheimer's Assoc.* 13, P420.

- Gunn, R.N., Lammertsma, A.A., Hume, S.P., et al., 1997. Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *Neuroimage* 6, 279–287.
- Hammers, A., Allom, R., Koepp, M.J., et al., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* 19, 224–247.
- Heeman, F., Yaqub, M., Alves, I.L., et al., 2021. Simulating the effect of cerebral blood flow changes on regional quantification of [18F]flutemetamol and [18F]florbetaben studies. *J. Cerebr. Blood Flow Metab.* 41, 579–589. doi:10.1177/0271678x20918029.
- Heeman, F., Yaqub, M., Lopes Alves, I., et al., 2019. Optimized dual-time-window protocols for quantitative [18F]flutemetamol and [18F]florbetaben PET studies. *EJNMMI Res.* 9, 32.
- Heeman, F., Hendriks, J., Lopes Alves, I., et al., 2020. [11C]PIB amyloid quantification: effect of reference region selection. *EJNMMI Res.* 10, 123.
- Heurling, K., Buckley, C., Van Laere, K., et al., 2015. Parametric imaging and quantitative analysis of the PET amyloid ligand [(18F)]flutemetamol. *Neuroimage* 121, 184–192.
- Hu, Z., Wang, W., Gualtieri, E.E., et al., 2007. An LOR-based fully-3D PET image reconstruction using a blob-basis function. In: *Proceedings of the 2007 IEEE Nuclear Science Symposium Conference Record*, pp. 4415–4418.
- Hu, Z., Renisch, S., Schweizer, B., et al., 2010. MR-based attenuation correction for whole-body PET/MR system. In: *Proceedings of the IEEE Nuclear Science Symposium Medical Imaging Conference*, pp. 2119–2122.
- Ichise, M., Liow, J.-S., Lu, J.-Q., et al., 2003. Linearized reference tissue parametric imaging methods: application to [11C]DASB positron emission tomography studies of the serotonin transporter in human brain. *J. Cerebr. Blood Flow Metab.* 23, 1096–1112.
- Jack, C.R., Knopman, D.S., Jagust, W.J., et al., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* 9, 119–128.
- Jagust, W.J., Eberling, J.L., Reed, B.R., et al., 1997. Clinical studies of cerebral blood flow in Alzheimer's disease. *Ann. N. Y. Acad. Sci.* 826, 254–262.
- Lammertsma, A.A., Hume, S.P., 1996. Simplified reference tissue model for PET receptor studies. *Neuroimage* 4, 153–158.
- Lammertsma, A.A., 2017. Forward to the past: the case for quantitative PET imaging. *J. Nucl. Med.* 58, 1019–1024.
- Logan, J., Fowler, J.S., Volkow, N.D., et al., 1996. Distribution volume ratios without blood sampling from graphical analysis of PET data. *J. Cerebr. Blood Flow Metab.* 16, 834–840.
- Lopes Alves, I., Collij, L.E., Altomare, D., et al., 2020. Quantitative amyloid PET in Alzheimer's disease: the AMYPAD prognostic and natural history study. *Alzheimer's Dement* doi:10.1002/alz.12069, Epub ahead of print 12 April.
- Mallik, A., Drzezga, A., Minoshima, S., 2017. Clinical Amyloid imaging. *Semin. Nucl. Med.* 47, 31–43.
- Nelissen, N., Van Laere, K., Thurfjell, L., et al., 2009. Phase 1 study of the Pittsburgh Compound B derivative 18F-Flutemetamol in healthy volunteers and patients with probable Alzheimer disease. *J. Nucl. Med.* 50, 1251–1259.
- Ottoy, J., Verhaeghe, J., Niemantsverdriet, E., et al., 2017. Validation of the semiquantitative static SUVR method for 18F-AV45 PET by pharmacokinetic modeling with an arterial input function. *J. Nucl. Med.* 58, 1483–1489.
- Price, J.C., Klunk, W.E., Lopresti, B.J., et al., 2005. Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh compound-B. *J. Cerebr. Blood Flow Metab.* 25, 1528–1547.
- Rask, T., Dyrby, T., Comerci, M., et al., 2004. PVElab: software for correction of functional images for partial volume errors. *Neuroimage* 22 Supplement.
- Salinas, C.A., Searle, G.E., Gunn, R.N., 2015. The simplified reference tissue model: model assumption violations and their impact on binding potential. *J. Cerebr. Blood Flow Metab.* 35, 304–311.
- Slifstein, M., Laruelle, M., 2000. Effects of statistical noise on graphic analysis of PET neuroreceptor studies. *J. Nucl. Med.* 41, 2083–2088.
- Son, S.H., Kang, K., Ko, P.-W., et al., 2020. Early-phase 18F-Florbetaben PET as an alternative modality for 18F-FDG PET. *Clin. Nucl. Med.* 45, e8.
- Tiepol, S., Hesse, S., Patt, M., et al., 2016. Early [18F]florbetaben and [11C]PiB PET images are a surrogate biomarker of neuronal injury in Alzheimer's disease. *Eur. J. Nucl. Med. Mol. Imaging* 43, 1700–1709.
- Verfaillie, S.C., Golla, S.S., Timmers, T., et al., 2021. Repeatability of parametric methods for [18F]florbetapir imaging in Alzheimer's disease and healthy controls: a test–retest study. *J. Cerebr. Blood Flow Metab.* 41, 569–578. doi:10.1177/0271678x20915403.
- Wu, Y., Carson, R.E., 2002. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. *J. Cerebr. Blood Flow Metab.* 22, 1440–1452.
- Yaqub, M., Tolboom, N., Boellaard, R., et al., 2008. Simplified parametric methods for [11C]PIB studies. *Neuroimage* 42, 76–86.