Biostatistical Estimation of Tau Threshold Hallmarks (BETTH) Algorithm for Human Tau PET Imaging Studies

Alexandra Gogola[1], Brian J. Lopresti[1], Dana Tudorascu[2], Beth Snitz[3], Davneet Minhas[1], Vincent Doré[4,5], Milos D. Ikonomovic[3,6], C. Elizabeth Shaaban[7], Cristy Matan[1], Pierrick Bourgeat[5], N. Scott Mason[1], Howard Aizenstein[2], Chester A. Mathis[1], William E. Klunk[2], Christopher C. Rowe[4], Oscar L. Lopez[3], Ann D. Cohen[2], and Victor L. Villemagne[2,4] for the Alzheimer's Disease Neuroimaging Initiative

[1]Department of Radiology, University of Pittsburgh, Pittsburgh, Pennsylvania; [2]Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania; [3]Department of Neurology, University of Pittsburgh, Pittsburgh, Pennsylvania; [4]Department of Molecular Imaging and Therapy, Austin Health, Melbourne, Victoria, Australia; [5]Health and Biosecurity, Commonwealth Scientific and Industrial Research Organisation, Melbourne, Victoria, Australia; [6]Geriatric Research Education and Clinical Center, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, Pennsylvania; and [7]Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania

A methodology for determining tau PET thresholds is needed to confidently detect early tau deposition. We compared multiple threshold-determining methods in participants who underwent either ^{18}F-flortaucipir or ^{18}F-MK-6240 PET scans. **Methods:** ^{18}F-flortaucipir (n = 798) and ^{18}F-MK-6240 (n = 216) scans were processed and sampled to obtain regional SUV ratios. Subsamples of the cohorts were based on participant diagnosis, age, amyloid-β status (positive or negative), and neurodegeneration status (positive or negative), creating older-adult (age ≥ 55 y) cognitively unimpaired (amyloid-β–negative, neurodegeneration-negative) and cognitively impaired (mild cognitive impairment/Alzheimer disease, amyloid-β–positive, neurodegeneration-positive) groups, and then were further subsampled via matching to reduce significant differences in diagnostic prevalence, age, and Mini-Mental State Examination score. We used the biostatistical estimation of tau threshold hallmarks (BETTH) algorithm to determine sensitivity and specificity in 6 composite regions. **Results:** Parametric double receiver operating characteristic analysis yielded the greatest joint sensitivity in 5 of the 6 regions, whereas hierarchic clustering, gaussian mixture modeling, and k-means clustering all yielded perfect joint specificity (2.00) in all regions. **Conclusion:** When ^{18}F-flortaucipir and ^{18}F-MK-6240 are used, Alzheimer disease–related tau status is best assessed using 2 thresholds, a sensitivity one based on parametric double receiver operating characteristic analysis and a specificity one based on gaussian mixture modeling, delimiting an uncertainty zone indicating participants who may require further evaluation.

Key Words: tau; PET; Alzheimer disease; ^{18}F-flortaucipir; ^{18}F-MK-6240

J Nucl Med 2023; 64:1798–1805
DOI: 10.2967/jnumed.123.265941

Alzheimer disease (AD) pathology, characterized by amyloid-β (Aβ) and hyperphosphorylated tau aggregation, starts accumulating decades before the onset of the clinical dementia phenotype (*1*). The development of imaging and fluid biomarkers has facilitated noninvasive detection and disease progression monitoring. Additionally, such biomarkers have proven useful for predicting clinical progression and the risk of cognitive decline, as well as for determining inclusion, target engagement, and outcome measures in clinical trials.

The Aβ (A), tau (T), and neurodegeneration (N) framework was developed to take advantage of these available imaging and fluid biomarkers and dichotomize them (as positive [+] or negative [−]) to discriminate between non-AD and AD continuum participants (*2*). The establishment of biomarker thresholds depends on the clinical or research question posed, which may benefit from a threshold that favors sensitivity or specificity. For example, if the research question aims at detecting early cortical tau deposition, it would be preferable to adopt a more sensitive threshold. On the other hand, if, for example, there is an antitau therapeutic trial, a more specific threshold would be preferable to ascertain robust levels of cortical tau to evaluate both target engagement and treatment efficacy. Previous work has defined biomarker thresholds for classifying cases as A+ or A− and as N+ or N−, examples of which include dichotomizing neuroimaging measures or cerebral spinal fluid or plasma measures of Aβ and either glucose metabolism or cortical thickness, respectively (*3*).

Although tau accumulation is associated with short-term clinical progression and cognitive decline rate in A+ individuals (*4*,*5*), a consensus method for tau threshold determination remains elusive (*6*). Another important aspect to consider is that clinical studies use many different tau PET tracers, which vary in degree of nonspecific binding, off-target binding, dynamic range, and kinetic behavior (*7*). These tau tracers include ^{18}F-flortaucipir (Tauvid [Eli Lilly and Co.], previously known as ^{18}F-AV-1451 and ^{18}F-T807) (*8*), ^{18}F-MK-6240 (*9*), ^{18}F-THK-5317, ^{18}F-THK-5351, ^{11}C-PBB3, ^{18}F-RO-948, ^{18}F-PI-2620, ^{18}F-GTP1, and ^{18}F-PM-PBB3 (*7*). This diversity complicates imaging result interpretation, tau load measure comparison, and tau PET signal standardization for determining tau status.

There are several potential methods for determining the appropriate tau PET threshold: cluster analysis, receiver operating characteristic (ROC) analysis, iterative outlier detection, z scores (as with CenTauR$_z$ (*10*)), gaussian mixture modeling, and control group percentiles (*3*,*11*). Although it would be ideal to establish a single tau status threshold based on a universal tau load scale, as with centiloids for Aβ (*12*), there have been few head-to-head studies of tau tracers on which to base such a scale (*13*,*14*). One proposed

Received Apr. 26, 2023; revision accepted Aug. 3, 2023.
For correspondence or reprints, contact Alexandra Gogola (alexandragogola@pitt.edu).
Published online Sep. 14, 2023.

scale, termed CenTauR (*10*), defines a universal mask for sampling across tracers and a method for standardizing tau load indices to a common scale. Until a universal scale is adopted, it may be best to establish a methodology for tau threshold determination.

Varying regional sampling strategies have been used to identify pathologic tau accumulation considering the various tau pathologic phenotypes and tracer-specific off-target binding patterns. Previously suggested regions involve combinations of the entorhinal, parahippocampal, middle and inferior temporal, lateral occipital, fusiform, supramarginal gyrus, and inferior parietal regions as well as the amygdala and banks of the superior temporal sulcus (*3*,*15*,*16*). One proposed composite region, based on its being a locus of early AD-related tau deposition, even in the absence of Aβ (*17*), is the mesial temporal lobe, which comprises the entorhinal, parahippocampal, fusiform, and amygdala regions. Because mesial temporal lobe tau deposition is associated with primary age-related tauopathy (*18*), a common pathology in older adults, other composite regions have been proposed that may be more specific for early AD-related tau deposition. One example, the meta temporal region proposed by Jack et al. (*2*), combines the mesial temporal lobe subregions with the middle and inferior temporal gyri.

The present work describes the biostatistical estimation of tau threshold hallmarks (BETTH) algorithm: our examination of different biostatistical approaches to establish tau thresholds for the 2 most widely used tau agents, ^{18}F-flortaucipir and ^{18}F-MK-6240, as tested in 6 composite regions and against clinically diagnosed test cases with the goal of identifying the threshold-determining methods that yield the greatest regional sensitivity and specificity for early AD-related tau deposition. Specifically, we aim at potentially detecting cortical tau deposition at the presymptomatic stages of the disease.

MATERIALS AND METHODS

Human Participants

This study was approved by the Institutional Review Boards of all participating institutions. Informed written consent was obtained from all participants at each site.

The ^{18}F-flortaucipir data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu/). This dataset was collected by 42 participating sites using 30 different PET scanner models (*19*). The identification numbers of the ADNI participants can be found in Supplemental Table 1 (supplemental materials are available at http://jnm.snmjournals.org). The ^{18}F-MK-6240 dataset was acquired from Cerveau Technologies, Inc., and collected by 4 participating sites using 2 PET scanner models.

All participants were at least 55 y old; underwent ^{18}F-flortaucipir or ^{18}F-MK-6240 tau PET, Aβ PET, structural T1 MRI, and cognitive testing; and received a consensus clinical diagnosis. For the ^{18}F-flortaucipir cohort, Aβ load was reported in centiloids and Aβ status was determined with a cutoff of 20 centiloids for optimal sensitivity (*20*). For the ^{18}F-MK-6240 cohort, categoric Aβ status (A+ or A−) was reported by each participating site. For all participants, neurodegeneration status was determined using MRI-based composite cortical thickness measures (CT composite) applying a cutoff of 2.7 mm for optimal sensitivity (*3*). Cognition was reported as the Mini-Mental State Examination score, and clinical diagnosis was reported as normal cognition with no subjective cognitive complaint, mild cognitive impairment, or AD.

Data Selection

Dichotomized samples of participants from the cohorts were selected to maximize the likelihood that they would be T− or T+. Cognitively unimpaired (CU) individuals were clinically categorized as no-subjective-cognitive-complaint participants who were A− and N−. Cognitively impaired (CI) individuals were clinically categorized as AD or mild-cognitive-impairment participants who were A+ and N+. To limit potential bias from the significant differences ($P < 0.05$) in age, cognition, and clinical diagnosis between the dichotomized samples (Supplemental Table 2), we subsampled the dichotomized ^{18}F-MK-6240 set by selecting only participants who had continuous measures of Aβ load (centiloid scores), collected under the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing. Then, using the R (version 4.2.1) (*21*) MatchIt tool (*22*) and variable ratio matching, we drew participants from the dichotomized ^{18}F-flortaucipir set to match the subsampled ^{18}F-MK-6240 set on the basis of age, Mini-Mental State Examination score, global centiloid indices, and CT composite indices, with secondary consideration given to the distributions of sex and diagnostic prevalence. Figure 1 shows the subsampling process.

Data Analyses

Brain parcellation atlases for PET image sampling were obtained from the MR images as previously described (*23*,*24*) using FreeSurfer

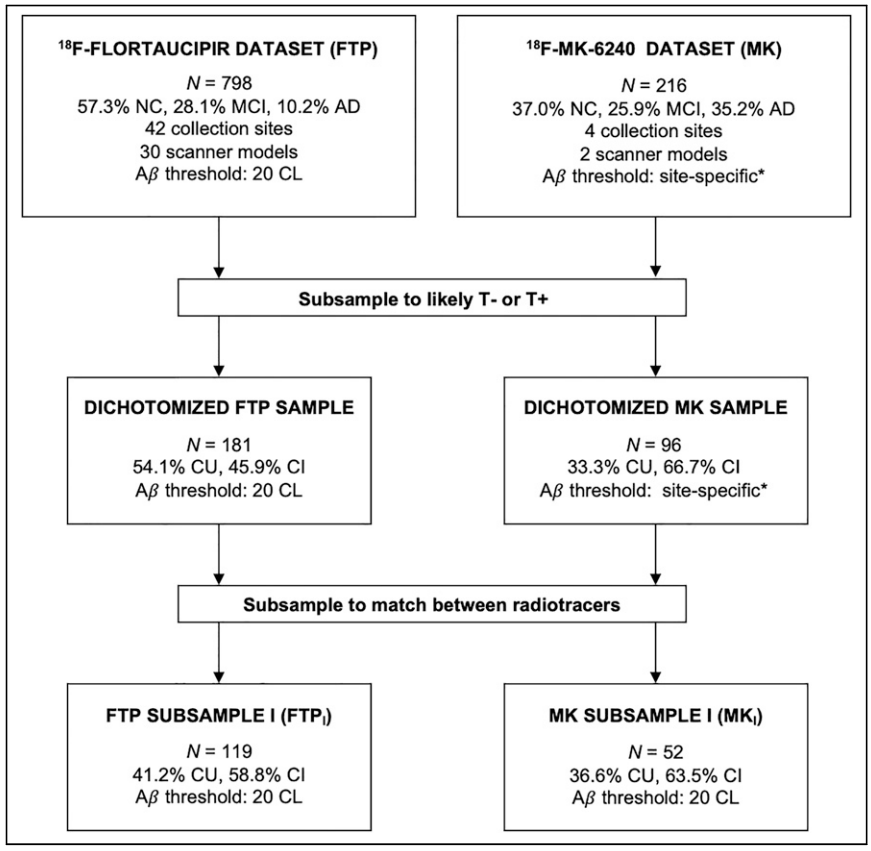

FIGURE 1. Subsampling process for threshold testing. *Aβ status was reported by each site depending on its individual method for threshold determination. CL = centiloids; MCI = mild cognitive impairment; NC = normal cognition with no subjective cognitive complaint.

version 7.1. Images were summed over 80–100 min for ^{18}F-flortaucipir (*25*,*26*) and 90–110 min for ^{18}F-MK-6240 (*27*). ^{18}F-flortaucipir images, processed by the Commonwealth Scientific and Industrial Research Organization Health and Biosecurity Business Unit, Department of Industry Innovation and Science, Australian Government, were rigidly aligned to their corresponding MR images using the robust block-matching registration software Mirorr (*28*). ^{18}F-MK-6240 images were registered to their corresponding MR images using rigid-body registration with the normalized mutual information cost function (*29*). All registrations were manually inspected for accuracy. The FreeSurfer parcellation template (Desikan-Killiany-Tourville atlas (*30*)) was used to sample summed PET images, and a volume-weighted average of FreeSurfer region counterparts (Table 1) was calculated for each of the 6 composite regions investigated: amygdala, inferior temporal region, lateral occipital region, lateral temporal region, mesial temporal region, and meta temporal region (*3*,*15*,*16*). SUV ratios were calculated using cerebellar gray matter as a reference (*9*,*31*). A CT composite was calculated as a surface area–weighted composite of the FreeSurfer-derived entorhinal, fusiform, inferior temporal, and middle temporal lobe CT values (*3*). To preserve the raw image data, no harmonization methods were applied to the reported data. However, to evaluate potential site- and scanner-specific effects, we conducted the same analyses on adjusted SUV ratios from the dichotomized subsample adjusted using the R ComBat tool (*32*), a harmonization tool designed to minimize scanner effects.

Methods for Tau Threshold Determination

Each subsample (subsampled ^{18}F-flortaucipir set, subsampled ^{18}F-MK-6240 set) was randomly split into training and testing sets, preserving matching, using R (70% and 30%, respectively) to avoid overfitting. Using the training set, we evaluated several approaches to determine and compare thresholds for both radiotracers: 90th percentile of CU participants, 95th percentile of CU participants, CenTauR$_z$ 1.5 (mean + 1.5 SDs of the CU group mean) (*10*), CenTauR$_z$ 2.0 (mean + 2 SDs of the CU group mean) (*10*), gaussian mixture modeling (*33*), hierarchic clustering (*34*), iterative outlier detection (*35*), k-means clustering (*36*), ROC method (*37*), nonparametric and parametric double ROC method (*38*), and Youden index (*39*).

Statistical Methods

All statistics were calculated using R. Statistical differences were assessed using a Mann–Whitney *U* test. The ability to discriminate tau signal between CU and CI participants was assessed using effect size (Cohen *d*). Each threshold-determining approach was assessed using sensitivity and specificity by applying the determined thresholds to the testing set.

RESULTS

Table 2 shows the participant characteristics. After matching, there were no statistically significant differences in distributions or between metrics.

Regional SUV ratio means, SDs, and effect sizes are shown in Table 3. For both tracers, the amygdala and the mesial temporal region showed the greatest effect sizes, or differences between the CI and CU means.

Figure 2 displays each tracer's sensitivities—and Figure 3, each tracer's specificities—for the combinations of 12 threshold-determining methods and 6 composite region pairs. When considering joint (subsampled ^{18}F-flortaucipir set plus subsampled ^{18}F-MK-6240 set) sensitivity and specificity, we found that different methods performed optimally for each region. The 90th percentile and parametric double ROC methods in the mesial temporal region yielded the highest joint sensitivity (1.86). The 90th percentile, CenTauR$_z$ 1.5, nonparametric and parametric double ROC methods in the amygdala, and the parametric double ROC method in the inferior temporal region were slightly lower (1.81), followed by the parametric double ROC method in the meta temporal region (1.71), the Youden index in the lateral temporal region (1.70), and the parametric double ROC method in the lateral occipital region (1.65). Gaussian mixture modeling, hierarchic clustering, and k-means clustering each had a joint sensitivity of 2.00 in all regions with the addition of CenTauR$_z$ 2 in the amygdala and iterative outlier detection in the amygdala, mesial temporal region, and meta temporal region.

Application of the parametric double ROC method for a sensitivity threshold and gaussian mixture modeling for a specificity

TABLE 1
Composite Regions of Interest

Composite region	Region	Sub-region	FreeSurfer region
Meta temporal	Mesial temporal	Amygdala	L amygdala
			R amygdala
			L entorhinal
			R entorhinal
			L fusiform
			R fusiform
			L parahippocampal
			R parahippocampal
	Lateral temporal	Inferior temporal	L inferior temporal
			R inferior temporal
			L middle temporal
			R middle temporal
		Lateral occipital	L lateral occipital
			R lateral occipital

TABLE 2
Characteristics of CU and CI Participants for ^{18}F-Flortaucipir and ^{18}F-MK-6240 Matched Subsamples

Characteristic	Subsampled ^{18}F-flortaucipir set		Subsampled ^{18}F-MK-6240 set	
	A−N− CU	A+N+ CI	A−N− CU	A+N+ CI
n	49 (41.2%)	70 (58.8%)	19 (36.5%)	33 (63.5%)
Sex	49.0% F; 51.0% M	42.9% F; 57.1% M	47.4% F; 52.6% M	45.5% F; 54.5% M
AD/MCI	—	47.1% AD; 52.9% MCI	—	51.5% AD; 48.5% MCI
Age (y)	72.78 ± 4.79	76.70 ± 7.85	73.89 ± 4.76	74.06 ± 6.84
Mini-Mental State Examination score	29.12 ± 1.00	24.46 ± 4.02	28.47 ± 1.27	24.61 ± 2.95
CT composite (mm)	2.79 ± 0.07	2.54 ± 0.11	2.78 ± 0.07	2.56 ± 0.11
Centiloids	2.84 ± 8.61	88.54 ± 28.70	2.82 ± 8.09	98.92 ± 26.29

MCI = mild cognitive impairment.

threshold yielded the following respective threshold pairs: 1.30 and 1.58 for ^{18}F-flortaucipir and 0.78 and 1.29 for ^{18}F-MK-6240 in the amygdala, 1.26 and 1.48 for ^{18}F-flortaucipir and 1.09 and 1.63 for ^{18}F-MK-6240 in the inferior temporal region, 1.15 and 1.42 for ^{18}F-flortaucipir and 1.11 and 1.73 for ^{18}F-MK-6240 in the lateral occipital region, 1.24 and 1.49 for ^{18}F-flortaucipir and 1.07 and 1.63 for ^{18}F-MK-6240 in the lateral temporal region, 1.23 and 1.33 for ^{18}F-flortaucipir and 1.01 and 1.46 for ^{18}F-MK-6240 in the mesial temporal region, and 1.18 and 1.36 for ^{18}F-flortaucipir and 1.04 and 1.53 for ^{18}F-MK-6240 in the meta temporal region.

When all 72 combinations of thresholding methods and sample regions were considered, neither tracer showed a sensitivity advantage, with ^{18}F-flortaucipir showing a higher sensitivity than ^{18}F-MK-6240 in 34 of 72 (47.2%) comparisons, ^{18}F-MK-6240 showing a higher sensitivity than ^{18}F-flortaucipir in 33 of 72 (45.8%), and the two being equivalent in 5 of 72 (6.9%). ^{18}F-flortaucipir outperformed ^{18}F-MK-6240 in specificity, showing an advantage in 39 of 72 (54.2%) comparisons, but the specificity was evenly matched in 23 of 72 (31.4%) comparisons.

The relative performance of threshold method sensitivities or specificities was not modified in either the unmatched dichotomized samples (Supplemental Table 3; Supplemental Figs. 1 and 2) or the ComBat harmonized samples (Supplemental Table 4; Supplemental Figs. 3 and 4).

DISCUSSION

When considering AD biomarker thresholds, it is important to note that the same threshold may not be optimal in all circumstances (*11*). Indeed, for the centiloid scale, a value of approximately 20 centiloids results in the greatest sensitivity for detecting early Aβ pathology (sensitivity threshold), but a threshold of

TABLE 3
Matched Composite Region Summary

Region	CU SUV ratio		CI SUV ratio		
	Mean	σ^2	Mean	σ^2	*d*
Subsampled ^{18}F-flortaucipir set					
Amygdala	1.18	0.11	1.64	0.33	1.87
Inferior temporal	1.18	0.09	1.67	0.47	1.43
Lateral occipital	1.10	0.09	1.35	0.38	0.90
Lateral temporal	1.17	0.08	1.63	0.46	1.39
Mesial temporal	1.15	0.08	1.52	0.31	1.63
Meta temporal	1.13	0.08	1.40	0.32	1.19
Subsampled ^{18}F-MK-6240 set					
Amygdala	0.67	0.06	1.770	0.67	2.30
Inferior temporal	1.06	0.07	1.936	0.88	1.41
Lateral occipital	1.12	0.11	1.673	0.84	0.93
Lateral temporal	1.04	0.07	1.834	0.83	1.35
Mesial temporal	0.94	0.07	1.808	0.68	1.80
Meta temporal	1.00	0.07	1.821	0.75	1.54

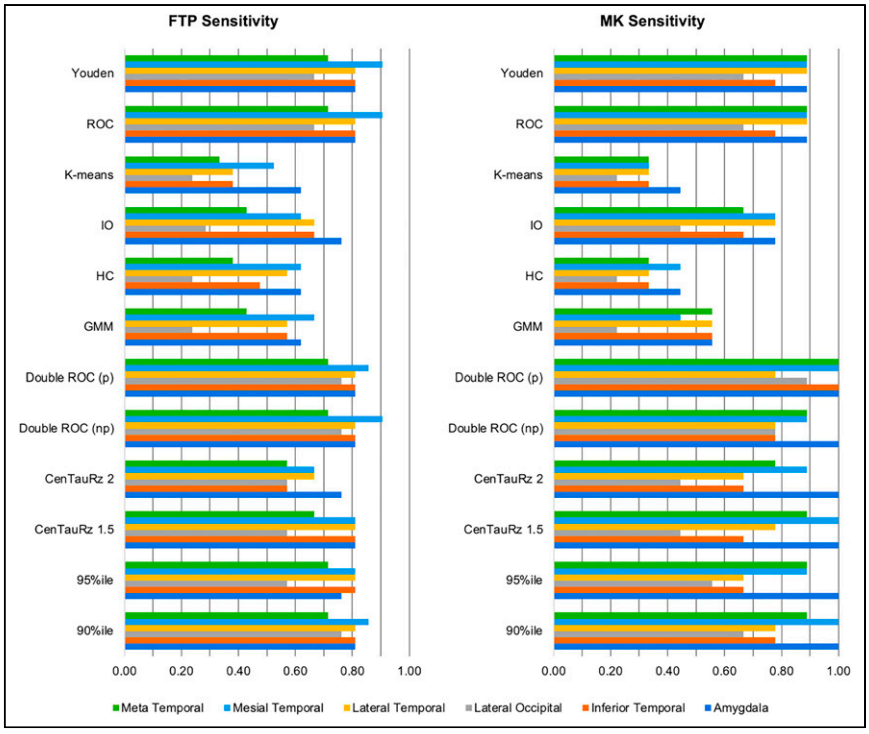

FIGURE 2. Matched threshold sensitivities for 6 composite regions when using each of 12 threshold-determining methods. CenTauR$_z$ 1.5 = mean + 1.5 SDs of CU group mean; CenTauR$_z$ 2.0 = mean + 2 SDs of CU group mean; GMM = gaussian mixture modeling; HC = hierarchic clustering; IO = iterative outlier detection.

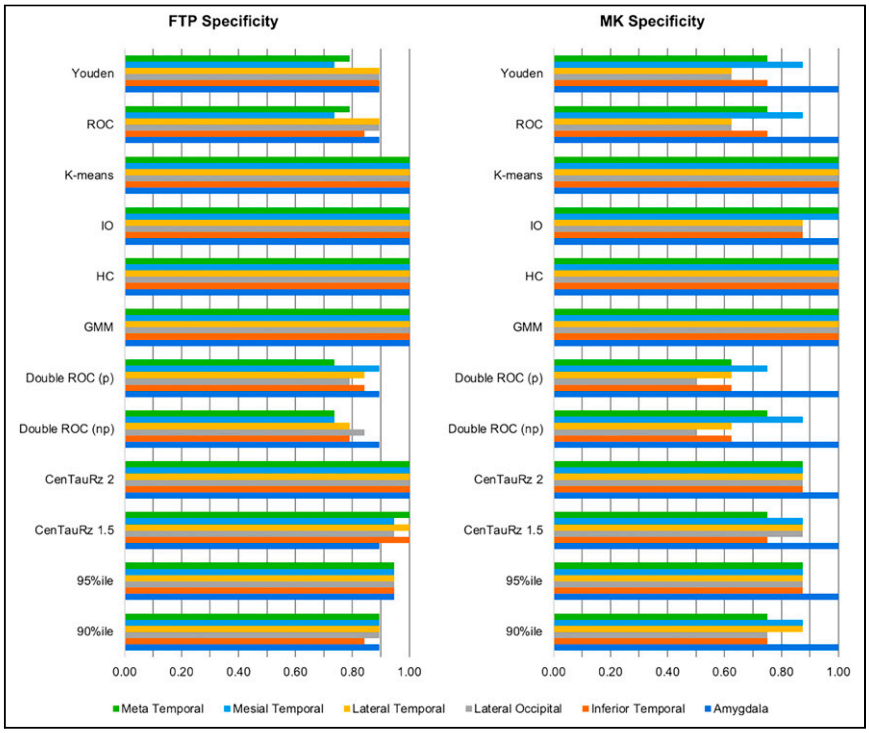

FIGURE 3. Matched threshold sensitivities and specificities for 6 composite regions when using each of 12 threshold-determining methods. CenTauR$_z$ 1.5 = mean + 1.5 SDs of CU group mean; CenTauR$_z$ 2.0 = mean + 2 SDs of CU group mean; GMM = gaussian mixture modeling; HC = hierarchic clustering; IO = iterative outlier detection.

approximately 50 centiloids better predicts the neuropathologic and clinical diagnosis of AD (specificity threshold) (*40*). Because the initial goal of the present work was to identify the threshold-determining methods that yield the greatest sensitivity and specificity for early AD-related tau deposition, and given the level of disagreement between the sensitivities and specificities of the various threshold-determining methods, it may be prudent to adopt distinct tau thresholds optimized for either sensitivity or specificity for AD-related tau as demanded by the application, similar to what has been proposed for establishing Aβ positivity thresholds using the centiloid scale (*41*).

In assessing the 72 threshold-determining-method and composite-region pairs for joint sensitivity between tracers, we found that none of the threshold methods yielded the highest sensitivity for every region, although the parametric double ROC method showed the highest joint sensitivity for all but the lateral temporal region, where it was slightly lower (0.11) than the Youden or ROC method (Fig. 2). For joint specificity, there was even more overlap in optimal methods for each region. Gaussian mixture modeling, hierarchic clustering, and k-means clustering all yielded perfect joint specificities (2.0) for all 6 regions examined. We believe that the difference in optimal sensitivity and specificity threshold-determining methods can be attributed, in part, to the regional SUV ratio distributions between the CU and CI samples for each tracer (Supplemental Figs. 5 and 6). Given that a sizeable proportion of CI participants presents with low tau, as previously reported (*42*,*43*), even after dichotomizing and matching of samples, there was significant overlap in SUV ratios between CU and CI participants, with clearly identifiable groups of lower and higher tau load within the CI participants. These distributions pose a challenge in fully separating the CI participants from the CU participants. Compared with the other methods tested, the parametric double ROC method is well positioned to set the optimal sensitivity threshold because it does not rely solely on the CU group, as many others do, which ultimately favors specificity. Further, the threshold in this specific instance was based on achieving an accuracy of 95%, and given the increased prevalence of the CI participants here (>58%), sensitivity was inherently targeted over specificity. Between the success of the parametric double ROC method across regions and its innate advantages, this method can be used to set the sensitivity

threshold regardless of region. Conversely, gaussian mixture modeling, hierarchic clustering, and k-means clustering—being clustering methods—proved advantageous for specificity. They cluster lower tau load as one group and higher tau load as a second group, setting the threshold at the split between low- and high-tau-load CI participants, grouping all CU participants with the low-tau CI participants. Although gaussian mixture modeling, hierarchic clustering, and k-means clustering all yield perfect specificity, gaussian mixture modeling may be the best single method to use across all regions because it sets the lowest perfect specificity threshold.

Several composites have been suggested as potential sampling regions for determining tau status. Although the algorithm used in this work—BETTH—does not determine the optimal region for evaluating tau status, it does highlight their differences, particularly with respect to the chosen sensitivity method (parametric double ROC method). The mesial temporal region yielded the greatest sensitivity, closely followed by the amygdala and the inferior temporal region. As shown in Table 1, the mesial temporal region comprises a combination of Braak I and III regions, and the amygdala is one of the Braak III regions. Thus, it is unsurprising that they would have greater effect sizes as shown in Table 3, given that these regions start accumulating tau earlier than do the inferior temporal region, lateral temporal region, and meta temporal region (regions included in Braak IV) or the lateral occipital region (Braak V). In fact, the mesial temporal lobe is frequently a site of early tau deposition in the AD pathologic process, before the development of cognitive symptoms (*17*). However, its association with primary age-related tauopathy raises concerns about its ability to differentiate AD-related from age-related tau deposition (*18*). Further, the Braak and Delacourte stages capture, to varying degrees, the neuropathologic distribution of tau (*44*,*45*), but their application to PET studies may fail to consistently capture tau deposition because the regions are either too small and subject to partial-volume effects (*46*,*47*), as may be of concern for the amygdala, or too large and subject to dilution of focal specific PET signal (*47*). Although the meta temporal region (*3*) may not be able to capture some atypical AD presentations such as posterior cortical atrophy (*48*), it seems capable of detecting early tau deposition and discriminating between AD-related and age-related tau deposition through its combination of the mesial temporal lobe and the inferior and middle temporal neocortical regions (*43*), the latter of which is associated with cortical AD tau pathology (Braak IV and later) but not primary age-related tauopathy. Further, the meta temporal region captures the 3 AD pathologic subtypes (*49*), as well as nonstereotypic tau deposition (*50*), while discriminating AD from non-AD neurodegeneration (*43*) and is the most sensitive for longitudinal analysis of tau deposition (*51*,*52*).

Our previous work compared the relative performance of ^{18}F-flortaucipir and ^{18}F-MK-6240 in terms of visual assessments, off-target binding, and dynamic range, where we found ^{18}F-MK-6240 to have approximately a 2-fold greater dynamic range and lower nonspecific binding than ^{18}F-flortaucipir across Braak pathologic stage regions (*14*), suggesting that ^{18}F-MK-6240 may be more useful for detecting early tau signal and small interval changes (*53*). The differences in dynamic range are of particular interest for the determination of tau thresholds. When comparing the 2 tracers' performance in differentiating between high and low tau load using the optimal specificity threshold, as set by gaussian mixture modeling, we found that ^{18}F-flortaucipir and ^{18}F-MK-6240 performed equivalently, each yielding perfect specificities. However, when comparing sensitivities, as determined using the parametric double ROC method, we found that ^{18}F-MK-6240 outperformed ^{18}F-flortaucipir by a mean sensitivity of 0.15 ± 0.10 in all regions but the lateral temporal. This characteristic may support the use of ^{18}F-MK-6240 over ^{18}F-flortaucipir, particularly when sensitivity to early tau deposition is paramount.

The implementation of the BETTH threshold-determining approach in the present study has the benefit of being evaluated for 2 tau tracers (in datasets selected to be as cleanly impaired or unimpaired as possible) and for matched samples. However, the present study has some limitations. In subsampling the CU and CI participants and matching, we reduced the sample sizes 2-fold and eliminated more ambiguously classified participants. Additionally, part of the selection process relied on third-party reporting either of clinical diagnosis or Aβ status, which likely used different criteria and methods for determination. Future work will need to assess the BETTH approach in expanded datasets without selection bias. Because race, ethnicity, education, and apolipoprotein E status were not provided with the original data, those factors could not be considered in our matching process, leaving any existing bias in our matched cohorts. Future work should conduct the same analyses on datasets in which those demographics are included and can be accounted for. The success of the thresholds was evaluated relative to the assumption that CU participants were T− and CI participants were T+. Future work would benefit from evaluating the success of the thresholds relative to visual reads, although visual identification of early tau deposits, while more sensitive (*54*), may be complicated by off-target signal (*55*). The goal of the present work was to establish a universal method for determining tau thresholds, but only ^{18}F-flortaucipir and ^{18}F-MK-6240 were included in our evaluation. To ensure the generalizability of our methods, it will be necessary to extend BETTH to other tau tracers to validate the conclusions drawn here. Further, these thresholds reflect the idiosyncratic pharmacokinetic and pharmacologic properties (affinity, nonspecific binding, etc.) of the tracers used, as well as small differences introduced using PET scanners with different sensitivities and resolutions, reflecting the situation today and for these 2 tau tracers. The introduction and implementation of digital PET scanners, with greater sensitivities and higher resolutions, will require the revision of these thresholds under the new conditions. Finally, the whole cerebellar gray matter was used as a reference region, to keep processing uniform between both radiotracers. However, the calculated SUV ratios may be impacted by the off-target binding characteristics of each tracer. As progress is made toward a common scale indexing tau load, it will be necessary to investigate the effects that universal tau cortical and reference regions that account for the off-target binding of tracers (*56*) have on the resultant thresholds.

CONCLUSION

Given that cortical tau is intimately and imminently associated with subsequent neurodegeneration and cognitive decline, it is imperative to detect AD-related cortical tau accumulation as early as possible. We have proposed the BETTH algorithm for assessing different approaches in various composite regions for determination of the optimal threshold to detect early tau deposition. Although the present work evaluated several threshold-determining methods and tissue-sampling strategies, BETTH is extensible and can be applied to other biomarkers. At this time, the present study suggests that AD-related tau status is best determined using a parametric double ROC–determined sensitivity threshold and a gaussian mixture modeling–determined specificity threshold. Each pair represents an uncertainty zone indicating participants who may need further evaluation.

DISCLOSURE

This work was supported financially by National Institute of Aging grants P50 AG005133, RF1 AG025516, R01 AG052446, R01 AG052521, and P01 AG025204. ^{18}F-flortaucipir data collection and sharing for this project were funded by the ADNI (National Institutes of Health grant U01 AG024904) and DOD ADNI (Department of Defense award W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Araclon Biotech, BioClinica, Inc., Biogen, Bristol-Myers Squibb Company, CereSpir, Inc., Cogstate, Eisai Inc., Elan Pharmaceuticals, Inc., Eli Lilly and Company, EuroImmun, F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc., Fujirebio, GE Healthcare, IXICO Ltd., Janssen Alzheimer Immunotherapy Research & Development, LLC, Johnson & Johnson Pharmaceutical Research & Development, LLC, Lumosity, Lundbeck, Merck & Co., Inc., Meso Scale Diagnostics, LLC, NeuroRx Research, Neurotrack Technologies, Novartis Pharmaceuticals Corporation, Pfizer Inc., Piramal Imaging, Servier, Takeda Pharmaceutical Company, and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. ^{18}F-MK-6240 data were supplied by Cerveau Technologies. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

Data used in preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The investigators within the ADNI contributed to the design and implementation of ADNI or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators is available online (https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

KEY POINTS

QUESTION: Can a single threshold-determining method yield optimal sensitivity and specificity for both ^{18}F-flortaucipir and ^{18}F-MK-6240?

PERTINENT FINDINGS: A comparison of 12 threshold-determining methods in 6 composite regions revealed that no method performed optimally in terms of both sensitivity and specificity. In most cases, doing particularly well in terms of sensitivity meant doing poorly in terms of specificity and vice versa, likely because of the degree of overlap in the unimpaired and impaired participant groups.

IMPLICATIONS FOR PATIENT CARE: Given the prevalence of low tau in Aβ-positive CI individuals, no single threshold for tau can be both highly sensitive and highly specific. Therefore, the threshold selected will depend on the clinical or research question posed.

REFERENCES

1. Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* 2013;12:357–367.
2. Jack CR, Bennett DA, Blennow K, et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology.* 2016;87:539–547.
3. Jack CR, Wiste HJ, Weigand SD, et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement.* 2017;13:205–216.
4. Ossenkoppele R, Pichet Binette A, Groot C, et al. Amyloid and tau PET-positive cognitively unimpaired individuals are at high risk for future cognitive decline. *Nat Med.* 2022;28:2381–2387.
5. Strikwerda-Brown C, Hobbs DA, Gonneaud J, et al. Association of elevated amyloid and tau positron emission tomography signal with near-term development of Alzheimer disease symptoms in older adults without cognitive impairment. *JAMA Neurol.* 2022;79:975–985.
6. Schöll M, Lockhart SN, Schonhaut DR, et al. PET imaging of tau deposition in the aging human brain. *Neuron.* 2016;89:971–982.
7. Leuzy A, Chiotis K, Lemoine L, et al. Tau PET imaging in neurodegenerative tauopathies: still a challenge. *Mol Psychiatry.* 2019;24:1112–1134.
8. Chien WT, Leung SF, Yeung FK, Wong WK. Current approaches to treatments for schizophrenia spectrum disorders, part II: psychosocial interventions and patient-focused perspectives in psychiatric care. *Neuropsychiatr Dis Treat.* 2013;9:1463–1481.
9. Hostetler ED, Walji AM, Zeng Z, et al. Preclinical characterization of ^{18}F-MK-6240, a promising PET tracer for in vivo quantification of human neurofibrillary tangles. *J Nucl Med.* 2016;57:1599–1606.
10. Villemagne VL, Leuzy A, Bohorquez SS, et al. CenTauR: toward a universal scale and masks for standardizing tau imaging studies. *Alzheimers Dement (Amst).* 2023;15:e12454.
11. Villemagne VL, Lopresti BJ, Doré V, et al. What Is T+? A Gordian Knot of Tracers, Thresholds, and Topographies. *J Nucl Med.* 2021;62:614–619.
12. Klunk WE, Koeppe RA, Price JC, et al. The centiloid project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement.* 2015;11:1–15.e4.
13. Smith R, Schöll M, Leuzy A, et al. Head-to-head comparison of tau positron emission tomography tracers [^{18}F]flortaucipir and [^{18}F]RO948. *Eur J Nucl Med Mol Imaging.* 2020;47:342–354.
14. Gogola A, Minhas DS, Villemagne VL, et al. Direct comparison of the tau PET tracers ^{18}F-flortaucipir and ^{18}F-MK-6240 in human subjects. *J Nucl Med.* 2022;63:108–116.
15. Mishra S, Gordon BA, Su Y, et al. AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: defining a summary measure. *Neuroimage.* 2017;161:171–178.
16. Maass A, Landau S, Baker SL, et al. Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. *Neuroimage.* 2017;157:448–463.
17. Weigand AJ, Bangen KJ, Thomas KR, et al. Is tau in the absence of amyloid on the Alzheimer's continuum? A study of discordant PET positivity. *Brain Commun.* 2020;2:fcz046.
18. Crary JF, Trojanowski JQ, Schneider JA, et al. Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathol (Berl).* 2014;128:755–766.
19. Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's Disease Neuroimaging Initiative 3: continued innovation for clinical trial improvement. *Alzheimers Dement.* 2017;13:561–571.
20. Rowe C, Amadoru S, Dore V, et al. Correlation of amyloid PET in centiloid units with neuropathological findings in Alzheimer's disease [abstract]. *J Nucl Med.* 2018;59(suppl 1):482.
21. R: a language and environment for statistical computing. The R Foundation website. https://www.r-project.org/. Accessed August 25, 2023.
22. Ho DEH, Imai K, King G, Stuart EA. Nonparametric preprocessing for parametric causal inference. *J Stat Softw.* 2011;42:1–28.
23. Okonkwo DO, Puffer RC, Minhas DS, et al. [^{18}F]FDG, [^{11}C]PiB, and [^{18}F]AV-1451 PET imaging of neurodegeneration in two subjects with a history of repetitive trauma and cognitive decline. *Front Neurol.* 2019;10:831.
24. Fischl B, van der Kouwe A, Destrieux C, et al. Automatically parcellating the human cerebral cortex. *Cereb Cortex.* 2004;14:11–22.
25. Shcherbinin S, Schwarz AJ, Joshi A, et al. Kinetics of the tau PET tracer ^{18}F-AV-1451 (T807) in subjects with normal cognitive function, mild cognitive impairment, and Alzheimer disease. *J Nucl Med.* 2016;57:1535–1542.
26. Baker SL, Lockhart SN, Price JC, et al. Reference tissue–based kinetic evaluation of ^{18}F-AV-1451 for tau imaging. *J Nucl Med.* 2017;58:332–338.
27. Pascoal TA, Shin M, Kang MS, et al. In vivo quantification of neurofibrillary tangles with [^{18}F]MK-6240. *Alzheimers Res Ther.* 2018;10:74.

28. Rivest-Hénault D, Dowson N, Greer PB, Fripp J, Dowling JA. Robust inverse-consistent affine CT-MR registration in MRI-assisted and MRI-alone prostate radiation therapy. *Med Image Anal.* 2015;23:56–69.
29. Studholme C, Hawkes DJ, Hill DLG. Normalized entropy measure for multimodality image alignment. Presented at: Medical Imaging 1998: Image Processing; June 24, 1998; San Diego, CA.
30. Klein A, Tourville J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci.* 2012;6:171.
31. Marquié 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.
32. Orlhac F, Eertink JJ, Cottereau AS, et al. A guide to ComBat harmonization of imaging biomarkers in multicenter studies. *J Nucl Med.* 2022;63:172–179.
33. Reynolds DA. Gaussian mixture models. In: *Encyclopedia of Biometrics.* Springer; 2009:741.
34. Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. *WIREs Data Mining Knowl Discov.* 2012;2:86–97.
35. Aizenstein HJ, Nebes RD, Saxton JA, et al. Frequent amyloid deposition without significant cognitive impairment among the elderly. *Arch Neurol.* 2008;65:1509–1517.
36. Likas A, Vlassis N, Verbeek JJ. The global k-means clustering algorithm. *Pattern Recognit.* 2003;36:451–461.
37. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.
38. Greiner M, Sohr D, Göbel P. A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *J Immunol Methods.* 1995;185:123–132.
39. Thiele C, Hirschfeld G. cutpointr: improved estimation and validation of optimal cutpoints in R. *J Stat Softw.* 2021;98:1–27.
40. Amadoru S, Doré V, McLean CA, et al. Comparison of amyloid PET measured in centiloid units with neuropathological findings in Alzheimer's disease. *Alzheimers Res Ther.* 2020;12:22.
41. van der Kall LM, Truong T, Burnham SC, et al. Association of β-amyloid level, clinical progression, and longitudinal cognitive change in normal older individuals. *Neurology.* 2021;96:e662–e670.
42. Rowe CC, Doré V, Krishnadas N, et al. Tau imaging with ^{18}F-MK6240 across the Alzheimer's disease spectrum. medRxiv website. https://www.medrxiv.org/content/10.1101/2022.02.13.22270894v1. Published February 15, 2022. Accessed August 25, 2023.
43. Ossenkoppele R, Rabinovici GD, Smith R, et al. Discriminative accuracy of [^{18}F]flortaucipir positron emission tomography for Alzheimer disease vs other neurodegenerative disorders. *JAMA.* 2018;320:1151–1162.
44. Braak H, Braak E. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging.* 1995;16:271–278.
45. Delacourte A, David JP, Sergeant N, et al. The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer's disease. *Neurology.* 1999;52:1158–1165.
46. Schmidt ME, Chiao P, Klein G, et al. The influence of biological and technical factors on quantitative analysis of amyloid PET: points to consider and recommendations for controlling variability in longitudinal data. *Alzheimers Dement.* 2015;11:1050–1068.
47. Schwarz AJ, Shcherbinin S, Slieker LJ, et al. Topographic staging of tau positron emission tomography images. *Alzheimers Dement (Amst).* 2018;10:221–231.
48. Ossenkoppele R, Schonhaut DR, Schöll M, et al. Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain.* 2016;139:1551–1567.
49. Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol.* 2011;10:785–796.
50. Seemiller J, Bischof GN, Hoenig MC, et al. Indication of retrograde tau spreading along Braak stages and functional connectivity pathways. *Eur J Nucl Med Mol Imaging.* 2021;48:2272–2282.
51. Schwarz C, Therneau T, Przybelski S, et al. Tau positivity: comparing flortaucipir meta-ROI vs. maximum of regional Z-scores. Presented at: 14th Human Amyloid Imaging; January 17, 2020; Miami, FL.
52. Schwarz CG, Therneau TM, Weigand SD, et al. Selecting software pipelines for change in flortaucipir SUVR: balancing repeatability and group separation. *Neuroimage.* 2021;238:118259.
53. Pascoal TA, Benedet AL, Tudorascu DL, et al. Longitudinal ^{18}F-MK-6240 tau tangles accumulation follows Braak stages. *Brain.* 2021;144:3517–3528.
54. Provost K, Iaccarino L, Soleimani-Meigooni DN, et al. Comparing ATN-T designation by tau PET visual reads, tau PET quantification, and CSF PTau181 across three cohorts. *Eur J Nucl Med Mol Imaging.* 2021;48:2259–2271.
55. Fleisher AS, Pontecorvo MJ, Devous MD Sr, et al. Positron emission tomography imaging with [^{18}F]flortaucipir and postmortem assessment of Alzheimer disease neuropathologic changes. *JAMA Neurol.* 2020;77:829–839.
56. Minhas D, Gogola A, Lopresti B, et al. Considerations for a universal tau PET reference region. Presented at: Alzheimer's Association Tau 2022 Global Conference; February 22–23, 2022; online conference.