



# Test–retest repeatability of [<sup>18</sup>F]Flortaucipir PET in Alzheimer’s disease and cognitively normal individuals

Tessa Timmers<sup>1,2</sup>, Rik Ossenkoppele<sup>1,3</sup>, Denise Visser<sup>2</sup>, Hayel Tuncel<sup>2</sup>, Emma E Wolters<sup>1,2</sup>, Sander CJ Verfaillie<sup>2</sup> , Wiesje M van der Flier<sup>1,4</sup>, Ronald Boellaard<sup>2</sup>, Sandeep SV Golla<sup>2</sup> and Bart NM van Berckel<sup>2</sup>

## Abstract

The aim of this study was to investigate the test–retest (TRT) repeatability of various parametric quantification methods for [<sup>18</sup>F]Flortaucipir positron emission tomography (PET). We included eight subjects with dementia or mild cognitive impairment due to Alzheimer’s disease and six cognitively normal subjects. All underwent two 130-min dynamic [<sup>18</sup>F]Flortaucipir PET scans within  $3 \pm 1$  weeks. Data were analyzed using reference region models receptor parametric mapping (RPM), simplified reference tissue method 2 (SRTM2) and reference logan (RLogan), as well as standardized uptake value ratios (SUV<sub>r</sub>, time intervals 40–60, 80–100 and 110–130 min post-injection) with cerebellar gray matter as reference region. We obtained distribution volume ratio or SUV<sub>r</sub>, first for all brain regions and then in three tau-specific regions-of-interest (ROIs). TRT repeatability (%) was defined as  $|\text{retest} - \text{test}| / (\text{average}(\text{test} + \text{retest})) \times 100$ . For all methods and across ROIs, TRT repeatability ranged from (median (IQR)) 0.84% (0.68–2.15) to 6.84% (2.99–11.50). TRT repeatability was good for all reference methods used, although semi-quantitative models (i.e. SUV<sub>r</sub>) performed marginally worse than quantitative models, for instance TRT repeatability of RPM: 1.98% (0.78–3.58) vs. SUV<sub>r80–100</sub>: 3.05% (1.28–5.52),  $p < 0.001$ . Furthermore, for SUV<sub>r80–100</sub> and SUV<sub>r110–130</sub>, with higher average SUV<sub>r</sub>, more variation was observed. In conclusion, while TRT repeatability was good for all models used, quantitative methods performed slightly better than semi-quantitative methods.

## Keywords

[<sup>18</sup>F]Flortaucipir, test–retest repeatability, parametric methods, Alzheimer’s disease

Received 22 March 2019; Revised 18 July 2019; Accepted 8 August 2019

## Introduction

Abnormal accumulation of tau proteins in the brain is one of the characteristics of Alzheimer’s disease (AD).<sup>1,2</sup> It recently became possible to visualize and quantify tau pathology in vivo using positron emission tomography (PET).<sup>3,4</sup> To date, the most widely used PET tracer for tau pathology is [<sup>18</sup>F]Flortaucipir (formerly known as [<sup>18</sup>F]AV1451 or [<sup>18</sup>F]T807).

[<sup>18</sup>F]Flortaucipir PET scans have predominantly been evaluated with standardized uptake value ratios (SUV<sub>r</sub>s). Although this method has certain advantages, such as a short scan duration and computational simplicity, some important disadvantages should be taken

<sup>1</sup>Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

<sup>2</sup>Department of Radiology & Nuclear Medicine, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

<sup>3</sup>Clinical Memory Research Unit, Lund University, Lund, Sweden

<sup>4</sup>Department of Epidemiology and Biostatistics, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

## Corresponding author:

Tessa Timmers, Amsterdam UMC, location VUmc, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands.  
Email: t.timmers@vumc.nl

into account.<sup>5–9</sup> Because SUVr shows the normalized uptake over a specified time interval, it is dependent on regional tracer delivery (which in turn is dependent on blood flow), retention, and clearance rates.<sup>5</sup> In contrast, kinetic models, derived from dynamic scans with arterial sampling, allow computation of these factors and allow to disentangle the specific binding of the tracer to the target under investigation. Since arterial cannulation is an invasive procedure and arterial plasma derived input functions are susceptible to noise, methods using reference regions<sup>7,9</sup> are practically more feasible. Among these, quantitative reference region methods are the basis function approaches and parametric implementations of SRTM<sup>10</sup> receptor parametric mapping (RPM)<sup>11</sup> and simplified reference tissue model 2 (SRTM2<sup>12</sup>), and the linearization approach reference Logan (RLogan).<sup>13</sup>

To assess the reliability of *in vivo* kinetics of a PET tracer and quantification of (specific) tracer binding, test–retest (TRT) studies must be performed. Repeatability is especially of importance for longitudinal studies and when assessing treatment effects in clinical trials.<sup>6</sup> One previous study investigated repeatability of [<sup>18</sup>F]Flortaucipir, and reported a mean change ranging from 1.8% to 3.3% between test and retest SUV<sub>r80–100</sub> in a cortical composite region-of-interest (ROI) using a white matter-based reference region.<sup>14</sup> This study performed analyses using semi-quantitative measures only, thus the repeatability of quantitative parametric methods remains to be elucidated. The aim of the present study was therefore to investigate the TRT repeatability of various parametric reference region tissue methods for [<sup>18</sup>F]Flortaucipir PET in patients with AD and controls. We hypothesized that methods using dynamic scan data (e.g. RPM, SRTM2 and RLogan) would have better TRT repeatability than methods using static scan data (e.g. SUVr).

## Materials and methods

### Participants

We included 14 participants; 8 patients with mild cognitive impairment (MCI) or dementia due to AD and 6 cognitively normal subjects. MCI/AD patients were recruited from the Amsterdam Dementia Cohort.<sup>15</sup> All patients visited the memory clinic of the Amsterdam Alzheimer Center for a standardized dementia screening, including brain imaging and an extensive neuropsychological assessment. Diagnoses were established in a multidisciplinary meeting. MCI/AD participants were eligible if they met criteria for MCI due to AD<sup>16</sup> or AD dementia,<sup>17</sup> had positive amyloid- $\beta$  biomarkers (cerebrospinal fluid A $\beta$ <sub>42</sub> < 813 pg/

mL<sup>18</sup> and/or abnormal amyloid PET (<sup>11</sup>C-PiB  $n=2$ , <sup>18</sup>F-florbetaben  $n=1$ ) on visual read<sup>19</sup>) and had a Mini Mental State Examination (MMSE) score of 18 or higher. Cognitively normal subjects were recruited through local advertisements. Cognitively unimpaired participants were screened using MMSE and MRI. They were eligible for the study when MMSE score was  $\geq 27/30$  and no extensive atrophy was seen on MRI. Exclusion criteria for all participants were structural masses on MRI that were likely to interfere with segmentation, a history of severe traumatic brain injury and participation in a drug trial. The study protocol was approved by the Medical Ethics Committee of the Amsterdam University Medical Center, location VUmc, according to the Dutch law on medical scientific research on humans (Wet medisch-wetenschappelijk onderzoek met mensen). All participants signed an informed consent form.

### Image acquisition

All participants underwent one structural whole brain MRI and two [<sup>18</sup>F]Flortaucipir PET scans. The time interval between two [<sup>18</sup>F]Flortaucipir scans had a minimum of one week, and a maximum of four weeks. The maximum time interval between MRI and first [<sup>18</sup>F]Flortaucipir scan was six months. MRI scans were acquired on a 3.0T Philips Ingenuity Time-of-Flight PET/MR scanner. We obtained isotropic structural 3D T1-weighted images using a sagittal turbo field echo sequence (1.00 mm<sup>3</sup> isotropic voxels, repetition time = 7.9 ms, echo time = 4.5 ms, flip angle = 8°). Test and retest [<sup>18</sup>F]Flortaucipir PET scans were acquired on a Philips Ingenuity TF PET/CT scanner with identical procedures. Individual doses of [<sup>18</sup>F]Flortaucipir PET were prepared on site according to Avid Radiopharmaceuticals quality control criteria. Scan procedures started with a low-dose CT for attenuation correction. Subsequently, [<sup>18</sup>F]Flortaucipir (237 ± 15 MBq at test and 245 ± 18 MBq at retest): was injected simultaneously with the start of the first 60-min dynamic emission scan (specific activity 919 ± 283 MBq/μg at test and 789 ± 310 MBq/μg at retest). After a 20-min break, the low-dose CT was repeated and a second PET scan was acquired from 80 to 130 min post-injection. During these scans, head movements were restricted with the use of head bands. Movement was checked regularly using laser beams, and head position was corrected if necessary. All scans underwent quality control for substantial movement artefacts. The second scan session was co-registered to the first using Vinci software (Max Plank Institute, Cologne, Germany), and combined into a dataset of 29 frames, with a matrix size of 128 × 128 × 90 and a final voxel size of 2 × 2 × 2 mm<sup>3</sup>.

### Image analysis

T1 weighted MR images were co-registered to PET using Vinci software. We used the Hammers template,<sup>20</sup> incorporated in PVElab software,<sup>21</sup> to delineate gray matter ROIs on the co-registered MR images. Next, we generated parametric images using in-house built software (PPET)<sup>22</sup> and the cerebellar gray matter as reference region. We created parametric images using the linearization approach RLogan,<sup>13</sup> the basis function approaches RPM<sup>11</sup> and STRM2.<sup>12</sup> For RLogan, the time interval 40–130 min post-injection was used to generate distribution volume ratio (DVR) images. For RPM and SRTM2, 0–130 min data were used, with 30 basis functions with a range from 0.01 to 0.1<sup>23</sup> to generate non-displaceable binding potential (BP<sub>ND</sub>) images. Finally, we generated SUV<sub>r</sub> images for different time intervals (i.e. 40–60, 80–100, 110–130 min post-injection). For analyses on a whole-brain level, we used all volume-weighted bilateral regions extracted from the Hammers template, with the exception of cerebellum, brainstem, corpus callosum, and ventricles. In addition, Hammers template ROIs were combined to obtain tau-specific ROIs, reflecting early, intermediate and late stages of tau deposition:<sup>24</sup> medial temporal lobe (MTL; hippocampus, amygdala, and parahippocampus), lateral temporal lobe (LTL; anterior lateral temporal lobe, superior temporal lobe, middle & inferior temporal lobe and posterior temporal lobe), and a global measure (hippocampus, amygdala, parahippocampal and ambient gyri, anterior temporal lobe lateral part, superior temporal gyrus, middle and inferior temporal gyri, posterior temporal lobe, gyrus cingula posterior part, superior parietal gyrus, inferolateral remainder of parietal lobe, gyrus cinguli anterior part, middle frontal gyrus, orbito-frontal gyri, inferior frontal gyrus, superior frontal gyrus, lateral remainder of occipital lobe, lingual gyrus and cuneus).

In addition, to test whether our results would translate to models with another reference region, we used a white matter reference region to create parametric images. To this end, we used the T1 weighted scans which were transformed to PET space. With the use of the Hammers template, we extracted a white matter reference region. This region was eroded, to minimize the spill-in from adjacent gray matter regions. We used PPET software and eroded white matter as a reference region to create parametric images using RLogan, RPM, SRTM2, SUV<sub>r</sub><sub>40–60</sub>, SUV<sub>r</sub><sub>80–100</sub>, and SUV<sub>r</sub><sub>110–130</sub>. Similar to the GM reference region analysis, we combined Hammers ROIs to create tau-specific ROIs.

### Statistical analysis

TRT repeatability (%) was defined as  $|retest - test| / ((average (test + retest)) \times 100)$ . For RPM and STRM2, TRT repeatability was assessed using BP<sub>ND</sub> + 1, corresponding to DVR, in order to directly compare values to RLogan and SUV<sub>r</sub> methods. To compare TRT repeatability per reference region method (RLogan, RPM, SRTM2, SUV<sub>r</sub> for 3 time intervals) and to compare TRT repeatability per tau-specific region (MTL, LTL and global ROI), we used Kruskal–Wallis tests and post-hoc Mann–Whitney *U* testing with Bonferroni correction. To compare TRT repeatability between AD and cognitively normal subjects and between cortical and subcortical regions, we used Mann–Whitney *U* tests to account for non-normally distributed data. To compare the TRT repeatability between ROIs based on cerebellar gray versus white matter as reference region, we used Mann–Whitney *U* testing. We considered a *p* value <0.05 as significant.

For exploratory analyses, we calculated sample sizes using GPower v3.1. For these analyses, we used different values for expected change over time (ranging from 0.5% to 3%), to inform on longitudinal study designs. We determined sample sizes for the two most widely used quantitative (RPM) and semi-quantitative (SUV<sub>r</sub><sub>80–100</sub>) methods. We calculated the differences between two dependent means (matched pairs), with an  $\alpha$  (error probability) of 0.05 and a power (1- $\beta$  error probability) of 0.80. We used the SD of the relative percentage change  $((retest - test) / (average (test + retest)) \times 100)$ .

## Results

### Demographics

Demographics according to diagnostic group are provided in Table 1. Participants were relatively young, with a mean age of 65 ± 11 years in MCI/AD and

**Table 1.** Demographics.

	Cognitively normal subjects <i>n</i> = 6	MCI/AD <i>n</i> = 8
Age	64.9 ± 9.4	65.5 ± 10.5
Sex (M/F)	4/4	3/3
MMSE	28.5 ± 0.6	24.0 ± 3.0
Amyloid- $\beta$ status	2 A $\beta$ negative, 4 unknown	All A $\beta$ positive
Time lag between PET scans (days)	23.5 ± 7.4	19.3 ± 6.2

64 ± 9 years in cognitively normal subjects. By study design, all MCI/AD participants were amyloid-β positive. Two cognitively normal subjects were amyloid-β negative, whereas for the other cognitively normal subjects amyloid-β status was unknown. Time between consecutive PET scans was 21 ± 7 days.

### TRT repeatability across all Hammers ROIs

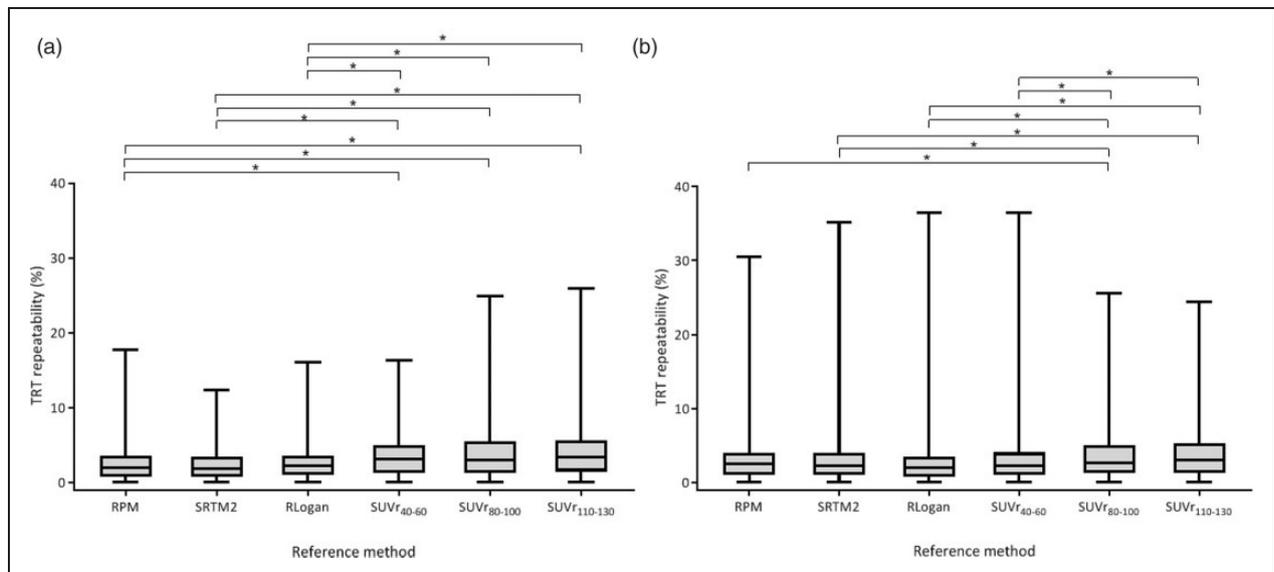
TRT repeatability ranged from (median (IQR)) 0.85% (0.15–2.22) to 3.88% (1.44–5.35) for RPM, 0.84% (0.68–2.15) to 3.95% (2.03–5.81) for SRTM2, 0.61% (0.23–2.63) to 3.61% (1.88–7.19) for RLogan, 1.36% (0.57–3.93) to 6.70% (1.38–9.20) for SUV<sub>r40–60</sub>, 0.97% (0.29–5.07) to 6.29% (3.16–10.23) for SUV<sub>r80–100</sub> and to 1.09% (0.33–2.51) to 6.84% (2.99–11.50) for SUV<sub>r110–130</sub> (Supplementary Table 1). TRT repeatability differed significantly between methods ( $H(5) = 86.3$ ,  $p < 0.0001$ ). Post-hoc analyses showed that the quantitative methods RPM, SRTM2, and RLogan performed better than semi-quantitative SUV<sub>r</sub> (for instance RPM: 1.98% (0.78–3.58) vs. SUV<sub>r80–100</sub>: 3.05% (1.28–5.52),  $p < 0.001$ ; Figure 1(a)). Moreover, for SUV<sub>r</sub>, more variation was observed than for quantitative methods. We observed no differences in TRT repeatability between RPM, SRTM2 and RLogan, nor in the TRT repeatability of SUV<sub>r</sub> between different time intervals. MCI/AD patients had worse TRT repeatability than cognitively

normal subjects when using SUV<sub>r80–100</sub> (3.76% (1.33–6.29) vs. 2.48% (1.21–4.21),  $p < 0.001$ ) or SUV<sub>r110–130</sub> (4.10% (2.19–6.28) vs. 2.62% (1.12–5.01),  $p < 0.001$ ), but not when using other reference region methods.

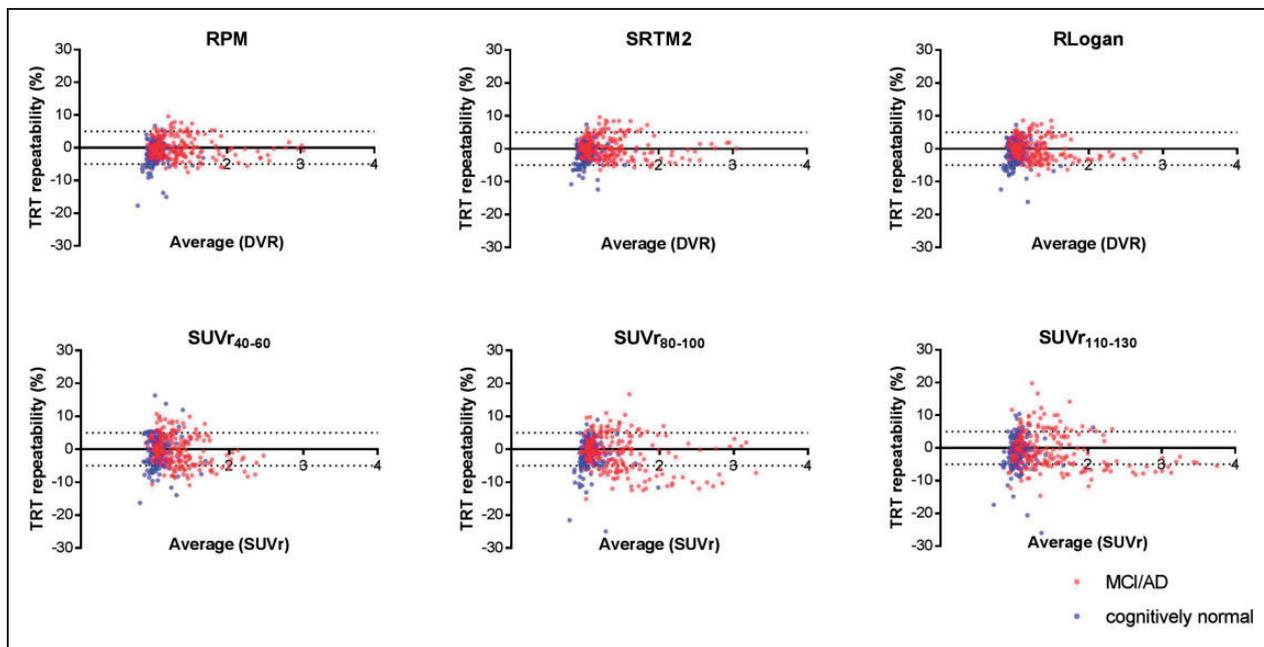
Figure 2 displays Bland Altman plots for cognitively normal subjects and AD participants, for the different reference region methods under investigation. For all methods, most data points fell within the 5% range, suggesting good repeatability. However, for SUV<sub>r</sub> at all three time intervals, more variation was observed than for RPM and SRTM2. For SUV<sub>r80–100</sub> and SUV<sub>r110–130</sub>, higher SUV<sub>r</sub> values were associated with greater variance. TRT repeatability was worse for ROIs with volumes in the lowest quartile (<2.82 cc; Supplementary Figure 1). TRT repeatability was worse in subcortical regions (e.g. nucleus caudatus, nucleus accumbens, putamen, thalamus and pallidum) compared to cortical regions (3.41% (1.48–5.91) vs. 2.36% (0.99–4.23),  $p < 0.001$ ).

### TRT repeatability in tau-specific ROIs

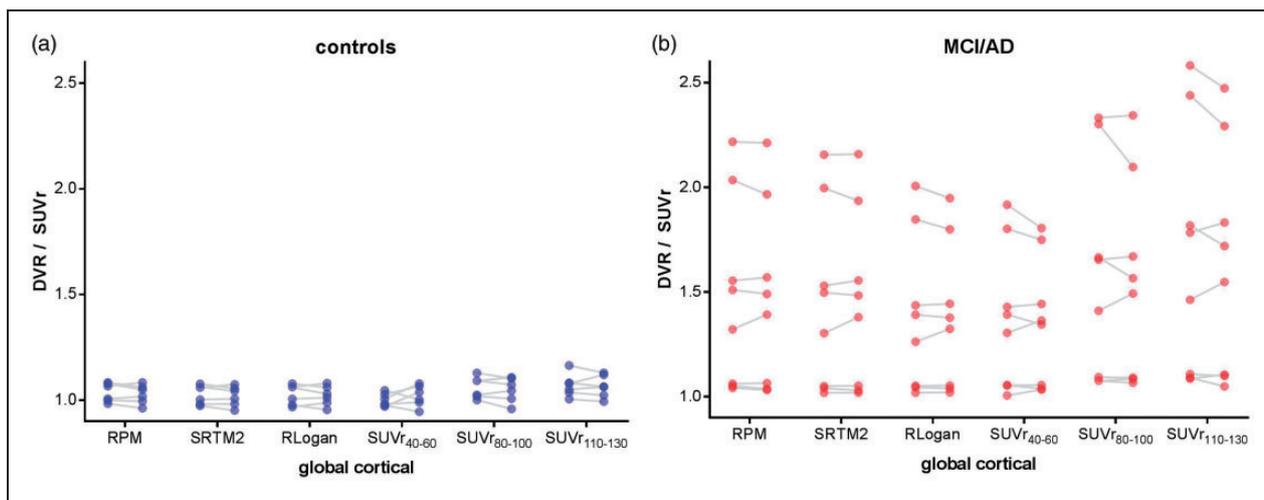
Supplementary Table 2 shows the parameter values of the total sample for the various parametric reference region methods for three ROIs reflecting early, intermediate, and late tau deposition. Parameter values for all parametric reference region methods at test and retest in the global region are displayed in Figure 3. Across all methods, differences between test and



**Figure 1.** Box and whiskers representing TRT repeatability for different reference region methods using cerebellar gray matter (a) and white matter (b) as reference. Displayed are box and whisker plots (median, IQR, lowest and highest value) representing TRT repeatability ( $100 \times |\text{retest} - \text{test}| / (0.5 \times (\text{test} + \text{retest}))$ ), y-axis for different reference region methods (x-axis) for all bilateral Hammers regions (excluding cerebellum, brainstem, corpus callosum and ventricles). Statistical significant differences between reference region methods are denoted with \* ( $p < 0.001$  (Mann–Whitney  $U$ , Bonferroni corrected)). Box (a) represents data obtained with cerebellar gray matter as a reference region. Box (b) represents data obtained with white matter as a reference region.



**Figure 2.** Bland Altman plots for all Hammers regions, using different reference region methods. Displayed are results for cognitively unimpaired subjects (blue) and MCI/AD participants (red) for bilateral Hammers ROIs, excluding cerebellum, brainstem, corpus callosum, and ventricles. TRT repeatability (y-axis) is represented as  $100 \times (\text{retest} - \text{test}) / (0.5 \times (\text{test} + \text{retest}))$ . The dotted lines represent an artificial boundary of 5%, which is considered as low variation and therefore good repeatability.



**Figure 3.** Test and retest values for various parametric methods the global ROI. Represented are individual DVR/SUVR values at test (left dot) and retest (right dot) in the global ROI for cognitively normal subjects (a) and MCI/AD (b). Test and retest values for each individual subject are connected by a line. For RPM, SRTM2, and RLogan, DVR values are presented.

retest values were small. SUVR values showed more variation between test and retest than the quantitative measures.

Table 2 displays the absolute TRT repeatability (%) in three tau specific ROIs in the total sample. Overall, TRT repeatability analysis yielded low values, ranging from (median (IQR)) 0.7% (0.4–5.2) to 4.3% (0.7–5.5). TRT repeatability did not differ by ROI (for instance:

for RPM  $H(2) = 2.2, p = 0.20$ ). Again, differences were observed between reference region methods ( $H(5) = 14.02, p < 0.05$ ). However, when we post-hoc compared the different reference region methods separately using Mann–Whitney  $U$  tests, differences did not survive correction for multiple testing. When we stratified for diagnosis, TRT repeatability in three AD-specific regions ranged from (median (IQR)) 0.9% (0.4–2.0)

**Table 2.** Median (IQR) TRT repeatability (%).

	RLogan	RPM	SRTM2	SUV <sub>r40-60</sub>	SUV <sub>r80-100</sub>	SUV <sub>r110-130</sub>
<b>Total sample</b>						
Medial temporal lobe	2.7 (1.7–4.1)	3.0 (1.3–3.9)	2.8 (1.2–4.0)	3.1 (1.4–6.8)	3.5 (2.4–6.0)	4.3 (0.7–5.5)
Lateral temporal lobe	2.1 (0.4–3.2)	1.1 (0.3–3.3)	1.2 (0.3–3.4)	3.5 (0.9–3.9)	0.7 (0.4–5.2)	2.2 (1.3–5.2)
Global	1.7 (0.5–2.7)	1.4 (0.7–2.3)	1.2 (0.1–2.1)	2.7 (1.2–4.0)	1.6 (0.9–4.7)	2.8 (1.3–4.6)
<b>MCI/AD</b>						
Medial temporal lobe	3.0 (2.4–4.3)	3.2 (2.1–4.5)	3.0 (1.6–4.8)	5.9 (2.3–7.5)	4.3 (2.3–8.1)	5.3 (0.9–7.8)
Lateral temporal lobe	2.4 (0.3–3.4)	1.1 (0.5–4.2)	1.2 (0.3–4.3)	3.5 (1.5–5.3)	2.6 (0.5–6.9)	5.1 (1.6–6.7)
Global	0.8 (0.2–2.9)	1.2 (0.4–2.9)	1.0 (0.1–2.7)	2.9 (1.1–4.2)	0.9 (0.6–6.0)	4.1 (2.0–5.6)
<b>Cognitively normal</b>						
Medial temporal lobe	1.9 (1.3–3.4)	1.8 (0.4–3.8)	2.4 (0.9–3.8)	1.8 (1.2–3.4)	3.1 (2.0–4.4)	3.5 (0.6–4.5)
Lateral temporal lobe	1.6 (0.7–2.7)	1.1 (0.3–2.3)	0.9 (0.4–2.0)	2.4 (0.6–3.9)	0.4 (0.1–3.0)	1.6 (0.4–2.6)
Global	1.9 (1.4–2.5)	1.9 (0.9–2.3)	1.6 (0.3–2.1)	1.9 (1.3–4.0)	1.9 (1.4–2.7)	1.5 (1.2–3.2)

TRT: test–retest; IQR: interquartile range; RLogan: reference logan; RPM: receptor parametric mapping; SRTM2: simplified reference tissue method 2; MCI: mild cognitive impairment; AD: Alzheimer's disease. TRT repeatability is defined as  $|retest - test| / (average (test + retest)) \times 100$ .

to 3.5% (0.6–4.5) in cognitively normal subjects and 0.8% (0.2–2.9) to 5.9% (2.3–7.5) in AD. For SUV<sub>r110–130</sub>, differences in TRT repeatability between AD and cognitively normal subjects were observed (5.06% (1.80–5.84) vs. 1.80% (1.05–3.56),  $p < 0.01$ ). This can also be denoted from Figure 4, which shows Bland Altman plots for three tau-specific ROIs for all reference region methods.

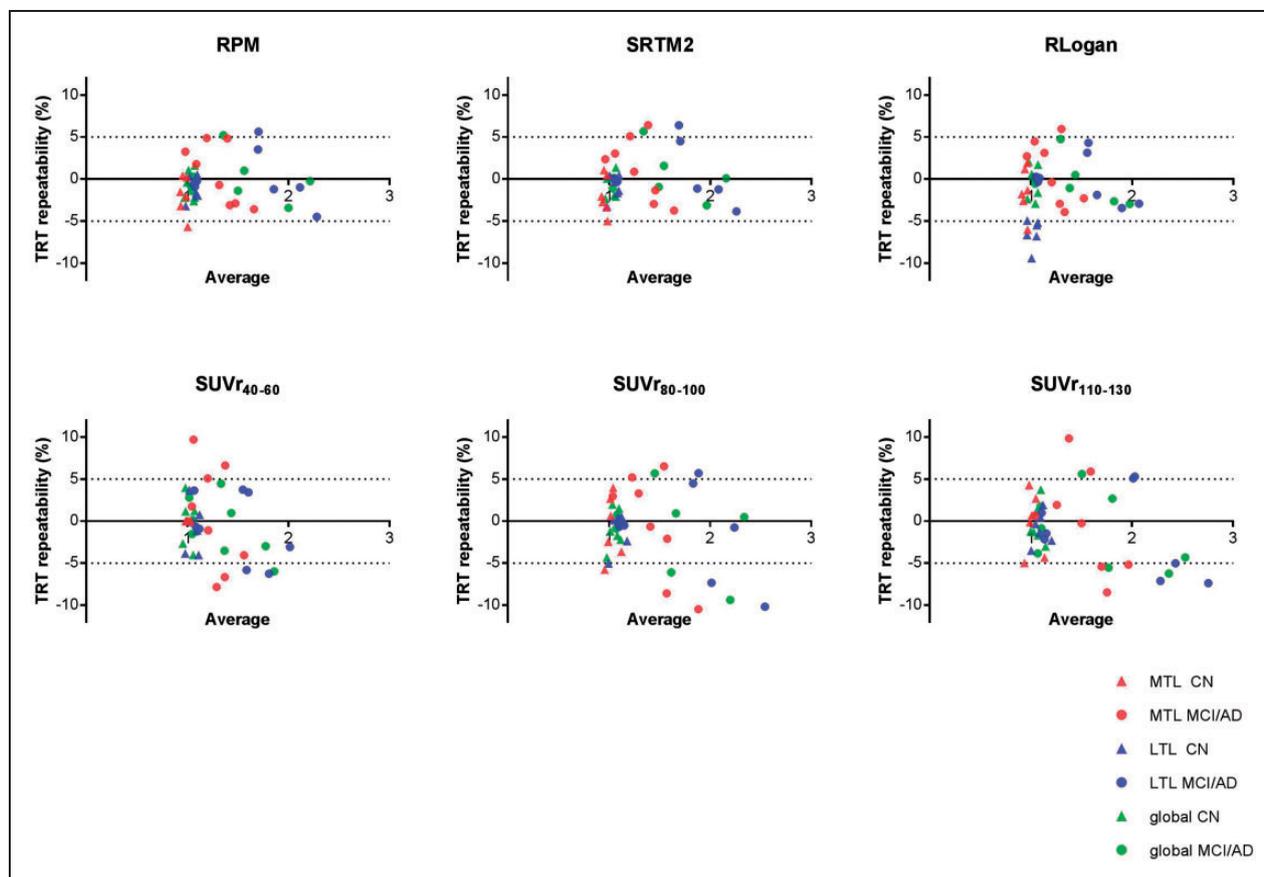
### White matter reference region

We performed additional analyses to assess whether TRT repeatability was consistent when using a different, i.e. white matter reference region. Parameter values for test and retest values for all six parametric reference methods when using white matter as a reference region are shown in Supplementary Table 3. Supplementary Table 4 displays the TRT repeatability for all parametric reference methods with white matter as reference region and Supplementary Figure 2 shows the corresponding Bland Altman plots. When using a white matter reference region, TRT repeatability yielded similarly low values as obtained when using cerebellar gray matter reference region. Figure 1(a) shows TRT repeatability using a cerebellar gray matter reference region, and Figure 1(b) shows TRT repeatability using a white matter reference region. When comparing TRT

repeatability between cerebellar gray matter and white matter for three tau-specific ROIs for all subjects, no differences in TRT repeatability for different reference methods were observed. There was one exception; TRT repeatability for MTL SUV<sub>r40–60</sub> using white matter reference region was lower than TRT repeatability using cerebellar gray matter (MTL SUV<sub>r40–60</sub> cerebellar gray matter reference: 3.12 (1.41–6.77) vs. MTL SUV<sub>r40–60</sub> white matter reference: 1.59 (0.55–3.12),  $p < 0.05$ ). When stratifying for diagnosis, this difference was only present for MCI/AD (MTL SUV<sub>r40–60</sub> cerebellar gray matter: 5.86 (2.32–7.54) vs. MTL SUV<sub>r40–60</sub> white matter reference: 1.41 (0.66–2.68),  $p < 0.05$ ) and not for controls. Similar to TRT repeatability for cerebellar gray matter, TRT repeatability with white matter as reference region differed significantly between methods ( $H(5) = 44.4$ ,  $p < 0.0001$ ). Post-hoc analyses showed that the quantitative methods RPM, SRTM2, and RLogan performed better than semi-quantitative SUV<sub>r80–100</sub> and SUV<sub>r110–130</sub> (for instance SRTM2: 2.26% (1.05–4.00) vs. SUV<sub>r80–100</sub>: 2.71% (1.24–5.08),  $p < 0.01$ ; Figure 1(b)).

### Sample size calculation

We performed sample size calculations to explore if the required sample size would differ between methods



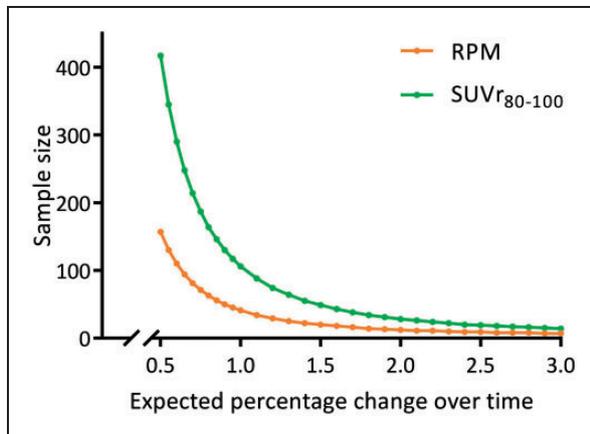
**Figure 4.** Bland Altman plots for all six reference region methods in three tau-specific regions-of-interest. Displayed are results for cognitively unimpaired subjects (CN, triangles) and MCI/AD participants (dots) for MTL (red), LTL (blue) and global ROI (green). TRT repeatability (y-axis) is represented as  $100 \times (\text{test} - \text{retest}) / (0.5 \times (\text{test} + \text{retest}))$ .

when minor changes (i.e. 0.5–3%) over time are expected. Figure 5 shows that with small changes in effect size, large differences in required sample sizes were observed. For instance, with an expected change over time of 0.5%, a sample size of 417 was needed when using  $\text{SUVR}_{80-100}$ , as opposed to 157 when using RPM (Figure 4). However, for greater expected changes over time the sample size calculations converged between methods.

## Discussion

This study evaluated TRT repeatability of [ $^{18}\text{F}$ ]Flortaucipir for various parametric reference region methods. We observed a good repeatability between test and retest for all reference region methods. Test–retest repeatability was comparable between tau-specific regions and all brain regions. However, small differences between quantitative methods, using dynamic scan data, and semi-quantitative methods obtained using static scan data were observed. For semi-quantitative methods, increased variation was seen with higher underlying tau load.

Based on these results one could argue that  $\text{SUVR}$  is a suitable method to quantify tau pathology with [ $^{18}\text{F}$ ]Flortaucipir due to its high correlation with quantitative measures obtained from full kinetic modeling,<sup>23,25,26</sup> and good test–retest properties. Taken into account its practical feasibility,  $\text{SUVR}$  might be acceptable for analyzing cross-sectional data. However, our data revealed two drawbacks of this method. First, for  $\text{SUVR}$ , worse TRT repeatability was seen with higher underlying tau load, which could have implications for the validity of the tracer, especially in AD patients. This is underscored by the finding that when using  $\text{SUVR}_{80-100}$  and  $\text{SUVR}_{110-130}$ , TRT repeatability was worse for AD patients than for cognitively normal subjects. Second, it should be noted that  $\text{SUVR}$  values in the present study are obtained from dynamic scans, allowing us to precisely set the  $\text{SUVR}$  uptake time interval. This is not a realistic scenario in clinical practice. A recent multi-center oncology study showed that the uptake time using a static imaging protocol varied significantly, with only ~30% of scans starting within 5 min of the intended time after injection.<sup>27</sup> Therefore, it is likely that the observed differences



**Figure 5.** Sample size calculations. Displayed are sample size calculations for RPM and  $SUVR_{80-10}$  (y-axis) for different effect sizes (i.e. expected difference between groups or change over time, x-axis).

between test and retest for  $SUVR$  in the present study are less pronounced than they will be in a situation where only static scans are used. This is specifically the case for  $[^{18}F]$ Flortaucipir, as this tracer does not reach steady state at 130 min post-injection,<sup>25</sup> meaning that small variations in  $SUVR$  time frames may result in large bias and poor repeatability.

For longitudinal data, semi-quantitative methods are not the most optimal reference region method.<sup>8,9,28</sup> As opposed to methods using dynamic scan data and quantitative approaches,  $SUVR$  does not take into account between-subject differences in tracer wash-in and wash-out.<sup>5</sup> The slightly higher variation in terms of test–retest repeatability for  $SUVR$  could partly be caused by day-to-day differences in flow<sup>29</sup> and thus in tracer delivery. In patients with AD, flow changes can occur with disease progression,<sup>30–32</sup> and these changes in flow may affect target and reference areas differently, thereby impacting  $SUVR$ .<sup>28,33,34</sup> Furthermore, flow changes might also occur as a result of therapeutic agents which may cause neuro-inflammation and disrupt blood–brain barrier integrity.

It could be argued that quantitative methods are not feasible for clinical trials with large sample sizes, since these methods require long and expensive dynamic scanning protocols. However, although differences in standard deviation between quantitative and semi-quantitative methods are small, we showed that these small differences had large implications on required sample sizes, when the expected difference between groups or change over time (i.e. the effect size) was small (for instance 0.5%, comparable to values

observed in amyloid-positive cognitively normal individuals in Jack et al.<sup>35</sup>), i.e. when the effect size has the same order of magnitude as the repeatability. However, with larger expected changes over time, this difference between quantitative and semi-quantitative methods was diminished. To date, only two studies evaluated tau progression over time using  $[^{18}F]$ Flortaucipir PET. In cognitively unimpaired subjects, increases of 2–3%<sup>36</sup> or 0.5% (amyloid- $\beta$  positives only<sup>35</sup>) in  $SUVR$  per year were observed. In AD patients, greater increases were observed, ranging from 3%<sup>35</sup> to 3–6%  $SUVR$  per year.<sup>36</sup> Although the differences in yearly change between cognitively normal subjects and AD patients are convincing, especially in cognitively normal subjects, yearly change rates may not exceed TRT repeatability and should be interpreted with caution.

Only one other study describes the TRT repeatability of  $[^{18}F]$ Flortaucipir.<sup>14</sup> In this study, 16 patients with MCI or AD dementia and 5 controls were included, who only underwent static test and retest  $[^{18}F]$ Flortaucipir PET scans. Data were analyzed using two different reference regions: the cerebellar crus and the parametric estimate of reference signal intensity (PERSI) reference region, which was derived for each subject separately and based on the peak frequency of voxel intensity within a pre-defined white matter region.<sup>37</sup> Relative percentage change for  $SUVR_{80-100}$  was comparable to our results, with values ranging from  $0.07 \pm 2.3\%$  (frontal lobe) to  $-1.15 \pm 4.4\%$  (posterior hippocampus) for PERSI reference region and  $0.00 \pm 4.5\%$  (frontal lobe) to  $-0.56 \pm 5.04\%$  (parietal lobe) for cerebellar crus. For  $[^{18}F]$ -labeled amyloid tracers, percentage test–retest variability between 1.5% and 7.5% has been reported for  $SUVR$ .<sup>38</sup> In line with our results, for  $[^{11}C]$ PiB, TRT repeatability was worse for semi-quantitative methods ( $\sim 7\%$  for  $SUVR$ ) than for quantitative methods ( $\sim 3\%$  for SRTM2).<sup>39</sup> However, observed differences between quantitative and semi-quantitative methods were less pronounced in the current study. We observed no differences in TRT repeatability when using a white matter reference region. This is not in line with Southeikal et al.,<sup>37</sup> who report that test–retest repeatability was better when using PERSI than when using cerebellar gray matter as reference region.

Our findings build upon previous reports, further validating the use of  $[^{18}F]$ Flortaucipir PET for the assessment of underlying tau load.  $[^{18}F]$ Flortaucipir's kinetics are best described by a two tissue reversible model<sup>25</sup> with blood volume parameter ( $2T4k_{VB}$ ).<sup>26,40</sup> When comparing parameters derived from  $[^{18}F]$ Flortaucipir's full kinetic model with various reference tissue models against that estimated using a

RPM correlated best.<sup>23</sup> With quantitative methods performing better than semi-quantitative methods in terms of test–retest repeatability, taken together, RPM might be the most optimal reference region method to quantify tau load using [<sup>18</sup>F]Flortaucipir PET. The main drawback of [<sup>18</sup>F]Flortaucipir is its known off-target binding in the choroid plexus, which could affect tracer quantification in nearby structures such as the entorhinal cortex and hippocampus.<sup>41,42</sup> The worse TRT repeatability we observed in subcortical structures, could possibly partly reflect different off-target binding at test and retest scans. However, another explanation could be the small volumes of subcortical structures, which could have resulted in worse TRT repeatability. Furthermore, TRT repeatability in hippocampus yielded no exceptional values (Supplementary Table 1).

Among the strengths of this study are the use of dynamic [<sup>18</sup>F]Flortaucipir PET scans, which allowed us to compare tau load as measured with quantitative versus semi-quantitative approaches. The present study is limited by the small sample size. As we did not have arterial input data available, the TRT repeatability of full kinetic models is not known.

In conclusion, test–retest repeatability for [<sup>18</sup>F]Flortaucipir was good for all reference methods used. In terms of repeatability between test and retest, quantitative methods performed slightly better than semi-quantitative methods and should be preferred when small (<3%) changes over time are expected.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Amsterdam Alzheimer Center is supported by Alzheimer Nederland and Stichting VUmc funds. This study is funded by ZonMW-Memorabel (project no 733050203).

### Acknowledgements

Research of Amsterdam Alzheimer Center is part of the Neurodegeneration program of Amsterdam Neuroscience. [<sup>18</sup>F]Flortaucipir PET scans were made possible by Avid Radiopharmaceuticals Inc.

### Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: TT, RO, DV, HT, EW, SV, RB, and SG report no disclosures. WvdF received grant support from ZonMW, NWO, EU-FP7, Alzheimer Nederland, CardioVascular Onderzoek Nederland, Stichting Dioraphte, Gieskes-Strijbis Fonds, Boehringer Ingelheim, Piramal Neuroimaging, Roche BV, Janssen Stellar, Combinostics.

All funding is paid to the institution. WvdF holds the Pasman chair. BvB has received research funding from the Alzheimer Association, American Health Assistance Foundation, Alzheimer Nederland, CTMM, ZonMW, Janssen Stellar, AVID Radiopharmaceuticals, NOW and EU-FP7. In addition, he is a trainer for Piramal Neuroimaging, GE and AVID Radiopharmaceuticals. All funding is paid to his institution.

### Authors' contributions

TT acquired data; analyzed and interpreted data; drafted the manuscript. RO designed and conceptualized study; interpreted data, critically revised the manuscript. DV acquired data, critically revised the manuscript. HT acquired data, critically revised the manuscript. EW acquired data, critically revised the manuscript. SV critically revised the manuscript. WvdF critically revised the manuscript. RB designed and conceptualized study, interpreted data, critically revised manuscript. SV designed and conceptualized study, interpreted data, critically revised manuscript. BvB designed and conceptualized study, interpreted data, critically revised manuscript.

### ORCID iD

Sander CJ Verfaillie  <https://orcid.org/0000-0003-1820-3378>

### Supplemental material

Supplemental material for this paper can be found at the journal website: <http://journals.sagepub.com/home/jcb>

### References

1. Querfurth HW and LaFerla FM. Alzheimer's disease. *N Engl J Med* 2010; 362: 329–344.
2. Scheltens P, Blennow K, Breteler MM, et al. Alzheimer's disease. *Lancet* 2016; 388: 505–517.
3. Xia CF, Arteaga J, Chen G, et al. [(18)F]T807, a novel tau positron emission tomography imaging agent for Alzheimer's disease. *Alzheimers Dement* 2013; 9: 666–676.
4. Hall B, Mak E, Cervenka S, et al. In vivo tau PET imaging in dementia: pathophysiology, radiotracer quantification, and a systematic review of clinical findings. *Ageing Res Rev* 2017; 36: 50–63.
5. Carson RE, Channing MA, Blasberg RG, et al. Comparison of bolus and infusion methods for receptor quantitation: application to [18F]cyclofoxy and positron emission tomography. *J Cereb Blood Flow Metab* 1993; 13: 24–42.
6. Ossenkoppele R, Prins ND and van Berckel BN. Amyloid imaging in clinical trials. *Alzheimers Res Ther* 2013; 5: 36.
7. Gunn RN, Slifstein M, Searle GE, et al. Quantitative imaging of protein targets in the human brain with PET. *Phys Med Biol* 2015; 60: R363–R411.
8. Heurling K, Buckley C, Van Laere K, et al. Parametric imaging and quantitative analysis of the PET amyloid ligand [(18)F]flutemetamol. *Neuroimage* 2015; 121: 184–192.

9. Lammertsma AA. Forward to the past: the case for quantitative PET imaging. *J Nucl Med* 2017; 58: 1019–1024.
10. Lammertsma AA and Hume SP. Simplified reference tissue model for PET receptor studies. *Neuroimage* 1996; 4: 153–158.
11. Gunn RN, Lammertsma AA, Hume SP, et al. Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *Neuroimage* 1997; 6: 279–287.
12. Wu Y and Carson RE. Noise reduction in the simplified reference tissue model for neuroreceptor functional imaging. *J Cereb Blood Flow Metab* 2002; 22: 1440–1452.
13. Logan J, Fowler JS, Volkow ND, et al. Distribution volume ratios without blood sampling from graphical analysis of PET data. *J Cereb Blood Flow Metab* 1996; 16: 834–840.
14. Devous MD Sr, Joshi AD, Navitsky M, et al. Test–retest reproducibility for the tau PET imaging agent Flortaucipir F 18. *J Nucl Med* 2018; 59: 937–943.
15. van der Flier WM and Scheltens P. Amsterdam dementia cohort: performing research to optimize care. *J Alzheimers Dis* 2018; 62: 1091–1111.
16. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer’s disease: recommendations from the National Institute on Aging–Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimers Dement* 2011; 7: 270–279.
17. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer’s disease: recommendations from the National Institute on Aging–Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimers Dement* 2011; 7: 263–269.
18. Tijms BM, Willemse EAJ, Zwan MD, et al. Unbiased approach to counteract upward drift in cerebrospinal fluid amyloid-beta 1-42 analysis results. *Clin Chem* 2018; 64: 576–585.
19. Groot C, van Loenhoud AC, Barkhof F, et al. Differential effects of cognitive reserve and brain reserve on cognition in Alzheimer disease. *Neurology* 2018; 90: e149–e156.
20. Hammers A, Allom R, Koeppe MJ, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum Brain Mapp* 2003; 19: 224–247.
21. Svarer C, Madsen K, Hasselbalch SG, et al. MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. *Neuroimage* 2005; 24: 969–979.
22. Boellaard R, Yaqub M, Lubberink M, et al. PPET: A software tool for kinetic and parametric analyses of dynamic PET studies. *NeuroImage* 2006; 31: T62.
23. Golla SS, Wolters EE, Timmers T, et al. (2018). Parametric methods for [<sup>18</sup>F]flortaucipir PET. *J Cereb Blood Flow Metab*. Epub ahead of print 20 December 2018. DOI:10.1177/0271678X18820765.
24. Braak H and Braak E. Staging of Alzheimer’s disease-related neurofibrillary changes. *Neurobiol Aging* 1995; 16: 271–278; discussion 8–84.
25. Barret O, Alagille D, Sanabria S, et al. Kinetic modeling of the tau PET tracer (18)F-AV-1451 in human healthy volunteers and Alzheimer disease subjects. *J Nucl Med* 2017; 58: 1124–1131.
26. Wooten DW, Guehl NJ, Verwer EE, et al. Pharmacokinetic evaluation of the Tau PET radiotracer (18)F-T807 ((18)F-AV-1451) in human subjects. *J Nucl Med* 2017; 58: 484–491.
27. Hristova I, Boellaard R, Vogel W, et al. Retrospective quality control review of FDG scans in the imaging sub-study of PALETTE EORTC 62072/VEG110727: a randomized, double-blind, placebo-controlled phase III trial. *Eur J Nucl Med Mol Imaging* 2015; 42: 848–857.
28. van Berckel BN, Ossenkoppele R, Tolboom N, et al. Longitudinal amyloid imaging using 11C-PiB: methodologic considerations. *J Nucl Med* 2013; 54: 1570–1576.
29. Bremner JP, van Berckel BN, Persoon S, et al. Day-to-day test-retest variability of CBF, CMRO<sub>2</sub>, and OEF measurements using dynamic 15O PET studies. *Mol Imaging Biol* 2011; 13: 759–768.
30. Alsop DC, Dai W, Grossman M, et al. Arterial spin labeling blood flow MRI: its role in the early characterization of Alzheimer’s disease. *J Alzheimers Dis* 2010; 20: 871–880.
31. Austin BP, Nair VA, Meier TB, et al. Effects of hypoperfusion in Alzheimer’s disease. *J Alzheimers Dis* 2011; 26: 123–133.
32. Binnewijzend MA, Kuijper JP, Benedictus MR, et al. Cerebral blood flow measured with 3D pseudocontinuous arterial spin-labeling MR imaging in Alzheimer disease and mild cognitive impairment: a marker for disease severity. *Radiology* 2013; 267: 221–230.
33. Cselenyi Z and Farde L. Quantification of blood flow-dependent component in estimates of beta-amyloid load obtained using quasi-steady-state standardized uptake value ratio. *J Cereb Blood Flow Metab* 2015; 35: 1485–1493.
34. Ottoy J, Verhaeghe J, Niemantsverdriet E, et al. A simulation study on the impact of the blood flow-dependent component in [18F]AV45 SUVR in Alzheimer’s disease. *PLoS One* 2017; 12: e0189155.
35. Jack CR Jr, Wiste HJ, Schwarz CG, et al. Longitudinal tau PET in ageing and Alzheimer’s disease. *Brain* 2018; 141: 1517–1528.
36. Harrison TM, La Joie R, Maass A, et al. Longitudinal tau accumulation and atrophy in aging and Alzheimer disease. *Ann Neurol* 2019; 85: 229–240.
37. Souhekal S, Devous MD Sr, Kennedy I, et al. Flortaucipir F 18 quantitation using parametric estimation of reference signal intensity. *J Nucl Med* 2018; 59: 944–951.
38. Vandenberghe R, Adamczuk K, Dupont P, et al. Amyloid PET in clinical practice: its place in the multi-dimensional space of Alzheimer’s disease. *Neuroimage Clin* 2013; 2: 497–511.
39. Tolboom N, Yaqub M, Boellaard R, et al. Test-retest variability of quantitative [11C]PIB studies in

- Alzheimer's disease. *Eur J Nucl Med Mol Imaging* 2009; 36: 1629–1638.
40. Golla SSV, Timmers T, Ossenkoppele R, et al. Quantification of tau load using [(18)F]AV1451 PET. *Mol Imaging Biol* 2017; 19: 963–971.
41. Marquie M, Normandin MD, Vanderburg CR, et al. Validating novel tau positron emission tomography tracer [F-18]-AV-1451 (T807) on postmortem brain tissue. *Ann Neurol* 2015; 78: 787–800.
42. Wolters EE, Golla SSV, Timmers T, et al. A novel partial volume correction method for accurate quantification of [(18)F] flortaucipir in the hippocampus. *EJNMMI Res* 2018; 8: 79.