
Flortaucipir F 18 Quantitation Using Parametric Estimation of Reference Signal Intensity

Sudeepti Southekal, Michael D. Devous Sr., Ian Kennedy, Michael Navitsky, Ming Lu, Abhinay D. Joshi, Michael J. Pontecorvo, and Mark A. Mintun

Avid Radiopharmaceuticals, Inc. (a wholly owned subsidiary of Eli Lilly and Company), Philadelphia Pennsylvania

PET imaging of tau pathology in Alzheimer disease may benefit from the use of white matter reference regions. These regions have shown reduced variability compared with conventional cerebellar regions in amyloid imaging. However, they are susceptible to contamination from partial-volume blurring of tracer uptake in the cortex. We present a new technique, PERSI (Parametric Estimation of Reference Signal Intensity), for flortaucipir F 18 count normalization that leverages the advantages of white matter reference regions while mitigating potential partial-volume effects. **Methods:** Subjects with a clinical diagnosis of Alzheimer disease, mild cognitive impairment, or normal cognition underwent T1-weighted MRI and flortaucipir imaging (to determine amyloid [A β] status) at screening and flortaucipir F 18 imaging at single or multiple time points. Flortaucipir F 18 images, acquired as 4 \times 5 min frames 80 min after a 370-MBq injection, were motion-corrected, averaged, and transformed to Montreal Neurological Institute (MNI) space. The PERSI reference region was calculated for each scan by fitting a bimodal gaussian distribution to the voxel-intensity histogram within an atlas-based white matter region and using the center and width of the lower-intensity peak to identify the voxel intensities to be included. Four conventional reference regions were also evaluated: whole cerebellum, cerebellar gray matter, atlas-based white matter, and subject-specific white matter. SUVr (standardized uptake value ratio) was calculated for a statistically defined neocortical volume of interest. Performance was evaluated with respect to test-retest variability in a phase 2 study of 21 subjects (5–34 d between scans). Baseline variability in controls (SD of SUVr and Δ SUVr) and effect sizes for group differences (Cohen *d*; A β -positive impaired vs. A β -negative normal) were evaluated in another phase 2 study with cross-sectional data (*n* = 215) and longitudinal data (*n* = 142/215; 18 \pm 2 mo between scans). **Results:** PERSI showed superior test-retest reproducibility (1.84%) and group separation ability (cross-sectional Cohen *d* = 9.45; longitudinal Cohen *d* = 2.34) compared with other reference regions. Baseline SUVr variability and Δ SUVr were minimal in A β control subjects with no specific flortaucipir F 18 uptake (SUVr, 1.0 \pm 0.04; Δ SUVr, 0.0 \pm 0.02). **Conclusion:** PERSI reduced variability while enhancing discrimination between diagnostic cohorts. Such improvements could lead to more accurate disease staging and robust measurements of changes in tau burden over time for the evaluation of putative therapies.

Key Words: flortaucipir; AV1451; T807; SUVr; reference region

J Nucl Med 2018; 59:944–951

DOI: 10.2967/jnumed.117.200006

The growing acceptance of molecular imaging biomarkers in Alzheimer disease (AD) research and clinical trials emphasizes the need for reliable and reproducible image quantification methods. Used independently or as an adjunct to visual interpretation by trained physicians, quantification in brain PET involves the determination of target-specific biomarker retention in the presence of confounds from physiologic factors (e.g., blood flow and delivery, radiopharmaceutical dose, route of administration, and body weight), technical factors (e.g., scanner model and acquisition parameters), and diffuse or focal nonspecific binding to brain structures that may differ between subjects and over time. In clinical trials enrolling large numbers of subjects, ideal quantification via tracer kinetic analysis (1,2), including noninvasive methods that obviate arterial blood sampling (3,4), can be prohibitively time-consuming and complicated. Therefore, an alternative approach is widely used that involves calculating the ratio of the average activity concentration in the target volume of interest (VOI) to that in a reference region—a tissue region devoid of specific binding of the radiotracer. The reference region serves to achieve intra- and intersubject count normalization, generating an intercomparable measure of SUV (5), that is, SUVr. Several studies (6–8) have validated the results from semiquantitative methods against kinetic models for AD biomarkers.

Reliable reference regions can be critically important to PET investigations. Errors in the mean activity concentration of the reference region directly translate to variability in the SUVr and can affect the power of a trial to detect meaningful signals, or subtle longitudinal changes in such signals, over the noise in the measurement.

Conventionally, tissue regions that are empirically known to contain a negligible number of specific binding sites are used as reference regions. The cerebellar cortex, which is expected to remain free of fibrillar amyloid over time, was selected for amyloid PET imaging with ¹¹C-Pittsburgh compound B (6) because cerebellar and cerebral gray matter (GM) show a similar clearance of Pittsburgh compound B and because this region is stable over time. Cerebellar GM regions are susceptible to noise because of their small size, low signal level, and proximity to the axial periphery of the PET scanner bore (where the scanner has a lower signal-detection sensitivity). With the goal of ameliorating

Received Aug. 14, 2017; revision accepted Nov. 3, 2017.
For correspondence or reprints contact: Sudeepti Southekal, Avid Radiopharmaceuticals, 3711 Market St., Philadelphia, PA 19104.
E-mail: southekal@avidrp.com
Published online Nov. 30, 2017.
COPYRIGHT © 2018 by the Society of Nuclear Medicine and Molecular Imaging.

some of these concerns, the whole cerebellum (including cerebellar white matter [WM]) was validated for ^{18}F -florbetapir in an autopsy study (9) and used to determine an amyloid (^{18}F -florbetapir) positivity threshold (10). Whole cerebellum is now routinely used for ^{18}F -florbetapir cross-sectional imaging (11) and is also currently recommended by the Centiloid Project (12) for the quantitation of amyloid burden with ^{11}C -Pittsburgh compound B. It has been noticed, however, that the location of the cerebellum within the approximately 15-cm-long axial PET field of view can make it vulnerable to truncation. For tau imaging, focal areas of uptake in the cerebellum (e.g., dentate nuclei) may also present a challenge.

The use of a subcortical WM reference region (centrum semiovale) for longitudinal studies with ^{18}F -florbetapir has been investigated by several groups (13–15). This region has the advantage of a shared axial location with the cortex and is typically larger (and hence less sensitive to noise) than cerebellar regions. WM regions showed lower variability than the cerebellar cortex, especially for longitudinal data. To take advantage of this reduced variability yet retain the merits of autopsy-validated SUVr referenced to cerebellum, a 2-step normalization process was proposed for longitudinal studies (16). Data are first normalized to whole cerebellum and then scaled by the ratio of subcortical WM at follow-up to subcortical WM at the first visit.

A phenomenon known as partial-volume effect (PVE) presents an additional challenge for fixed anatomic regions. PVE refers to blurring caused by the limited spatial resolution of PET scanners (on the order of 4–6 mm (17)), which leads to apparent spillover

or cross-contamination of counts between adjacent structures. If counts from areas with specific tracer binding spill in to the reference region, SUVr will be underestimated. This phenomenon is of particular concern for WM reference regions because of their proximity to cortical GM (containing specific tracer uptake in positive scans). The fraction of counts spilling in to the reference region depends on the degree of tracer binding (cortical SUVr) and the size of the structure with specific uptake (18,19). Because both cortical atrophy and uptake levels can change over time, a WM reference region can be affected by PVE differently at different time points, thus obfuscating longitudinal measures.

In this paper, we introduce and evaluate a novel technique, PERSI (Parametric Estimation of Reference Signal Intensity), to identify a subject-specific WM reference region for the tau biomarker flortaucipir F 18 that addresses the above concerns with traditional reference regions. PERSI is a subject-specific, data-driven technique that identifies voxels with apparent nonspecific flortaucipir F 18 uptake within a WM region, based on the signal intensity histogram of the region. The underlying assumption of the approach is that the WM in a flortaucipir F 18 PET image is composed of the following categories of voxels in differing proportions: voxels with negligible specific binding; voxels with contamination (or spillover) by counts from adjacent cortical tissues with specific uptake or regions with off-target binding; and rare occurrences of focal WM areas with elevated binding. Of these, the voxels with negligible specific binding are considered appropriate for count normalization across subjects and scans. The goal of

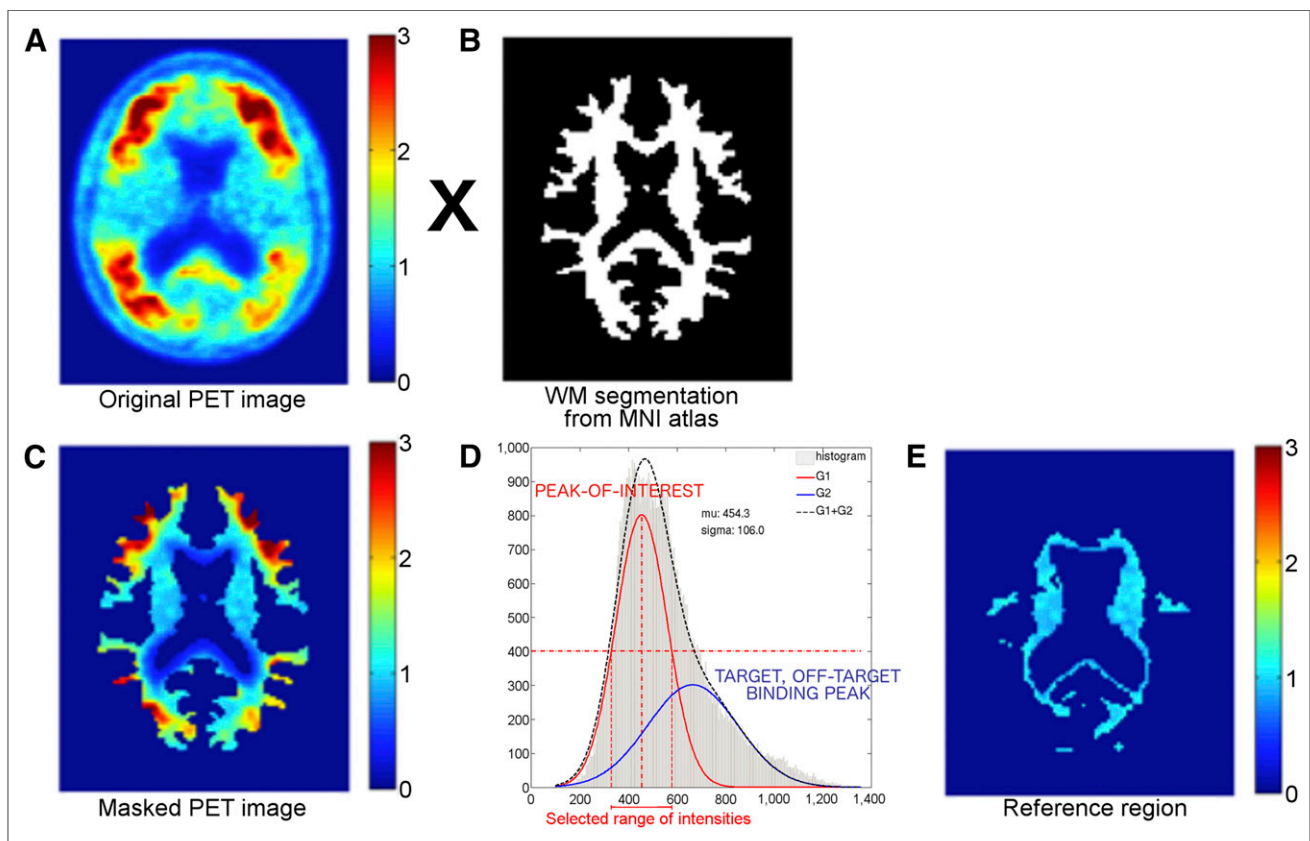


FIGURE 1. Graphical description of PERSI. Original PET image (A) is masked using WM segmentation from MNI atlas (B) to extract counts in WM (C). The resulting nonzero counts are plotted as histogram (D) and fit to bimodal gaussian distribution. Parameters of lower-intensity peak, shown in red, are used to select range of values to be used as reference region (E).

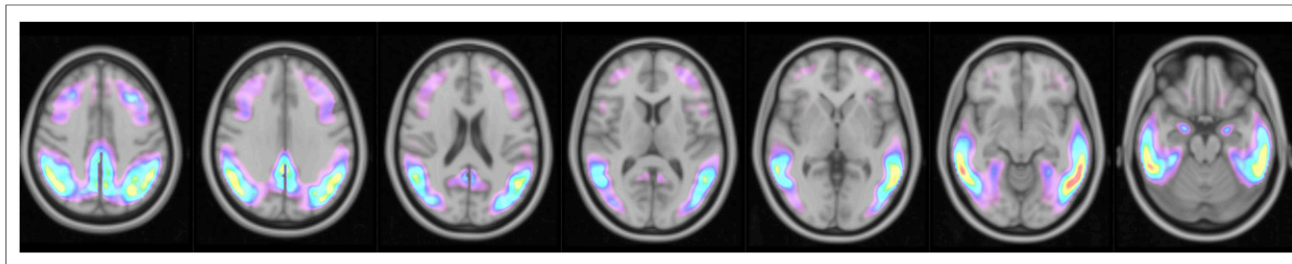


FIGURE 2. MUBADA VOI obtained using discriminant analysis to identify voxels predominantly associated with group separation between cohorts, overlaid on MNI template. This VOI is applied to measured PET data in weighted fashion to generate SUVr.

PERSI is to identify such voxels, which will generally be the majority within the WM intensity histogram and are expected to have a lower intensity than voxels from other categories within that histogram. These voxels are identified by fitting the histogram to a bimodal gaussian distribution (i.e., the sum of 2 gaussian distributions). The higher-intensity peak captures spill-in from tissues with elevated flortaucipir F 18 uptake, whereas the lower-intensity peak represents nonspecific (reference) signal intensity (Fig. 1). Because this technique is not constrained by anatomic boundaries, it maximizes the number of voxels, and hence minimizes the variance, of the reference region. Further, PERSI is designed to be relatively robust to issues related to positioning, registration, and PVE.

MATERIALS AND METHODS

Design Summary

We evaluated the performance of PERSI for flortaucipir F 18 relative to traditional reference regions using data from 2 previously reported studies (NCT01992380 (20) and NCT02016560 (21)), which were performed in accordance with the ethical standards of the institutional or national research committee and with the 1964 Helsinki declaration and its later amendments, or comparable ethical standards. No new subjects were administered flortaucipir F 18 for this work. The first of these previous studies was a test–retest study wherein subjects underwent flortaucipir F 18 scanning on 2 occasions within approximately 4 wk. These data were used to determine the stability of the reference region without changes in the underlying pathology over time. The second study was a phase 2 clinical trial with cross-sectional (3 cohorts) and longitudinal (18 ± 2 mo between scans) components. Knowledge of the amyloid (Aβ) status of the subjects in this study allowed us to evaluate the baseline variability of PERSI vis-à-vis traditional reference regions, since Aβ-negative (Aβ−) older controls are presumed not to accumulate tau pathology outside the mesial temporal lobe. The second study also allowed us to assess the cross-sectional and longitudinal performance of the technique at varying levels of impairment. Details of these analyses are described below.

Flortaucipir F 18 Image Acquisition and Preprocessing

Flortaucipir F 18 images were acquired as four 5-min dynamic frames, beginning 80 min after an injection of approximately 370 MBq. Frames 2–4 were rigidly registered to the first frame using the MCFLIRT tool provided by FSL software for interframe motion correction. For images acquired at a start time offset from 80 min after injection, time correction factors were calculated for each voxel as the slope of the linear regression line through the 4 time points, multiplied by the time offset (22). Motion- and time-corrected frames were then averaged and coregistered to the subject’s MR images using the FSL FLIRT linear registration tool. Next, the unified segmentation and normalization algorithm (23) in SPM8 was run to spatially normalize

the T1-weighted MR images acquired at screening to the Montreal Neurological Institute (MNI) brain template (24) while simultaneously generating probabilistic segmentations for the GM, WM, and cerebrospinal fluid.

PERSI Implementation

The MNI atlas was segmented using FSL to generate a binary WM image. PET images spatially normalized to MNI space were then masked with this image, and the resulting voxel intensities were plotted as a histogram. The histogram for each subject was fit to a bimodal gaussian distribution, as shown in Figure 1. The peak location (μ_1) and the sigma (σ_1) of the first Gaussian peak were extracted for analyses. Voxels with values within the full width at half maximum [FWHM = $\sigma_1 * 2\sqrt{(2\ln(2))}$] of μ_1 , i.e. $\mu_1 \pm$ FWHM, were included in the reference region for that subject.

SUVr Analysis

For comparison with PERSI, SUVr values were generated using 4 additional reference regions: whole cerebellum, (wholeCere), cerebellar GM (cereCrus), atlas-based WM (atlasWM), and subject-specific WM (ssWM). wholeCere was delineated manually on a template (10). cereCrus was derived from the Automated Anatomical Labeling (AAL) software VOI “cere-crus-1” and modified to avoid potential overlap with other structures by translating it inferiorly by 6 mm. For the atlasWM region, the MNI atlas was segmented to extract a binary WM mask. Then, 2 steps were undertaken to eliminate voxels susceptible to PVE spillover of counts from the cortex: (1) a mask made up of a combination all volumes from the AAL template, a cerebellum volume, and a brain stem volume was created and applied to the WM segmentation to remove the associated voxels; (2) the resulting volume was eroded using a 5-voxel box erosion in FSL. The same 2 steps were applied to generate ssWM regions, after a threshold was applied to subject-specific, probabilistic MRI segmentations to preserve voxels

TABLE 1
Cross-Sectional and Longitudinal Study Participants

Clinical diagnosis group	Aβ−	Aβ+	Total
Probable AD	16 (10)	30 (19)	46 (29)
MCI	49 (33)	47 (28)	96 (61)
Older CN (>50 y)	52 (47)	5 (5)	57 (52)
Younger CN (≤40 y)	16 (0)	0 (0)	16 (0)
All	133 (90)	82 (52)	215 (142)

MCI = mild cognitive impairment.
Data are number at visit 1, followed by number at visit 2 in parentheses.

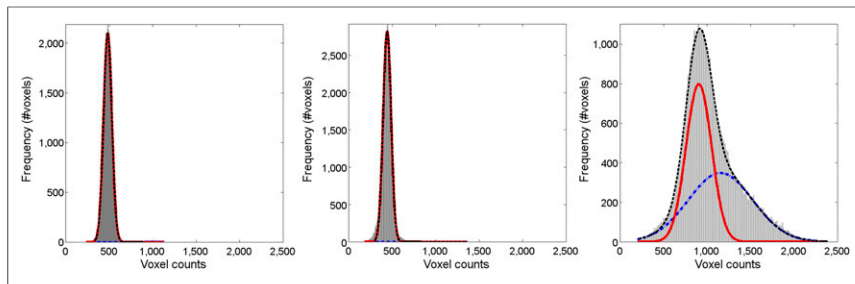


FIGURE 3. Typical WM histograms for 3 subjects: 36-y-old Aβ⁻ CN (left); 69-y-old Aβ⁻ CN (center); and 75-y-old Aβ⁺ AD (right). Red line denotes lower-intensity signal, blue line denotes higher-intensity peak, and black line denotes sum of 2 peaks (bimodal gaussian). For the 2 CNs, only 1 peak is observed. For subject with elevated flortaucipir F 18 uptake in GM, counts that have spilled into WM are captured by blue peak.

with more than a 50% probability of being WM. SUV_r values were calculated for a target VOI (Fig. 2) relative to the 5 candidate reference regions. This VOI was derived using a discriminant analysis technique (MUBADA) (25) to identify the voxels most relevant to separation by diagnostic group and Aβ status. More details are provided in a companion article (20).

Validation of PERSI Implementation for Longitudinal Data

For longitudinal datasets with more than 1 scan per subject, the PERSI algorithm was run separately for each scan, and a common reference region was generated from the voxels that were shared by the individual regions. Because this approach may not be practical for all studies, SUV_r was compared using both the combined reference region and the individual reference regions for each scan.

Evaluation of Test-Retest Reproducibility

Twenty-four subjects (10 with AD, 8 with mild cognitive impairment, and 6 with normal cognition [CN]) were imaged at baseline, followed by a repeat scan between 48 h and 4 wk later (20). Twenty-one of these subjects had technically adequate scans for inclusion in these analyses. The SD of SUV_r percentage difference was calculated and compared across reference regions as a measure of test-retest reproducibility. A paired *t* test was also performed to evaluate differences between test and retest scans.

Cross-Sectional and Longitudinal Analyses

In total, 215 subjects from a phase 2 study (21) were imaged at baseline, of which 142 returned for a follow-up scan 18 ± 2 mo after the screening visit (shown in parentheses in Table 1). Subjects also had screening-visit ¹⁸F-florbetapir scans that were interpreted as Aβ-

positive (Aβ⁺) or Aβ⁻ based on visual reads augmented by quantitative information, as previously described (26).

To assess reference region performance in CNs, SUV_r was calculated for 68 CN subjects (Table 1) determined to be Aβ⁻ (¹⁸F-florbetapir). Of these, 16 were between 20 and 40 y old (younger CNs), and the remainder were over 50 y old (older CNs). The accuracy and precision across reference regions were assessed, given an expected SUV_r of 1 for Aβ⁻ CN subjects with no specific uptake. For each reference region, a 1-sample *t* test was performed to test the hypothesis that the mean SUV_r for this cohort was 1.

Cross-sectional SUV_r, longitudinal SUV_r change, and group separation were evaluated

in subjects classified into groups based on clinical diagnosis (CN, mild cognitive impairment, or AD) and Aβ status (Aβ⁺ or Aβ⁻; Table 1). Mean SUV_r and SDs for each group were computed using each of the 5 reference regions. Baseline variability in ΔSUV_r was evaluated for the Aβ⁻ older CN cohort (younger CNs were not evaluated longitudinally), given an expected ΔSUV_r of 0. The impact of reference regions on Cohen *d* effect sizes for differences between Aβ⁺ impaired subjects (AD and mild cognitive impairment) and Aβ⁻ CNs was assessed for cross-sectional SUV_r and longitudinal SUV_r change (ΔSUV_r = SUV_r_{follow-up} - SUV_r_{baseline}).

RESULTS

Figure 3 shows representative signal intensity histograms (within the WM mask) used to derive the PERSI reference region for representative subjects clinically diagnosed as younger CN, older CN, or Aβ⁺ AD. For the 2 CNs, only a single peak was observed. For the subject with elevated flortaucipir F 18 uptake in the GM, counts that spilled into the WM formed a second peak. PERSI reference regions comprise a larger number of voxels than traditional reference regions, even though the volume of the PERSI region decreases with increasing count spillover from the GM (Table 2). The average PERSI region for Aβ⁺ subjects (though smaller than the average Aβ⁻ PERSI region because of spillover) is a factor of 2.5 larger than the whole cerebellum VOI.

Evaluation of Test-Retest Reproducibility

Test and retest datasets correlated significantly (*P* < 0.01) for all reference regions (Table 3; Fig. 4). Reproducibility (SD) varied, depending on reference region, with WM regions (1.8%–2.7%)

TABLE 2
Reference Region Sizes

Region	Size*
Whole cerebellum	15,878
Cerebellar GM	1,815
Pons	1,226
Atlas-based WM	8,384
Average subject-specific WM	1,140
Average PERSI region (Aβ ⁻)	44,222
Average PERSI region (Aβ ⁺)	38,954

*Number of 2 × 2 × 2 mm voxels.

TABLE 3
Comparison of Reference Regions Using Test-Retest Data

Reference region	R ²	Mean difference (%)	SD of difference (%)
Whole cerebellum	0.971	0.066	3.700
Cerebellar GM	0.961	0.042	4.179
Atlas-based WM	0.986	-0.764	2.616
Subject-specific WM	0.987	-0.869	2.663
PERSI	0.990	-0.165	1.841

All correlations were statistically significant, with *P* < 0.01.

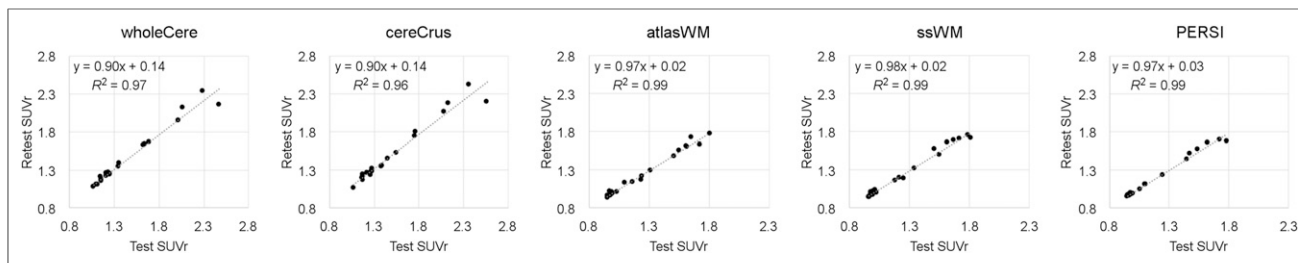


FIGURE 4. Comparison of test and retest conditions for 21 subjects using 5 reference regions. wholeCere = whole cerebellum; cereCrus = cerebellar GM; atlasWM = atlas-based WM; ssWM = subject-specific WM.

outperforming cerebellar regions (3.7%–4.2%). PERSI SUVR was the most reproducible SUVR evaluated. The absolute range of SUVR for WM regions was smaller than for cerebellar regions.

Validation of PERSI Implementation for Longitudinal Data

The 2 methods used to identify PERSI volumes in longitudinal datasets generated consistent results, as shown in Figure 5. Although the use of individual masks for each scan produced a slightly lower SUVR overall, SUVR correlated strongly ($R = 1$) across the disease spectrum.

Cross-Sectional and Longitudinal Analyses

Table 4 shows mean SUVR \pm SD for $A\beta^-$ younger and older CNs. For this cohort, presumed to have no tau pathology, all reference regions generated an SUVR consistent with expectations (expected SUVR = 1), indicating low baseline variability in flortaucipir F 18 SUVR overall. However, there appeared to be higher variability with cerebellar regions (7%–11%) than with WM regions (4%–7%). Again, PERSI outperformed other regions in both accuracy and reproducibility. One-sample t tests indicated that PERSI SUVR for older and younger CNs was not significantly

different from 1, whereas statistically significant differences were noted for the other regions.

Similarly, when measuring changes in SUVR over 18 mo for $A\beta^-$ older CNs (Table 4), PERSI registered the lowest mean and SD of change. This finding is consistent with the understanding that this demographic was not expected to accumulate additional tau over 18 mo. PERSI also generated the tightest Δ SUVR range (Table 4) for this demographic, again suggesting low baseline variability in the measurement of longitudinal changes using PERSI.

Box plots of cross-sectional SUVR grouped by clinical diagnosis and $A\beta$ status for each reference region are presented in Figure 6. Overall, flortaucipir F 18 SUVR was effective at distinguishing $A\beta^+$ impaired subjects from $A\beta^-$ subjects (21). Cerebellar reference regions again generated a larger range of SUVR than did WM. Unlike atlas-based or subject-specific WM, however, the smaller range of PERSI SUVR did not translate to a lowered ability to distinguish between groups, as revealed by the Cohen d effect sizes in Table 4. PERSI had the highest effect size for differentiating impaired from control cohorts in all reference regions tested. Although PERSI, atlas-based WM, and subject-specific WM had comparable test–retest reproducibility (Table 3) and baseline variability (Table 4), PERSI had an almost 2-fold increase in effect size.

Similar advantages were observed with PERSI in the measurement of Δ SUVR 18 mo after baseline. The low variability in Δ SUVR for $A\beta^-$ groups led to an improved signal-to-noise ratio for the measurement of Δ SUVR for impaired subjects, resulting in a greater effect size (Table 4). Longitudinal Cohen d was 2.34 with PERSI, 1.3 times higher than the next highest value (atlas-based WM). The occurrence of Δ SUVR substantially below 0 (considered biologically unlikely) for $A\beta^+$ subjects relative to $A\beta^-$ subjects was also reduced with PERSI, compared with the other candidate reference regions (Fig. 6). There was 1 $A\beta^-$ AD subject for whom Δ SUVR was less than 0 regardless of reference region, including PERSI; however, the quantification of this subject may have been affected by subject motion.

DISCUSSION

We have evaluated PERSI, a new inter- and intrasubject count normalization technique for flortaucipir F 18. PERSI estimates nonspecific (reference) signal intensity within WM by modeling the spill-in of counts from GM. Our analyses showed that PERSI was associated with the highest effect sizes for diagnostic group separation and the lowest variability among the reference regions evaluated, for cross-sectional as well as longitudinal measures. In the absence of a truth standard, PERSI generated

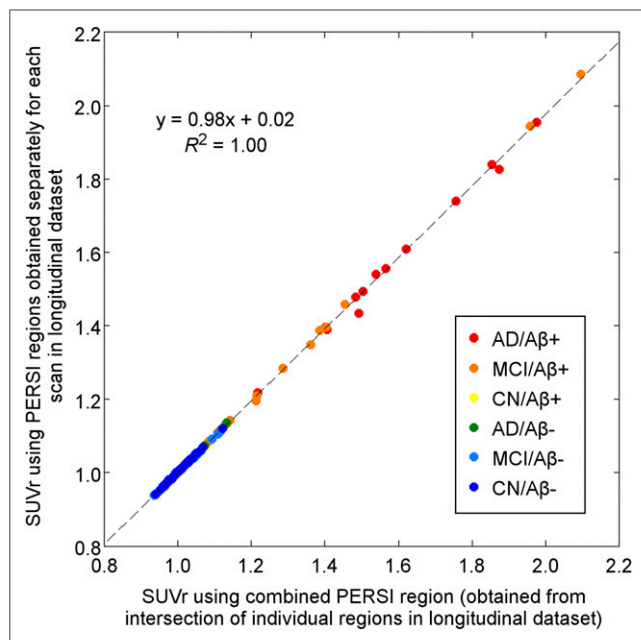


FIGURE 5. PERSI SUVR comparison for longitudinal data: individual vs. combined mask.

TABLE 4
Comparison of 5 Reference Regions Using Cross-Sectional and Longitudinal Data

Measure	Group	Whole cerebellum	Cerebellar GM	Atlas-based WM	Subject-specific WM	PERSI
SUVr, mean ± SD*	YCN (<i>n</i> = 16)	1.06 ± 0.08	1.10 ± 0.08	0.98 ± 0.06	0.99 ± 0.07	1.01 ± 0.04
	OCN (<i>n</i> = 52)	1.09 ± 0.07	1.12 ± 0.11	1.06 ± 0.07	1.06 ± 0.07	1.01 ± 0.04
ΔSUVr, mean ± SD†	OCN/Aβ- (<i>n</i> = 47)	0.01 ± 0.03	0.01 ± 0.04	0.01 ± 0.04	0.01 ± 0.04	0.00 ± 0.02
ΔSUVr, range‡	OCN/Aβ- (<i>n</i> = 47)	-0.05/0.13	-0.09/0.11	-0.09/0.16	-0.09/0.19	-0.06/0.09
Cohen d SUVr¶	Aβ+ AD & MCI vs. Aβ- CN	6.10	4.57	5.12	5.43	9.45
	Aβ+ AD vs. Aβ- CN	8.14	6.20	7.01	7.68	13.61
	Aβ+ MCI vs. Aβ- CN	4.80	3.53	3.91	3.99	6.79
Cohen d ΔSUVr§	Aβ+ AD & MCI vs. Aβ- CN	1.65	1.39	1.82	1.47	2.34
	Aβ+ AD vs. Aβ- CN	2.75	2.03	2.51	2.00	3.19
	Aβ+ MCI vs. Aβ- CN	0.99	0.95	1.34	1.11	1.76

*Baseline variability of SUVr measured in Aβ- CNs.

†Baseline variability of ΔSUVr measured in Aβ- CNs.

‡Range of ΔSUVr in Aβ- older CNs.

¶Cross-sectional effect sizes (Cohen d).

§Longitudinal effect sizes (Cohen d).

YCN = younger CNs; OCN = older CNs; MCI = mild cognitive impairment.

results consistent with expectations based on the pathophysiology of AD. PERSI should, in theory, also be applicable to other tauopathies.

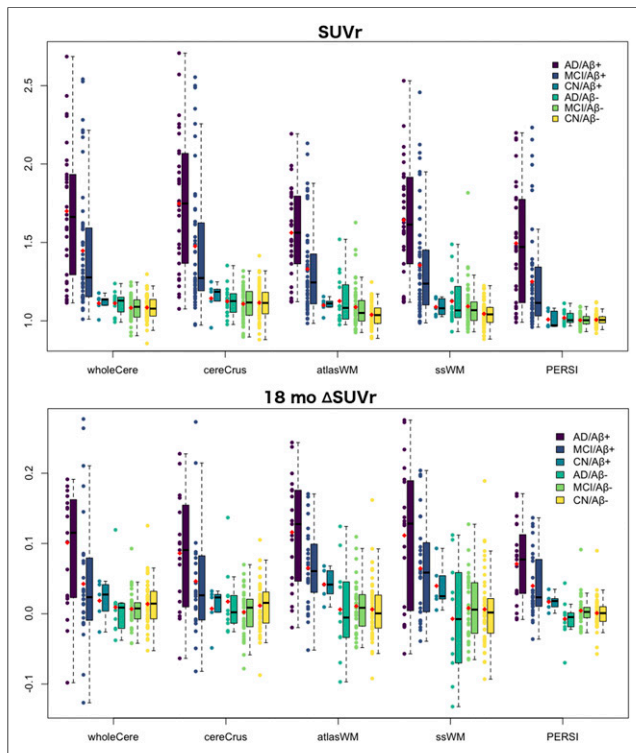


FIGURE 6. Comparison of SUVr and ΔSUVr using 5 reference regions and MUBADA as target region. Subjects are grouped by clinical diagnosis and Aβ status. Red dot denotes mean. wholeCere = whole cerebellum; cereCrus = cerebellar GM; atlasWM = atlas-based WM; ssWM = subject-specific WM.

Unique individual patterns of flortaucipir F 18 uptake (21) (and associated spillover) make it challenging to identify anatomic reference tissues for subjects across the disease spectrum. Issues with subject positioning and image truncation, errors in spatial normalization to an atlas, and unexpected areas of truly increased tracer uptake in selected reference tissues (e.g., dentate nuclei in cerebellum for flortaucipir F 18) may further compound the problem. Strategies to overcome PVE contamination in the reference region typically involve erosion or shrinking of WM segmentations, similar to our atlas-based and subject-specific WM regions, to minimize voxels with expected contamination. Our cross-sectional and longitudinal effect size results show that this strategy does not completely eliminate PVE. Further, such regions require several preprocessing steps that might be sensitive to noise. PERSI is designed to leverage the advantages of WM reference regions (27) while mitigating the potential negative impact of PVE. With PERSI, identification of the optimal reference signal is based on voxel intensity rather than anatomic location. Therefore, the reference signal is less likely to be affected by differences in uptake pattern or by image processing errors. Although PVE correction algorithms (19,28–31) also correct VOI counts for the effects of spillover, PERSI reduces potential contamination of the reference region when PVE correction is not applied.

The WM mask described in this article was derived from the MNI template and is common to all subjects. Alternatively, individual WM segmentations can be used as subject-specific masks if volumetric MRI scans are acquired for each subject. In our studies, data were analyzed using both subject-specific and template-based masks, and the results agreed. For ease of implementation and broader applicability in studies without MRI data acquisition, our implementation of PERSI used the template-based WM mask.

For longitudinal datasets, a common PERSI region was obtained for all scans of a given subject from the intersection of the individual PERSI masks. Although this additional step was executed to improve the consistency of our longitudinal results, its omission did not alter our conclusions.

PERSI reference regions had several advantages over other candidates. Our results showed excellent test–retest reproducibility (<2%) and equivalently low variability in ΔSUVr over 18 ± 2 mo using PERSI. However, baseline variability cannot be used in isolation to select a reference region, since the true variability of the data could be underestimated. PERSI surpassed other candidate regions in differentiating between groups stratified on the basis of clinical diagnoses and A β status. Atlas-based WM and subject-specific WM had variability similar to that of PERSI, but the group separation ability for these regions was inferior.

With all reference regions, some subjects appeared to show decreased flortaucipir F 18 binding over time. Although it is possible for tau to decrease in later disease stages because of loss of tau-producing neurons and atrophy, in these mostly mild cognitive impairment and early AD subjects the observed decreases are most likely artifacts of image processing methods. The occurrence of these negative ΔSUVr findings was reduced with PERSI.

The volume of the PERSI region (Table 2) varies depending on the number of voxels deemed to be contaminated by spillover, yet the smallest PERSI region seen in our datasets was substantially larger than traditional reference regions. Large regions are another contributor to low variability, further substantiating the suitability of PERSI for longitudinal measurements.

WM reference regions may suffer from quantitative biases due to count variations brought about by ischemic changes in subcortical WM. Limited literature (32) suggests that calcification after ischemic changes could result in elevated flortaucipir F 18 uptake. In theory, PERSI should be unaffected by focal irregularities in counts because the method specifically excludes areas with a signal intensity detectably different from the overall WM signal. This remains to be proven since ischemic changes in WM were not investigated in this work.

A drawback to PERSI is that it involves additional processing compared with standard reference regions. However, population-based SUV normalization methods have been proposed in the past. Turkheimer et al. (33) developed a supervised clustering method for the neuroinflammatory marker ^{11}C -PK11195 that assumed a weighted linear combination of 6 kinetic classes. A novel method (34) was recently introduced for tau imaging using ^{11}C -PBB3. This method involved first calculating SUVr images using a standard cerebellar cortex reference tissue method and then generating a new reference region for each subject using only the voxels with SUVr between the mean and 2 SDs below the mean SUVr of a database of healthy controls. These methods can provide advantages over traditional methods and, with rigorous evaluation, may prove to be critical to future trials.

CONCLUSION

PERSI reduced variability while enhancing discrimination between diagnostic cohorts. These improvements could lead to more accurate disease staging and robust measurements of changes in tau burden over time for evaluation of the efficacy of putative therapeutic interventions.

DISCLOSURE

All authors are employees of Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly and Company. Avid Radiopharmaceuticals sponsored this study. No other potential conflict of interest relevant to this article was reported.

REFERENCES

- Huang S, Phelps M. Principles of tracer kinetic modeling in positron emission tomography and autoradiography. In: Phelps ME, Mazziotta JC, Schelbert HR, eds. *Positron Emission Tomography and Autoradiography: Principles and Applications for the Brain and Heart*. New York, NY: Raven Press; 1986: 287–346.
- Mintun MA, Raichle ME, Kilbourn MR, Wooten GF, Welch MJ. A quantitative model for the in vivo assessment of drug binding sites with positron emission tomography. *Ann Neurol*. 1984;15:217–227.
- Logan J. Graphical analysis of PET data applied to reversible and irreversible tracers. *Nucl Med Biol*. 2000;27:661–670.
- Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ. Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *Neuroimage*. 1997;6:279–287.
- Thie JA. Understanding the standardized uptake value, its methods, and implications for use. *J Nucl Med*. 2004;45:1431–1434.
- Price JC, Klunk WE, Lopresti BJ, et al. Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh compound-B. *J Cereb Blood Flow Metab*. 2005;25:1528–1547.
- Shcherbinin S, Devous MD, Schwarz AJ, Joshi AD, Navitsky M, Mintun MA. Region-dependent kinetics of the tau PET tracer [^{18}F]-AV-1451 (T807) [abstract]. *Alzheimers Dement*. 2015;11(suppl):113.
- Wong DF, Rosenberg PB, Zhou Y, et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand ^{18}F -AV-45 (florbetapir F 18). *J Nucl Med*. 2010;51:913–920.
- Clark CM, Pontecorvo MJ, Beach TG, 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: 669–678.
- Joshi AD, Pontecorvo MJ, Clark CM, et al. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. *J Nucl Med*. 2012;53:378–384.
- Landau SM, Mintun MA, Joshi AD, et al. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann Neurol*. 2012;72:578–586.
- Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement*. 2015;11: 1.e4–15.e4.
- Wong K-P, Wardak M, Shao W, et al. Quantitative analysis of [^{18}F]FDNDP PET using subcortical white matter as reference region. *Eur J Nucl Med Mol Imaging*. 2010;37:575–588.
- Joshi A, Pontecorvo M, Navitsky MA, Kennedy IA, Mintun M, Devous MD. Measuring change in beta-amyloid burden over time using florbetapir-PET and a subcortical white matter reference region [abstract]. *Alzheimers Dement*. 2014;10 (suppl):902.
- Chen K, Roontiva A, Thiyyagura P, et al. Improved power for characterizing longitudinal amyloid- β PET changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region. *J Nucl Med*. 2015;56:560–566.
- Joshi AD, Pontecorvo MJ, Lu M, Skovronsky DM, Mintun MA, Devous MD. A semiautomated method for quantification of F 18 florbetapir PET images. *J Nucl Med*. 2015;56:1736–1741.
- Khalil MM. *Basic Science of PET Imaging*. New York, NY: Springer; 2016.
- Erlundsson K, Buvat I, Pretorius PH, Thomas BA, Hutton BF. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Phys Med Biol*. 2012;57: R119–R159.
- Thomas BA, Erlundsson K, Modat M, et al. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2011;38:1104–1119.
- Devous MD Sr, Joshi AD, Navitsky M, et al. Test–retest reproducibility for the tau PET imaging agent ^{18}F -AV-1451. *J Nucl Med*. December 28, 2017 [Epub ahead of print].
- Pontecorvo MJ, Devous MD, Navitsky M, et al. Relationships between flortaucipir PET tau binding and amyloid burden, clinical diagnosis, age and cognition. *Brain*. 2017;140:748–763.

22. Southekal S, Devous MD Sr, Navitsky M, Kennedy I, Joshi AD, Mintun MA. Correction for acquisition time discrepancies in SUVr analyses of ¹⁸F-AV1451 tau images. In: *10th Human Amyloid Imaging: Conference Program and Abstracts*. World Events Forum website. http://www.worldeventsforum.com/hai/wp-content/uploads/2016/12/HAI2016_Book-1-6-16.pdf. Published 2016. Accessed March 15, 2018.
23. Ashburner J, Friston KJ. Unified segmentation. *Neuroimage*. 2005;26:839–851.
24. Fonov VS, Evans AC, Mckinstry RC, Almlri CR, Collins L. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*. 2009;47.
25. Abdi H, Williams LJ, Beaton D, et al. Analysis of regional cerebral blood flow data to discriminate among Alzheimer’s disease, frontotemporal dementia, and elderly controls: multi-block barycentric discriminant (MUBADA) methodology. *J Alzheimers Dis*. 2012;31(suppl 3):S189–S201.
26. Pontecorvo MJ, Arora AK, Devine M, et al. Quantitation of PET signal as an adjunct to visual interpretation of florbetapir imaging. *Eur J Nucl Med Mol Imaging*. 2017;44:825–837.
27. Landau SM, Fero A, Baker SL, et al. Measurement of longitudinal β -amyloid change with ¹⁸F-florbetapir PET and standardized uptake value ratios. *J Nucl Med*. 2015;56:567–574.
28. Meltzer CC, Leal JP, Mayberg HS, Wagner HN, Frost JJ. Correction of PET data for partial volume effects in human cerebral cortex by MR imaging. *J Comput Assist Tomogr*. 1990;14:561–570.
29. Müller-Gärtner HW, Links JM, Prince JL, et al. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. *J Cereb Blood Flow Metab*. 1992;12:571–583.
30. Rousset OG, Ma Y, Evans A. Correction for partial volume effects in PET: principle and validation. *J Nucl Med*. 1998;39:904–911.
31. Sattarivand M, Kusano M, Poon I, Caldwell C. Symmetric geometric transfer matrix partial volume correction for PET imaging: principle, validation and robustness. *Phys Med Biol*. 2012;57:7101–7116.
32. Lockhart SN, Ayakta N, Winer JR, Joie RL, Rabinovici GD, Jagust WJ. Elevated ¹⁸F-AV-1451 PET tracer uptake detected in incidental imaging findings. *Neurology*. 2017;88:1095–1097.
33. Turkheimer FE, Edison P, Pavese N, et al. Reference and target region modeling of [¹¹C]-(R)-PK11195 brain studies. *J Nucl Med*. 2007;48:158–167.
34. Kimura Y, Endo H, Ichise M, et al. A new method to quantify tau pathologies with ¹¹C-PBB3 PET using reference tissue voxels extracted from brain cortical gray matter. *EJNMMI Res*. 2016;6:24.