Journal of Nuclear Medicine, published on September 29, 2022 as doi:10.2967/jnumed.122.264371

# A Visual Interpretation Algorithm for Assessing Brain Tauopathy with18-F MK-6240 Positron Emission Tomography

John P. Seibyl (1,2), Jonathan M. DuBois (3), Annie Racine (3), Jessica Collins (3), Qi Guo (4), Dustin Wooten (3), Eddie Stage (4), David Cheng (2), Roger N. Gunn (2), Lilly Porat (2), Alex Whittington (2), Phillip H. Kuo (5), Masanori Ichise (6), Robert Comley (4), Laurent Martarello (3), Cristian Salinas (3)

[1]Institute for Neurodegenerative Disorders, New Haven, CT
[2]Invicro, New Haven, CT
[3]Biogen, Cambridge, MA
[4]AbbVie, North Chicago, IL
[5]Univ. of Arizona, Tucson, AZ
[6] National Institute of Radiological Sciences
National Institutes for Quantum and Radiological Science and Technology, Chiba, Japan

Correspondence: John P. Seibyl, MD

jpseibyl@seibyl.com

+1-203-887-6574

Figures:4
Tables:5

Running Title: Visual Interpretation of 18-F MK-6240

Disclosures

John P. Seibyl is consultant to Biogen and Distinguished Scientist at Invicro.
David Cheng, Roger N. Gunn, Lilly Porat, and Alex Whittington are employees of Invicro.
Jonathan M. DuBois, Annie Racine, Jessica Collins, Laurent Martarello, and Cristian Salinas are employed by Biogen.
Qi Guo, Dustin Wooten, Eddie Stage, and Robert Comley are employed by AbbVie.
Phillip Kuo is a consultant and employed by Invicro.
Masanori Ichise is a consultant.
No other potential conflicts of interest relevant to this article exist.

**ABSTRACT**

Introduction: In vivo characterization of pathologic deposition of tau protein in the human brain by PET imaging is a promising tool in drug development trials of Alzheimer's disease (AD). 18-F MK-6240 is a radiotracer with high selectivity and sub-nanomolar affinity for neurofibrillary tangles (NFTs) that shows favorable non-specific brain penetration and excellent kinetic properties. The purpose of the present investigation was to develop a visual assessment method that provides both an overall assessment of brain tauopathy and regional characterization of abnormal tau deposition.

Methods: 18-F MK-6240 scans from 102 participants (including healthy volunteers, patients with AD or other neurodegenerative disorders) were reviewed by an expert nuclear medicine physician blind to participants' diagnosis to identify common patterns of brain uptake. This initial visual read method was field tested in a separate, non-overlapping cohort of 102 participants with 2 additional naïve readers trained on the method. Visual read outcomes were compared with semiquantitative assessments using volume of interest (VOI) SUVr.

Results: For the visual read eight gray-matter regions per hemisphere were assessed as negative (no abnormal uptake) or positive (1-25% of the region involved, 25-75% involvement, or >75% involvement), then characterized the tau binding pattern as positive or negative for evidence of tau, and if positive, whether brain uptake is in an AD pattern. Readers demonstrated agreement 94% of the time for overall positivity/negativity. Concordance on the determination of the regional binary outcomes (negative or positive) showed agreement of 74.3 % and FIeiss' k of 0.912. Using clinical diagnosis as the ground truth, readers demonstrated a sensitivity of 73-79% and specificity of 91-93, with combined reader concordance sensitivity of 80% and specificity 93%. Cortical regions average SUVr showed a robust correlation with visually-derived ratings of regional involvement ($r$ = .73, $p$ < .0001).

Conclusion: We developed a visual read algorithm for 18-F MK-6240 PET offering both determination of scan positivity and the regional degree of cortical involvement. These cross-sectional results show strong inter-reader concordance on both binary and regional assessments of tau deposition, as well as good sensitivity and excellent specificity supporting use as a tool for clinical trials.

## INTRODUCTION

The recent introduction of tau PET imaging biomarkers for clinical and research applications provides a powerful tool for corroborating the patterns of pathologic progression in Alzheimer's disease (AD) suggested by post mortem studies (1), as well as potentially offering a means for monitoring response to treatments designed to interrupt AD brain pathology (2, 3). Tau PET imaging shows good correlation between regional brain uptake and clinical and psychometric measures in cross-sectional studies.(4-7). Indeed, increased density and spread of abnormal uptake in tau PET images is consistent with progression of the disease occurring in early and mild AD patients and is associated with the degree of neuropsychological impairment.

Tau PET studies have corroborated findings from pathologic post-mortem examination of AD brains, which demonstrated initial cortical uptake in entorhinal cortex and medial temporal structures, extending to inferolateral temporal, superolateral temporal, and neocortical occipital, posterior cingulate, parietal, and frontal cortices (8). Temporal lobe structures, especially mesial temporal gyri and hippocampus are the earliest neocortical regions to manifest neurofibrillary tangles (9), suggesting visual read methods might focus particular attention here. The first- generation tau PET agent 18-F AV-1451, which was utilized in the largest clinicopathological study of AD and mild cognitive impairment (MCI) (10), has off target uptake in areas adjacent to the mesial temporal lobe, limiting the ability to assess tau pathology of this important region.

18-F MK-6240 is a second-generation tau PET imaging agent with high specificity and low off target binding in gray and white matter, representing improvements over first-generation tracers (11, 12). In cognitively normal individuals, 18-F MK-6240 demonstrates homogeneous uptake such that some structural features like the ventricles are visualized but without evidence of focal uptake in neocortex. Alzheimer's patients show a pattern of cortical uptake which is more intense, asymmetric, and focal (13), consistent with the distribution of the tau pathology reported in postmortem studies (9).

As the utility of tau PET radiotracers like 18-F MK-6240 expands, potential future clinical applications in AD may include: 1) aid in differential diagnosis of patients with cognitive impairment, 2) eligibility screening for long-term treatments to slow disease progression, 3) monitoring the effectiveness of such treatments, 4) assessing the course of disease, and 5) use as a prognostic biomarker potentially identifying at-risk cohorts. For any of these indications a robust method for visual assessment will be an important way to evaluate 18-F MK-6240 PET images,

especially in the clinical setting where an interpretation of the tau PET scan as negative or positive is a primary goal. However, unlike amyloid PET, tau uptake patterns demonstrate dynamic heterogeneity in individuals with AD, both in the spatial extent within the brain and the intensity of uptake within regions. This offers an opportunity for obtaining additional information from the visual read relevant to clinical research trials. Specifically, it may be possible to visually assess changes to the extent of uptake within regions over time, as well as between regions. Thus, it may be possible to elicit tau progression information from the visual interpretation.

The goals of the present study were to develop, field test, and refine a visual read method for 18-F MK-6240 PET as a potential tool for assessing *in vivo* brain tau accumulation, providing a readout of both tau positivity/negativity and the spatial extent of uptake within individual regions. The latter may be relevant to assessing within-patient changes in the context of clinical therapeutic drug trials where the tau PET signal might be expected to be unchanged or even decrease on serial imaging.

## MATERIALS AND METHODS

### Imaging data

Pooled imaging data provided under Informed consent from 204 participants with various diagnoses (Healthy controls (CN), Mild Cognitive Impairment (MCI) due to AD, AD dementia, and non-AD brain disorders) were obtained from Cerveau Technologies Inc. which gathered, curated, organized, and archived the data under contractual agreement from nine separate studies at nine clinical sites. All studies had Investigational Review Board approval and participants provided written informed consent for the procedures including handling of imaging data. Scans were received as reconstructed, scatter, randoms, and attenuation-corrected anonymized DICOM image volumes. Data from each center were acquired according to their own imaging protocol and therefore scan time windows were not guaranteed to overlap at later post injection times. Therefore, to generate an average static image for visual assessments, we used the most common overlapping late frame time window (6x5min frames), which was between 60 minutes and 90 minutes post-injection.

### Definition of MK-6240 uptake patterns

For developing the visual read method, we randomly selected a subset of the original database containing 102 participants (52 CN, 17 MCI, 29 AD, 4 non-AD neurological disease). 18-F MK-6240

averaged images were reviewed by an experienced nuclear medicine physician-researcher ("Reader 0") without a pre-specified examination protocol and blinded to each participants' diagnosis and imaging site. This process of unguided examination led to the identification and categorization of common uptake patterns that were generally consistent with current understanding of AD tau pathophysiology, and patterns that were better described as non-AD or off-target binding. Based on this review, an initial visual read procedure was developed for field testing and refinement using the second half of the image dataset (102 different, non-overlapping participants).

Visual assessment algorithm for 18-F MK-6240

The visual assessment of [18-F MK-6240 is a three-step process: 1) assess technical adequacy, 2) systematically review neocortical areas as for the presence and spatial extent of increased radiotracer uptake, and 3) apply a ruleset to step 2 findings for determination of positivity and classification as an AD pattern or not. Details are described in the Supplemental materials S-1.

Readers focus on eight pre-specified brain regions in each hemisphere of the cerebral cortex (16 regions total): hippocampus, mesial temporal, inferior temporal, lateral temporal, parietal, posterior cingulate, occipital, and frontal lobes (Figure 1). Temporal regions are grouped under the designation ""Cluster 1", while the extra-temporal cortical regions are grouped under the designation ""Cluster 2". Cluster 3 comprises subcortical regions (striatum/globus pallidus, thalamus, pons, dentate nucleus) suggestive of binding related to non-AD tauopathy.

For each of the 16 cortical regions the reader ascertains if there is abnormal increased radiotracer in the region relative to the cerebellum and informed by training examples. The reader also assigns a regional extent score expressed as a percent of the region showing abnormal increased uptake: none (0%), 1-25%, 26-75%, or >75% involvement of the region).

----- Figure 1----

Regional positivity is defined by either focal or confluent uptake involving at least 1-25% of the region (Table 1). Readers are asked to judge the extent (voxels with increased uptake within each region) rather than the intensity of uptake.

--Table 1—

The initial algorithm for assessing positivity was according to three rules below (Figure 2). As described below, this initial set of rules was refined further based on additional information and experience.

*Rule 1: Normal scan- no more than one region of focally increased radiotracer uptake in Clusters 1 or 2 (combined) and no regions of focal uptake in Cluster 3,*

*Rule 2: Positive, Alzheimer's pattern- two or more Cluster 1 and/or Cluster 2 positive regions, with at least one positive region in Cluster 1 and no positive regions in Cluster 3*

*Rule 3: Positive, non-Alzheimer's pattern- two or more Cluster 2 and/or Cluster 3 positive regions, with no positive regions in Cluster 1.*

----- Figure 2 -----

Readers

Two board-certified nuclear medicine physicians with research backgrounds in brain molecular imaging and some experience with flortaucipir PET, but naïve to 18-F MK-6240, served as testers of the read method. First, readers reviewed the scientific background, study rationale, and description of the read method with focus on brain region identification. Following a group case review and individual testing for competency, field testing with the two independent readers and one in-house expert reader was performed. A total of 112 test scans (102 cases + 10 repeat scans) were randomly presented without clinical or diagnostic information.

Scans were read in either linear gray or inverse gray scale. Readers were not provided MRIs, CTs or other structural information. Images were maintained on a DICOM server accessed by remote desktop software to run PMOD 3.8 (PMOD Technologies, Zurich) for visual display and adjustment of the PET scans. All interactions with the read platform were logged by PMOD. Readers recorded read findings in an electronic report form, which captured data, time, and user information

for each case. Reads were conducted by 2 readers over 2 days and 1 reader over 7 days, all initiated within one day of the training.

Evaluation of the read method

Binary readouts (positive/negative) for tau deposition by the two field test readers (Reader 1 (R1), Reader 2 (R2)) were compared to the ""gold standard" read by the internal nuclear medicine reader (Reader 0 (R0)). Concordance for visual assessment among all three readers were tallied for positive and negative cases and expressed as a percent. Cohen's kappa and Fleiss' kappa were used to assess reader by reader and group agreement correcting for chance agreement for the overall scan assessment. Rater agreement was also evaluated on a region-by-region basis for agreement between all three readers on both the binary determination and regional extent using Fleiss' kappa statistic (14). In addition, an exploratory overall tau visual regional extent score (VRES) was calculated for each region by assigning a value to the categorical region score of 1 for visual scores > 75%, 0.5 for visual scores of 26% to 75%, 0.25 for scores of < 25%, and 0 for visual scores of 0. Hence, a scan with complete bilateral uptake involving greater than 75% extension throughout each region has a total VRES of 16.

Other assessments included:

1. Intra-rater Test-retest reliability for scan positivity/negativity was determined for those scans (n=10) which were randomly presented twice to the readers.
2. Sensitivity, specificity, and area under the ROC curve were determined using site clinical diagnosis for those participants where this information was available (n=91).
3. Self-reported reader confidence in their assessments (Supplemental materials S2).

Comparison with Standard Uptake Value ratios

SUVr was used to compare the brain uptake in regions involved in Braak stages 1-6 and the VOI sampling of Jack, et al (15) (i.e., "Jack VOI") to binary and regional extent visual reads. T1-weighted MR images and 18-F MK-6240 PET scans were obtained from the Cerveau database. 18-F MK-6240 PET scans were processed and analyzed similar to visual read images, as described above. Details of the image processing are provided in Supplemental materials S3.

Statistical analyses

Comparison between demographics in the development and testing scan groups utilized descriptive statistics and unpaired t-tests. For visual reads, percent concordance against the read standard (Reader 0) was determined and pairwise and group reader agreement was assessed with Cohen's and Fleiss' kappa statistics, respectively. Readers' total VRES (0 -16) were compared with ANOVA and post hoc Welch's T. VRES totals were also correlated with SUVrs using Pearson's R. Further, individual visual reads, VRES analyses, and SUVrs using different VOI strategies underwent ROC curve for determination of sensitivity and specificity against the clinical site diagnosis as truth standard.

Refinement of the algorithm for assessing positivity

Following the blind read, a study close-out and image review with all readers assessed the adequacy of the initial ruleset for describing scans encountered in the presented cases. This information was supplemented by other data sources including published studies on the regional patterns of radiotracer uptake and review on another internal 18-F MK-6240 PET dataset of MCI and mild AD participants. Based on these learnings, the algorithm was refined. These refinements were then applied back to the original dataset for comparison to the preliminary algorithm.

**RESULTS**

Overall Binary Scan Assessment

The development (n=102) and testing (n=102) data sets were not significantly different with regards to age, self-reported gender, or diagnostic cohorts (Table 2). No scans were excluded from review by any reader for technical deficiencies.

----Table 2-----

Comparing overall binary tau positivity (original rules 2, 3) against the ""gold standard" by Reader 0, both Reader 1 and Reader 2 had a high level of concordance with complete agreement in 107 (95%) and 108 (95%) out of 112 scans (102 original + 10 repeated), respectively. All three readers were in complete agreement in 105 out of 112 cases (94%).

Cohen's k for pairwise comparisons were 0.964 and 0.955, respectively, for Reader 0 vs Reader 1 and Reader 0 vs Reader 2, indicating excellent agreement. Fleiss' kappa for inter-reader agreement among all three readers was 0.912, again indicating excellent agreement. Review of the 7 discordant cases show two primary causes; 1) technical issues like improper attenuation correction or reconstruction errors due to motion artifacts on scans judged to be still interpretable (n=5), and 2) uptake in the inferior and mesial temporal lobes that was incorrectly attributed to off-target meningeal uptake at the base of the skull (n=2) (Figure 3). Discrimination of this non-specific uptake from adjacent cortical regions may be improved by leveraging structural imaging (MR or CT) which was not allowed in the present study (Supplemental Figure 1).

----- Figure 3 -----

Regional assessments

Regional assessments were made in 1792 regions (112 cases x 16 regions). Evaluation of the sixteen regions for reader agreement on the binary determination of positivity/negativity for each region showed 3/3 rater agreement in 1329 regional reads (74.3% of regions), 2/3 agreement in 367 reads (20.4%), and 96 regions read (5.3%) with complete discordance. Over the sixteen regions, complete agreement between the three readers ranged from 69% to 86%. There was excellent agreement for some regions with Fleiss' kappa ranging from k= 0.726 (left mesial temporal) to k= 0.945 (right parietal) (Table 3).

Evaluation of reader agreement across the 4 possible regional responses (0, < 25%, 25-75%, >75%) showed substantial agreement for 8 of 16 regions and moderate agreement for the other 8 regions. The regions demonstrating highest reader agreement were the lateral temporal lobes and posterior cingulate cortex, while those regions demonstrating the least reader agreement were the occipital lobes, right hippocampus, and left mesial temporal lobe.

----- Table 3-----

Reproducibility

Reproducibility of the ten randomly selected cases for the binary determination of positive/negative was excellent, with Reader 0 and Reader 2 achieving 100% reproducibility and Reader 1 achieving 90%. The reproducibility of the binary positive/negative determination for the sixteen cortical regions for the 10 repeated case pairs was also very good, with readers showing 93.8%, 95.0%, and

98.8% (readers R0, R1, R2, respectively) self-agreement for the 10 case pairs of 16 regions. Overall, of the 480 region pairs (10 case pairs of 16 regions each for 3 readers) assessed with the regional extent score, 460 (95.8%) of the reads were scored identically. Regions which demonstrated the most intra-reader disagreement were left and right inferior temporal lobes which together had 8 intra-reader disagreements.

## Comparison with Clinical Diagnosis

In 91 of the 102 (89%) cases where a clear, well-supported clinical diagnosis of AD, MCI, or healthy volunteer was available from the site, we assessed binary visual reads, total VRES, and diagnostic sensitivity and specificity. Among the healthy volunteers 42 of 45 had negative visual reads, the AD participants demonstrated positive visual reads in 22 of 28 cases, while MCI positive visual reads were noted in 12 of 18 cases. The consensus read for the overall binary determination of scan positivity (2/3 or 3/3 readers agree) relative to a site diagnosis of cognitive impairment due to AD or AD dementia had a sensitivity = 81% and specificity = 93%. Comparing the visual read findings of each reader to the site clinical diagnosis as the ground truth demonstrated excellent specificity and moderate to good sensitivity for the visual read method (Table 4).

--- Table 4----

## Comparison with Standard Uptake Value ratios

SUVrs were calculated in 87 subjects who had a T1-weighted MRI for co-registration and regional segmentation. When parsed by visual read status, negative and positive read groups demonstrated significantly different mean SUVrs ($p<0.001$) for all individual Braak regions, Braak regional combinations (Braak 1-2, 3-4, 5-6), and Jack VOIs (Supplemental Table 1). Individual read data (Figure 4) demonstrate clustering of negative scans around SUVr=1 (no specific binding) while visually positive scans show higher SUVrs spread over a wider range. Although the majority of negative scans cluster around an SUVr of 1, we show a strong association ($r(102)=0.73$, $p< 0.0001$) between quantitative SUVrs from Braak 3-4 with semi-quantitative VRES summed across all readers (Supplemental figure 2). In addition, we examined the regional distribution of total VRES, showing a Braak-like staging, with a greater number or cases showing higher VRESs in the temporal regions compared to extra-temporal cortical regions (Supplemental figure 3).

---Figure 4---

ROC analyses were performed to assess different VOI template sampling for SUVr relative to the

clinical site diagnosis as gold standard. Comparing VOI sampling strategies to the site clinical diagnosis as the truth standard using the highest Youden Index for each analysis as the SUVr cut-off (Supplemental figure 4), again demonstrated excellent specificity and moderate sensitivity with Braak 3-4. Jack regions showed slightly higher sensitivity/specificity than the Braak 1-2 or Braak 5-6 analyses although this did not achieve statistical significance (p= 0.59).

Refinement of the algorithm for assessing positivity

The original algorithm for assessing positivity performed well in separating tau positive from tau negative scans consistently across readers. Nevertheless a few changes were suggested after reader review, review of additional 18-F MK-6240 PET datasets, and consideration of published literature (16). These resulted in the adjustments described in Table 5.

----- Table 5 -----

The revised algorithm was applied to the regional data with little effect on the determination of positivity by the readers. The final assessment was only changed in two scans, for one reader only. Both of which were reclassified from “Negative” to “Positive, atypical AD pattern”, improving the intrareader concordance from 90% to 100% for that reader. Overall concordance amongst the three readers showed 106/112 (95%) were in complete agreement on binary assessment of positivity.

## DISCUSSION

This study describes development, initial evaluation, and refinement of a visual read method for 18-F MK-6240 brain PET assessing the uptake patterns consistent with tau deposition in patients with Alzheimer's disease. The method provides two sets of related information, the binary determination of positive or negative for presence of pathological tau, and the regional extent of tau deposition in brain areas thought to be involved in the pathological progression of AD. The utility of this method lies in its potential use in clinical trials, particularly for eligibility assessments to confirm the presence of the targeted pathology, as well as to measure disease progression and treatment effect.

We demonstrated a high reader concordance for binary (positive/negative) and regional

assessment of brain uptake consistent with expected tau pathology. Discordance between readers was relatively low and due primarily to technical artifacts or the confusion of uptake at the base of the skull with inferior and mesial temporal lobes. As expected, absolute agreement for visual extent (4 possible categories) was lower but still substantial to moderate and varied by region. Agreement between readers was lower in the temporal lobe regions (excepting lateral temporal lobes) than other cortical regions (e.g., parietal, post. cingulate), perhaps due to more difficult anatomic localization relative to the more accessibly definable and larger regions and high meningeal uptake adjacent to inferomedial temporal regions.

Our data show robust test-retest reproducibility and good accuracy relative to clinical site diagnoses. The region with the lowest reproducibility was the inferior temporal lobes; signal in this region was infrequently misattributed to extracerebral uptake in surrounding regions. Off-target uptake of 18-F MK-6240 primarily involves meninges as well as uptake at the base of the calvarium, both of which had only minor impact on scan interpretation in this cohort and minimized with the aid of structural imaging. This compares favorably with 18-F AV-1451 where it is typically difficult to evaluate medial temporal structures due to the proximity of off-target uptake within the choroid plexus.

We observed a strong concordance between SUVr and the binary read with visually negative reads clustered around SUVr =1.0, consistent with negligible binding. By contrast, visually positive scans had higher SUVrs about twice the values of negative reads. Analysis of the semi-quantitative regional extent score (VRES) demonstrated more frequent occurrence of uptake in temporal cortical structures than extra-temporal regions, as predicted from the models of tau spread from the post-mortem data. However, these differences could also be due to the lack of normalization for region size, which ran from relatively small in the mesial temporal cortex to very large in frontal lobe. Partial volume error correction, which was not performed, would be expected to increase the differences between temporal and non-temporal regions.

Relative to clinical diagnosis, both the visual read and SUVR analysis showed excellent specificity and moderate to good sensitivity with the best combination of sensitivity/specificity (81% sensitivity/93% specificity) in the consensus read. Not surprisingly, the consensus read is the most commonly used PET interpretation method for eligibility assessments in clinical therapeutic trials, where amyloid PET negativity rates may be as high as 10-20% amongst individuals thought to have AD on clinical examination (17). In addition, lower sensitivity relative to clinical diagnosis is common

in cross-sectional datasets given the observed tendency for baseline clinical diagnosis to identify AD with higher sensitivity and lower specificity than imaging when the gold standard is final clinical diagnosis following longitudinal follow-up over a year or more (18). Both, the semiquantitative SUVR analysis and the visual read performed similarly for sensitivity and specificity, with visual reads having slightly higher Youden indices. While the SUVr cut-offs used here represent an optimized Youden index (see Supplemental figure 2), they may not be ideal for addressing other research questions, which may require a different point on the ROC curve.

Limitations

Although the readers were naïve to 18-F MK-6240 PET, they were brain imaging specialists with experience reading other tau and amyloid PET tracers. Reader selection was intentional in order to evaluate a read method designed for research and clinical trial visual assessment of brain tau burden with regional information for tracking change and comparing patients rather than the simple binary determination of positive/negative determination normally required in routine clinical use. Hence, the study readers may have more readily handled the difficult task of accurately identifying subregions within the temporal lobes and other neocortical regions than readers with less experience. Future studies will aim to evaluate the method and training paradigm with less experienced readers to determine generalizability of the binary and regional extent aspects of the method. Perhaps more germane to the routine clinical use is that a simpler version of this method can be employed for the easier task of determining overall scan positivity.

Limitations of this study are a lack of longitudinal data, no pathological diagnostic confirmation, and limited data on amyloid status or clinical measures to support the accuracy of clinical diagnosis. Moreover, our development sample included a limited number or MCI patients (n=17), which may limit the applicability of the algorithm for clinical diagnosis and clinical trial inclusion. Specifically, in cases with lower 18-F MK-6240 binding, such as in MCI, off-target binding in meningeal tissue could theoretically lower accuracy of visual reads in proximal brain regions. Future research will aim to include a better characterized sample, more representative of patients screened for initial memory complaints. Furthermore, there are exceptions to the rules, specifically logopenic primary progressive aphasia will score as typical AD owing to unilateral left hemispheric uptake in cluster regions 1 and 2. The case used as an example of positive scan, non-AD pattern in figure 2 raises additional points. Here there is

uptake in the caudate bilaterally and also in the right occipital and parietal regions. Is this an atypical AD or non-AD tauopathy? In these rare cases additional data may prove to clarify and adjust the algorithm. Finally, we included midbrain in Cluster 3 while instructing readers not to rate the substantia nigra as positive even though it is localized in midbrain. This may be confusing for some, although the readers very rarely endorsed midbrain regardless of the evident uptake in the substantia nigra.

The selection of the time window of 60 to 90 minutes, was driven by the availability of the data and may not be ideal for separating negative from positive scans. In high binding regions of positive scans, SUVrs may continue to rise past 90 minutes as washout proceeds (13). When visually comparing early versus later PET images, there is some minor residual background uptake in some of the scans. However, this is most apparent in the positive scans which are typically easier to interpret than negative scans because of the relatively low off- target uptake. In summary, we believe that given the small impact on scan interpretability, it was preferable to open the time window rather than reducing the size of the cohort.

## Future work

Further validation of this read method would be needed for greater confidence in the utility of the visual assessment toward the intended research application. As discussed above, the present data has resulted in a minor adjustment of the algorithm to be more reflective of the potential range of cases that a reader may encounter. While we are confident that the proposed algorithm captures most AD-related cases, minor adjustments may be needed to better marry the scan phenomenology with the intended cohort. Future development should also focus on evaluation of patients with serial 18-F MK-6240 scans with months to years of longitudinal data, interpretation of scans over a wider range of tauopathies, evaluation in a larger group of readers with greater range of experience reading PET images, and 18-F MK-6240 PET with clinical metrics and post-mortem brain pathology.

## Conclusions

This work provides a method for the visual interpretation of 18-F MK-6240 brain PET images. This initial cross-sectional study demonstrated that experienced readers can utilize this

algorithm for robust and reproducible visual interpretation of 18-F MK-6240 brain scans. The overall determination of scan positivity may be useful in aiding diagnosis or enhancing the accuracy of clinical trial enrollment. Scan information derived from cortical regional assessment may be most valuable for within-patient evaluation of change over time, as well as for determining the efficacy of new treatments designed to alter the course of progression.

ACKNOWLEDGEMENTS

The authors acknowledge the very helpful scientific input and logistical support provided by Ms. Megan Stark. The authors also acknowledge the following investigators and their funding organizations for contributing subjects' data used in this study: Davangere P. Devanand, M.D., Department of Psychiatry, Columbia University Irving Medical Center, New York. Grant support from R01 AG055422; Eric D. Hostetler, PhD, Merck & Co., Inc.; Keith Johnson, MD, Department of Radiology and Neurology, Massachusetts General Hospital, Harvard Medical School; Sterling C. Johnson, PhD, Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health. Grant support from AG021155, AG062285, AG027161, AG062715, AG062167, and S10OD025245; William C.Kreisl, MD, Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University. Grant support from R01AG063888 and K23AG052633; Pedro Rosa- Neto, MD, PhD, Translational Neuroimaging Laboratory, McGill University Research Centre for Studies in Aging, Alzheimer's Disease Research Unit, Douglas Research Institute, Le Centre intégré universitaire de santé et de services sociaux (CIUSSS) de l'Ouest-de-l'Île-de- Montréal; Department of Neurology and Neurosurgery, Psychiatry and Pharmacology and Therapeutics, McGill University; Christopher C. Rowe, MD, Department of Molecular Imaging and Therapy Austin Health, Australian Dementia Network, University of Melbourne, Australian Imaging, Biomarker and Lifestyle study of aging (AIBL), Florey Institute of Neuroscience and Mental Health. Grant support from NHMRC APP1132604, APP1140853, and APP1152623; Yaakov Stern, PhD, Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University: Grant support from 5R01 AG038465. This study was funded by Biogen.

KEY POINTS

QUESTION: Can a visual read method for brain tau deposition using 18-F MK-6240 PET provide robust overall positivity/negativity and information about the regional extent of abnormal tracer uptake?

PERTINENT FINDINGS: Readers can utilize this algorithm for robust and reproducible visual interpretation of 18-F MK-6240 brain scans at the whole brain and regional levels.

IMPLICATIONS FOR PATIENT CARE: This read method may be a useful tool for AD clinical drug development as means to enhancing the accuracy of clinical trial enrollment, for within-patient evaluation of change over time, and for determining the efficacy of new treatments designed to alter the course of progression.

FIGURES & TABLES

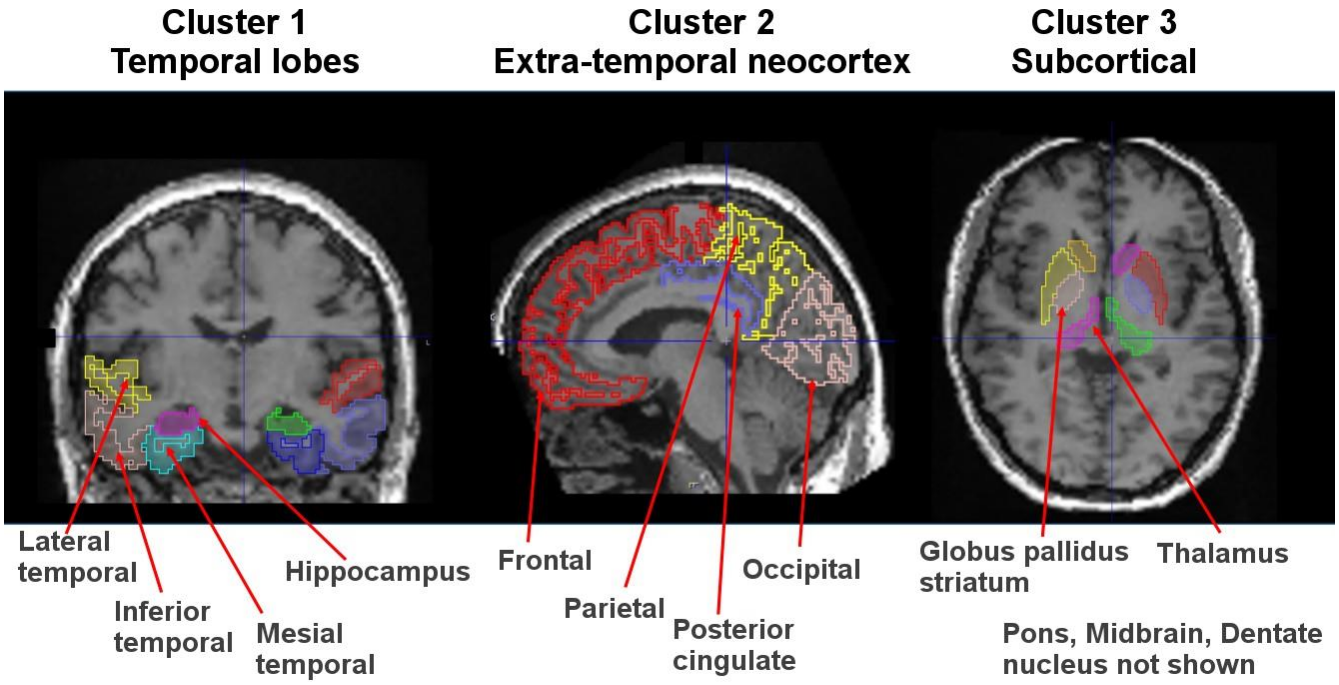



Figure 1 Regions for visual read are outlined and overlaid on T1 MRI for anatomic reference.

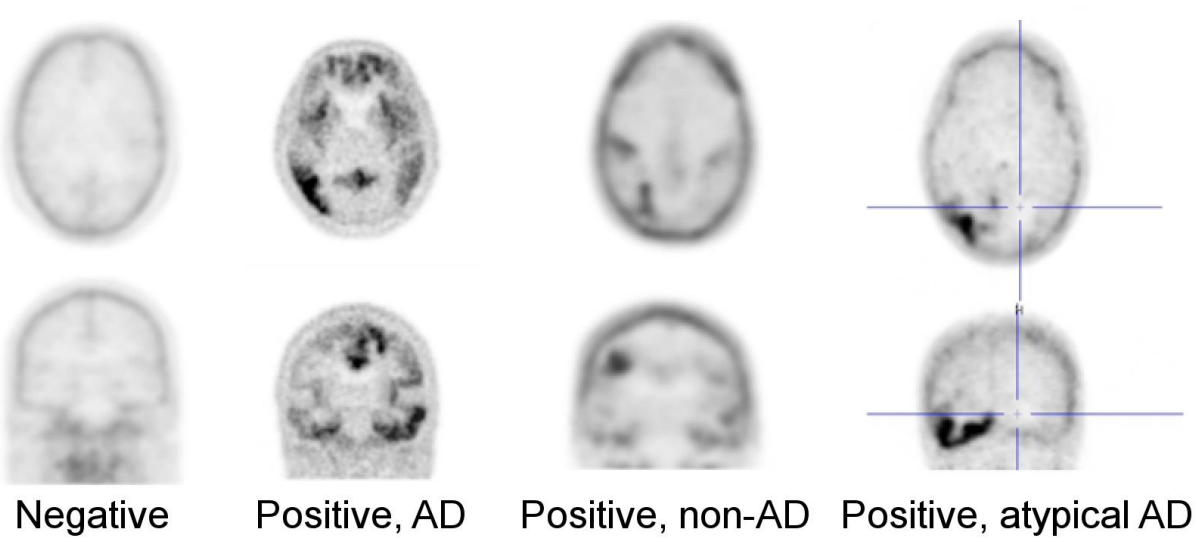



Figure 2 18-F MK-6240 PET in Healthy Volunteers, Alzheimer and non-Alzheimer Tauopathy. Note that “Positive, atypical AD” was added in the refined algorithm. Also, the non-AD tauopathy is an extremely rare instance of cluster 3 and cluster1,2 positivity
classified on basis of the striatal finding. See Discussion section.

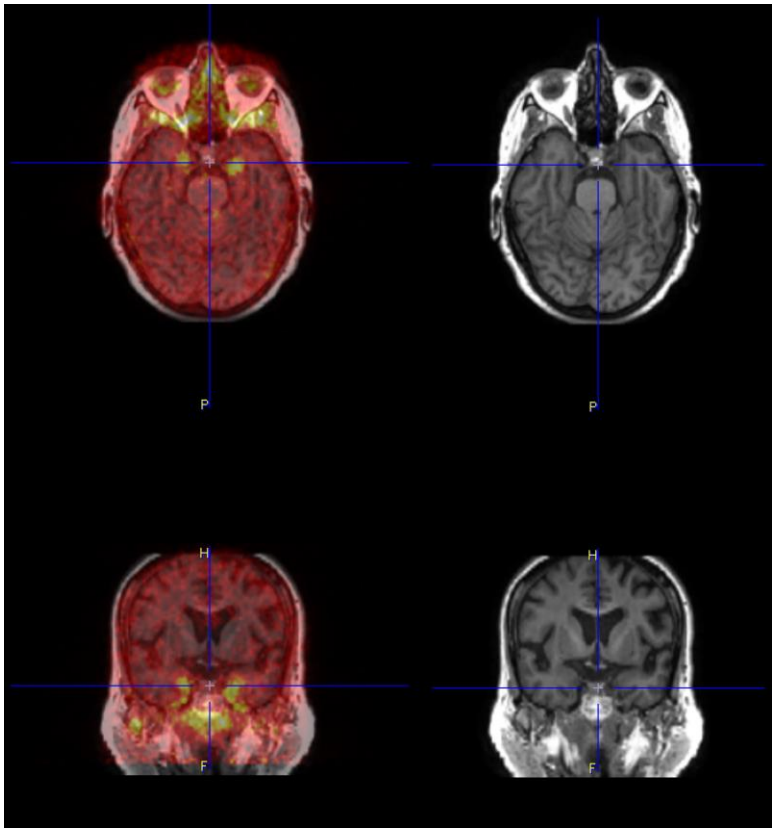



Figure 3. Difficult case shows bilateral anterior mesial temporal uptake which can be confused with off target uptake in the meninges and the
floor of the calvarium.

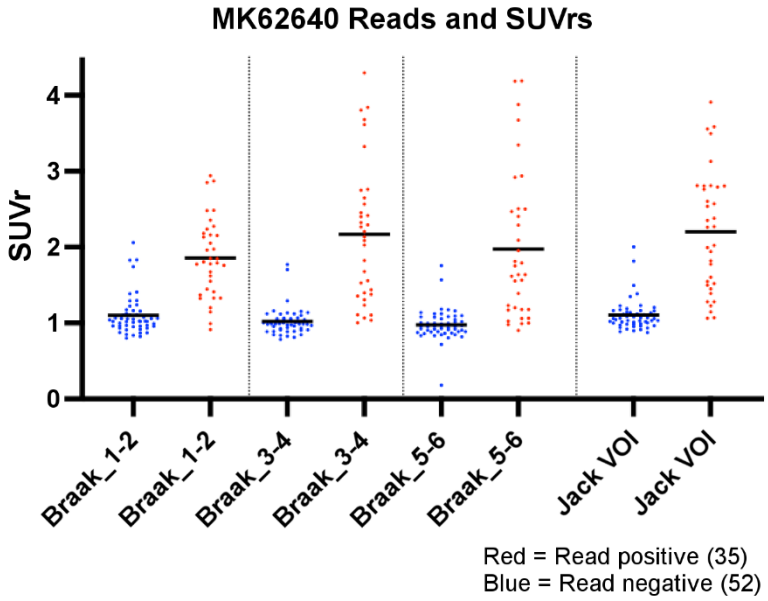



Figure 4 Standard uptake value ratios for different regional VOIs parsed by visual read negative or positive.

## Tables

*Table 1 Regions for Visual Assessment*

| Brain area | Included Regions | Visual Rating | No. regions | Rationale |
|---|---|---|---|---|
| Cluster 1 Temporal lobes | Hippocampus Mesial temporal Inferior temporal Lateral temporal | No uptake (0%) Uptake 1-25% extension 26% to 75% >75% extension | 8 regions 4 each left and right hemisphere | Earliest cortical regions involved in AD per Braak staging |
| Cluster 2 Extra-temporal Neocortex | Occipital<br>Post.cingulate<br>Parietal<br>Frontal | No uptake (0%) Uptake 1-25% extension 26% to 75% >75% extension | 8 regions 4 each left and right hemisphere | Next regions involved in AD |
| Cluster 3 Subcortical Area | Striatum-globus Thalamus Dentate nucleus Pons Midbrain | Presence/absence | 5 regions | May be positive in non-AD tauopathies |

Table 2 Demographic Data

| DEVELOPMENT DATASET | | | | | TEST DATASET | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ***GROUP*** | ***N*** | AGE Yrs mean(SD) | GENDER male | GENDER female | ***GROUP*** | ***N*** | AGE Yrs mean(SD) | GENDER male | GENDER female |
| AD | 29 | **72.4** 9.9 | 20 | 9 | AD | 24 | **70.4** 10.7 | 12 | 12 |
| MCI | 17 | **71.2** 7.1 | 11 | 6 | MCI | 21 | **69.9** 8.0 | 12 | 9 |
| HEALTHY | 52 | **66.4** 12.1 | 25 | 27 | HEALTHY | 45 | **68.6** 7.5 | 18 | 27 |
| OTHER | 4 | **63.3** 4.6 | 0 | 4 | OTHER | 12 | **65.9** 9.4 | 5 | 7 |

Table 3. Reader Agreement for Visual Assessment Positive/Negative for Regional Tau

| Regional Binary Agreement Positive/Negative | | | Complete Agreement on Spatial Extent Scoring | | |
|---|---|---|---|---|---|
| Region | Fleiss' kappa | Interpretation | Region | Fleiss' kappa | Interpretation |
| Right Parietal | 0.945 | Almost perfect agreement | Right Lateral Temporal | 0.748 | Substantial agreement |
| Right Frontal | 0.929 | | Left Lateral Temporal | 0.716 | |
| Right Lateral Temporal | 0.907 | | Right Post. Cingulate | 0.702 | |
| Left Parietal | 0.906 | | Left Posterior Cingulate | 0.678 | |
| Left Posterior Cingulate | 0.904 | | Right Parietal | 0.655 | |
| Left Lateral Temporal | 0.888 | | Left Parietal | 0.621 | |
| Right Post. Cingulate | 0.87 | | Right Frontal | 0.618 | |
| Left Frontal | 0.801 | Substantial agreement | Left Frontal | 0.602 | |
| Left Occipital | 0.801 | | Right Inferior Temporal | 0.597 | Moderate agreement |
| Left Hippocampus | 0.783 | | Left Inferior Temporal | 0.576 | |
| Right Inferior Temporal | 0.782 | | Right Hippocampus | 0.552 | |
| Right Occipital | 0.759 | | Right Mesial Temporal | 0.552 | |
| Right Hippocampus | 0.756 | | Left Hippocampus | 0.551 | |
| Left Inferior Temporal | 0.753 | | Left Mesial Temporal | 0.545 | |
| Right Mesial Temporal | 0.742 | | Right Occipital | 0.537 | |
| Left Mesial Temporal | 0.726 | | Left Occipital | 0.503 | |

Overall binary assessment kappa = 0.912

TABLE 4. Sensitivity/Specificity of MK-6240 Visual Reads and SUVR analyses using clinical diagnosis as standard-of-truth

| Method | Reader | Sensitivity | Specificity | ′ouden Index |
|---|---|---|---|---|
| Visual Read | R0 | 0.79 | 0.93 | 0.72 |
| | R1 | 0.75 | 0.93 | 0.68 |
| | R2 | 0.79 | 0.91 | 0.70 |
| | Consensus | 0.81 | 0.93 | 0.74 |
| **Visual** | Cut-off | | | |
| **VRES** | 1.5 | 0.742 | 0.890 | 0.63 |
| **SUVr** | Cut-off | | | |
| **B1-2** | 1.4 | 0.649 | 0.888 | 0.54 |
| **B3-4** | 1.3 | 0.645 | 0.963 | 0.61 |
| **B5-6** | 1.2 | 0.611 | 0.960 | 0.57 |
| **Jack** | 1.4 | 0.650 | 0.880 | 0.53 |

Table 5. Initial and revised algorithms for interpreting MK-6240 PET

| Scan Assessment | Original Algorithm | Refined Algorithm | Reason for adjustment |
|---|---|---|---|
| Negative | No more than one region positive in the cortex | All clusters negative | *Allowing one region to be positive was to prevent calling scans positive when meningeal uptake close to the inferolateral temp lobes could be misread as a positive region. This was dropped with improved methods/instructions to identify this confound.* |
| Positive, AD pattern | Evidence of increased radiotracer uptake in two cortical regions with at least one region in temporal lobes | At least one cluster 1 region positive and no cluster 3 positive regions | *Revised following observations of multiple cases with isolated cluster 1 abnormality in just one region* |
| Positive, Atypical AD pattern | Not assessed | Evidence of increased radiotracer uptake in one or more regions in cluster 2 but not clusters 1 or 3 | *This rare pattern is noted from review of other MK- 6240 datasets and is an expected pattern based on the published literature off other tau PET tracers* |
| Positive, Non-AD pattern | Any positive scan not fitting AD criteria | Evidence of increased radiotracer uptake in one or more regions in clusters 1-3 with at least one region in cluster 3 | *Formalized uniform assessment for regions involved in non-AD tauopathies* |

# REFERENCES


1. Braak H, Del Tredici K. Top-Down Projections Direct the Gradual Progression of Alzheimer-Related Tau Pathology Throughout the Neocortex. Adv Exp Med Biol. 2019;1184:291-303.
2. Dronse J, Fliessbach K, Bischof GN, von Reutern B, Faber J, Hammes J, et al. In vivo Patterns of Tau Pathology, Amyloid-beta Burden, and Neuronal Dysfunction in Clinical Variants of Alzheimer's Disease. J Alzheimers Dis. 2017;55(2):465-71.
3. Knopman DS, Lundt ES, Therneau TM, Vemuri P, Lowe VJ, Kantarci K, et al. Joint associations of beta-amyloidosis and cortical thickness with cognition. Neurobiol Aging. 2018;65:121-31.
4. Bao W, Jia H, Finnema S, Cai Z, Carson RE, Huang YH. PET Imaging for Early Detection of Alzheimer's Disease: From Pathologic to Physiologic Biomarkers. PET Clin. 2017;12(3):329-50.
5. Bejanin A, Schonhaut DR, La Joie R, Kramer JH, Baker SL, Sosa N, et al. Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. Brain. 2017;140(12):3286-300.
6. Buckley RF, Hanseeuw B, Schultz AP, Vannini P, Aghjayan SL, Properzi MJ, et al. Region-Specific Association of Subjective Cognitive Decline With Tauopathy Independent of Global beta-Amyloid Burden. JAMA Neurol. 2017;74(12):1455-63.
7. Hansson O, Grothe MJ, Strandberg TO, Ohlsson T, Hagerstrom D, Jogi J, et al. Tau Pathology Distribution in Alzheimer's disease Corresponds Differentially to Cognition-Relevant Functional Brain Networks. Front Neurosci. 2017;11:167.
8. Schwarz AJ, Yu P, Miller BB, Shcherbinin S, Dickson J, Navitsky M, et al. Regional profiles of the candidate tau PET ligand 18F-AV-1451 recapitulate key features of Braak histopathological stages. Brain. 2016;139(Pt 5):1539-50.
9. Braak H, Del Tredici K. Potential Pathways of Abnormal Tau and alpha-Synuclein Dissemination in Sporadic Alzheimer's and Parkinson's Diseases. Cold Spring Harb Perspect Biol. 2016;8(11).
10. Meyer PF, Pichet Binette A, Gonneaud J, Breitner JCS, Villeneuve S. Characterization of Alzheimer Disease Biomarker Discrepancies Using Cerebrospinal Fluid Phosphorylated Tau and AV1451 Positron Emission Tomography. JAMA Neurol. 2020;77(4):508-16.
11. Hostetler ED, Walji AM, Zeng Z, Miller P, Bennacef I, Salinas C, et al. Preclinical Characterization of 18F-MK-6240, a Promising PET Tracer for In Vivo Quantification of Human Neurofibrillary Tangles. J Nucl Med. 2016;57(10):1599-606.
12. Koole M, Lohith TG, Valentine JL, Bennacef I, Declercq R, Reynders T, et al. Preclinical Safety Evaluation and Human Dosimetry of [(18)F]MK-6240, a Novel PET Tracer for Imaging Neurofibrillary Tangles. Mol Imaging Biol. 2020;22(1):173-80.
13. Betthauser TJ, Cody KA, Zammit MD, Murali D, Converse AK, Barnhart TE, et al. In Vivo Characterization and Quantification of Neurofibrillary Tau PET Radioligand (18)F-MK-6240 in Humans from Alzheimer Disease Dementia to Young Controls. J Nucl Med. 2019;60(1):93-9.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
15. Jack CR, Wiste HJ, Botha H, Weigand SD, Therneau TM, Knopman DS, et al. The bivariate distribution of amyloid-beta and tau: relationship with established neurocognitive clinical syndromes. Brain. 2019;142(10):3230-42.
16. Phillips JS, Nitchie FJt, Da Re F, Olm CA, Cook PA, McMillan CT, et al. Rates of longitudinal change in (18) F-flortaucipir PET vary by brain region, cognitive impairment, and age in atypical Alzheimer's disease. Alzheimers Dement. 2022;18(6):1235-47.
17. Chapleau M, Iaccarino L, Soleimani-Meigooni D, Rabinovici GD. The Role of Amyloid PET in Imaging Neurodegenerative Disorders: A Review. J Nucl Med. 2022;63(Suppl 1):13S-9S.
18. Veitch DP, Weiner MW, Aisen PS, Beckett LA, DeCarli C, Green RC, et al. Using the Alzheimer's Disease Neuroimaging Initiative to improve early detection, diagnosis, and treatment of Alzheimer's disease. Alzheimers Dement. 2022;18(4):824-57.

SUPPLEMENTAL MATERIALS

## S-1. Detailed visual assessment algorithm for [$^{18}$F]MK-6240

The visual assessment of [18F]MK-6240 is a three-step process. First, readers adjusted the image to standardize the orientation, windowing, and thresholding of the gray scale per explicit instructions for consistency when viewing images in all three planes (axial, coronal, and sagittal). Readers were not permitted to sample voxel values, draw regions of interest, or develop a quantitative measure. Next, technical adequacy for visual review was systematically queried to determine the overall quality of the scan for the following: positioning, motion, attenuation correction, field of view, image filtering, and count rate. Next the image volumes are reoriented in a standardized manner and the upper threshold on the window is set to 70-80% of maximum, depending upon the characteristics of the scan. An inverse linear grayscale is the recommended scale to use for interpretation based on previous data which suggests biases in applying many color scales including monochrome color scales.

Once the images are processed as described the next step is careful interrogation of neocortical regions for the presence or absence of focal uptake suggestive of tau deposition in eight pre-specified brain regions in each hemisphere of the cerebral cortex (16 regions total): hippocampus, mesial temporal, inferior temporal, lateral temporal, parietal, posterior cingulate, occipital, and frontal lobes. These regions are organized into clusters for temporal regions (Cluster 1), extra-temporal cortical regions are grouped under the designate (Cluster 2), and an additional Cluster 3 comprising subcortical regions suggestive of binding related to non-AD tauopathy.

For each of these 16 cortical regions the reader ascertains first, if there is abnormal increased radiotracer in the region relative to the cerebellum and as informed by a basic sense of abnormal levels of tissue uptake provided by the training. Uptake in the cerebellar gray is an internal reference for nonspecific distribution of the tracer; as a rule of thumb uptake greater than 1.5 times the cerebellar gray would be in keeping with increased [$^{18}$F]MK-6240 binding. Next, the reader determines for each region, what percentage of the region appears to show, if any, abnormal increased uptake in one of four categories: none (0%), 1-25%, 26-

75%, or >75% involvement of the region.

Region by region assessment begins on the coronal views where the left and right temporal lobes are easily displayed. There are four regions of the temporal area read in sequence: hippocampus, mesial temporal lobe, inferior temporal lobe, lateral temporal lobe (Cluster 1). In assessing these regions readers identify focal uptake which may be asymmetric or limited to medial and inferior structures or medial structures alone. The sagittal image allows visualization of the anatomic distribution of abnormal uptake in the lateral temporal regions where there are distinct linear areas of increased uptake that track the lobar anatomy. The axial images are next viewed for initial assessment of the parietal, occipital, posterior cingulate and frontal regions (Cluster 2) for evidence of increased tracer uptake relative to the cerebellar gray. Regional positivity is defined by either focal or confluent uptake involving at least 1-25% of the region (Table 1). Readers are asked to judge the extent of abnormal increased uptake within the region rather than intensity of that uptake. The extension refers to the number of voxels with increased uptake, while the intensity is how high the uptake is in any given voxel or cluster of voxels with respect to cerebellar gray matter. Next, the reader reviews subcortical brain uptake (striatum/globus pallidus, thalamus, pons, midbrain, dentate nucleus; collectively “Cluster 3”). Finally, the reader assesses common areas of “off-target” uptake which might confound regional evaluation, including: meninges, substantia nigra, venous sinuses and other vascular uptake including scalp, bony structures at the skull base, eyeballs, and salivary glands. The last step is the application of the algorithm to the regional positivity findings to address the questions is this scan positive or negative for evidence of brain tau? If positive is it in an Alzheimer's pattern? If it is an Alzheimer's pattern is it typical or atypical?

## S-2. Reader Confidence

Self-rated confidence was measured with a five-point Likert scale where 1 is no confidence and 5 is complete confidence. Of the total of 336 scans read, 295 scans (87.8%) were read with complete confidence while 41 scans (12.2%) were read with less than complete confidence; of those, 21 scans (6.2%) were read with moderate or less (score < 4) confidence. Of those scans scoring 4 or less, 17 (41.5%) were read as negative, 19 (46.3%) as positive, AD pattern, and 5 (12.1%) as positive, non-AD pattern. While differences between readers were small given the high number of completely confident reads, the number of reads rated <5 was less consistent, with readers reporting between 3 and 25 cases as less than completely confident. However, reader confidence had no bearing on the reader's concordance with other readers.

## S-3. Details of image processing for SUVR determination

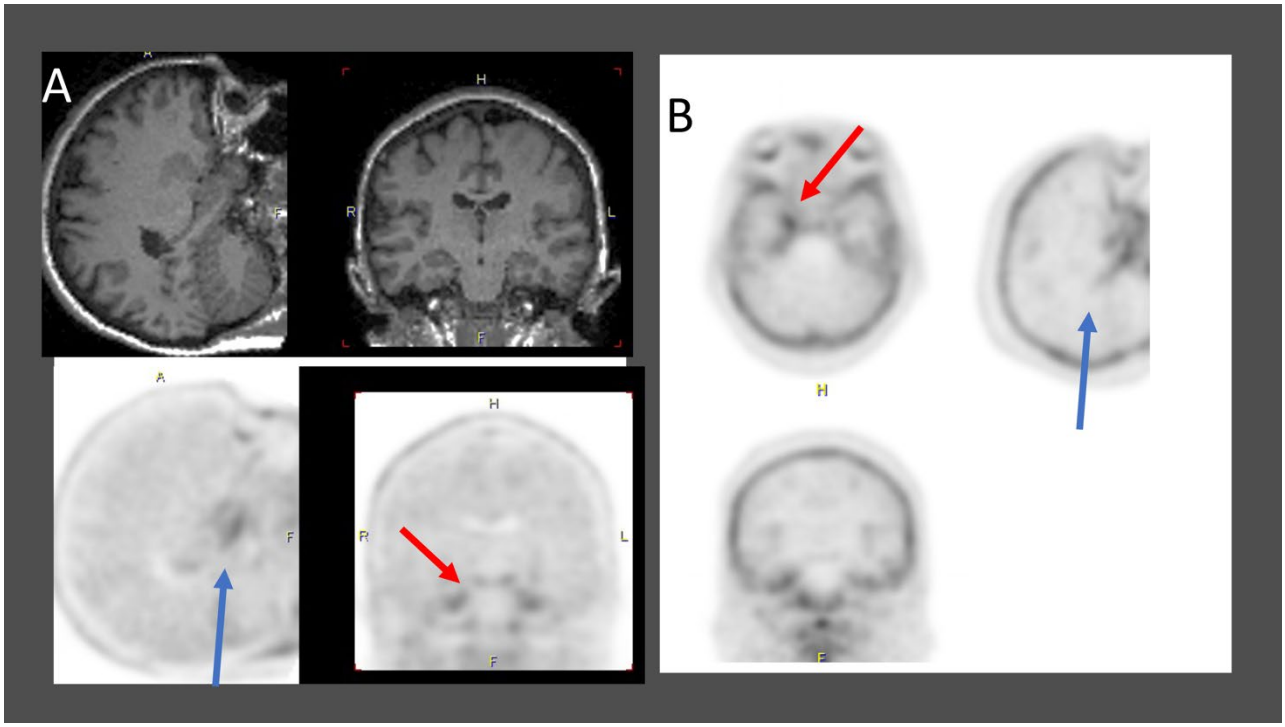
[$^{18}$F]MK6240 images were nonlinearly registered into MNI152 space using the subjects T1 MRI scan as part of a diffeomorphic nonlinear registration (DARTEL). First, the T1 structural MRI images were segmented into gray matter and white matter using SPM12. DARTEL then uses these tissue probability maps to create flow-fields which provide the parameters required to spatially normalize any images which are co-registered to the MRI image into MNI152 space. Each PET image was registered to the corresponding MRI using a rigid-body registration. Finally, the individuals' DARTEL flow-field was applied without modulation resulting in [$^{18}$F]MK6240 images in MNI152 space. The normalized maps were spatially smoothed with an 8mm full width at half maximum Gaussian kernel. SUVr images for [$^{18}$F]MK6240 were generated by dividing all intensities in the image by the mean uptake value of the reference region, for which we utilized the ventrolateral cerebellum VOI from the CIC atlas.

Supplemental Table 1

Standard uptake value ratios for different brain regions and volume of interest sampling strategies parsed by visual read status.

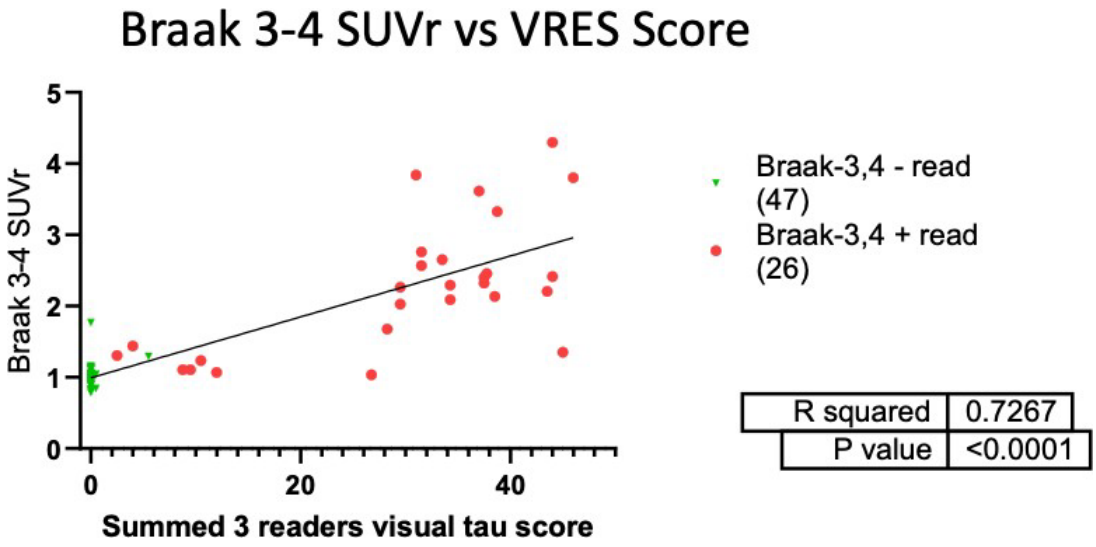
| TABLE 1 | Negative Read | | Positive Read | | Welch's t |
|---|---|---|---|---|---|
| SUVr | Mean | SD | Mean | SD | P |
| Braak_1 | 1.24 | 0.28 | 2.02 | 0.57 | <0.0001 |
| Braak_2 | 0.89 | 0.29 | 1.62 | 0.49 | <0.0001 |
| Braak_3 | 1.06 | 0.17 | 2.04 | 0.88 | <0.0001 |
| Braak_4 | 1.01 | 0.18 | 2.20 | 0.94 | <0.0001 |
| Braak_5 | 0.99 | 0.22 | 2.05 | 1.03 | <0.0001 |
| Braak_6 | 0.92 | 0.17 | 1.63 | 0.77 | <0.0001 |
| Braak_1-2 | 1.10 | 0.26 | 1.86 | 0.52 | <0.0001 |
| Braak_3-4 | 1.02 | 0.18 | 2.17 | 0.91 | <0.0001 |
| Braak_5-6 | 0.98 | 0.21 | 1.97 | 0.97 | <0.0001 |
| Jack VOI | 1.10 | 0.21 | 2.20 | 0.80 | <0.0001 |

Supplemental Figure 1



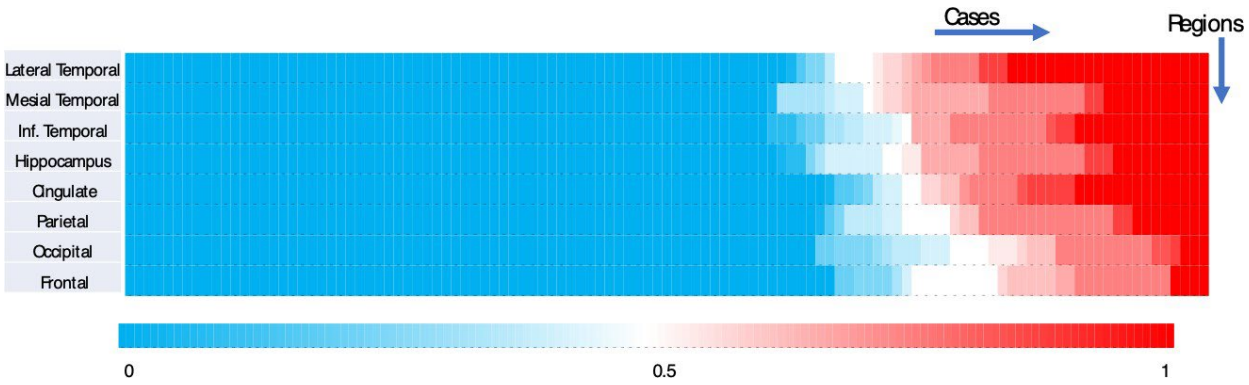

Supplemental Figure 1 Two difficult cases (A and B) with discordant reads. Case A lower gray scale 18-F MK-6240 shows small foci of uptake bilaterally in the mesial temporal lobes on the coronal view (red arrow) which has the appearance of meningeal uptake, but on sagittal view is linear and conforms to temporal lobe. This is confirmed on MRI. B is a patient with asymmetric uptake in the mesial aspect of the right temporal lobe (red arrow). The linear appearance on the sagittal images (blue arrow) is consistent with temporal binding.

Supplemental Figure 2



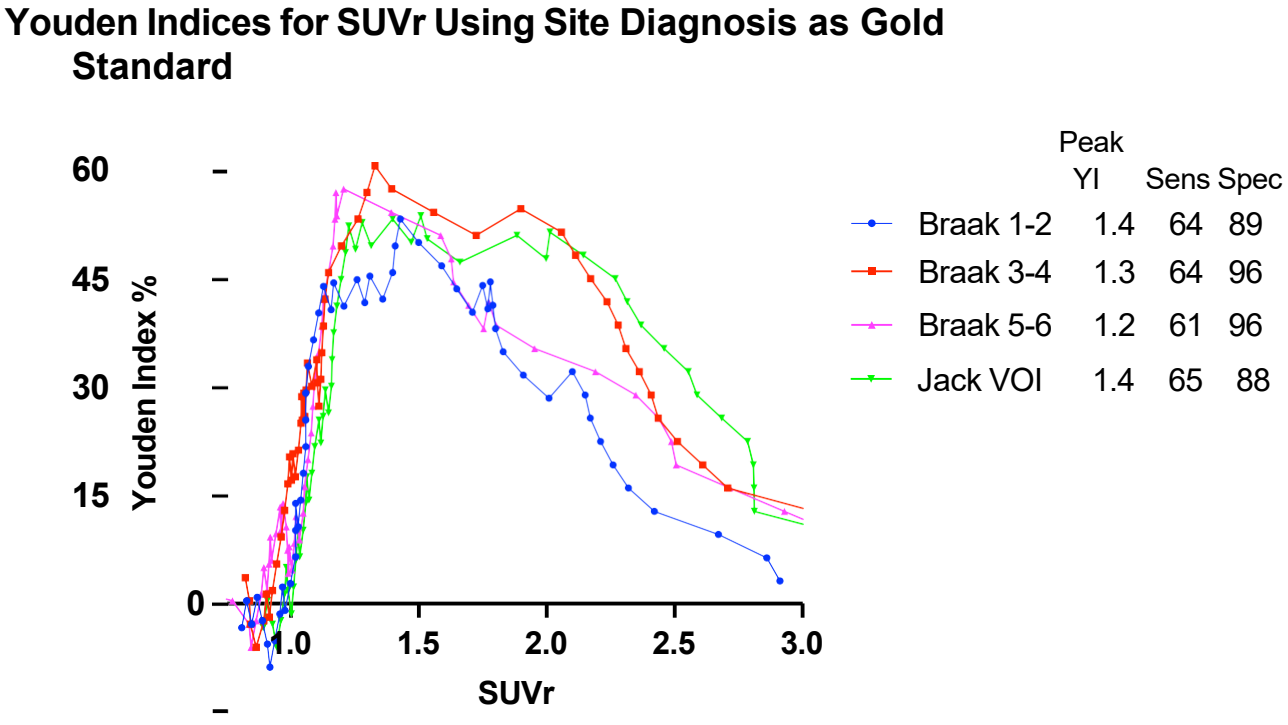

Supplemental Figure 2. Braak 3-4 SUVrs plotted against the three reader summed visual reader extent score (VRES) and fit to a linear regression. Note the large majority of visually negative scans had visual tau scores = 0 and SUVrs clustering around 1.

Supplemental Figure 3




Supplemental Figure 3. Composite mean cortical tau regional visual uptake score calculated for all 102 research participants by merging left and right hemisphere homotypic regions and averaging all three readers scores. Scores for each subject and region range from 0 (no uptake) to 1 (>75% of region involved).

Supplemental Figure 4




Supplemental Figure 4 Sensitivity/specificity analysis for different SUVr sampling templates using site clinical diagnosis as the truth standard. Youden index is obtained by ROC curve and calculated as the sum of (sensitivity and specificity)-1, expressed as a percent for different SUVr strategies and cut-offs.