HHS Public Access

Author manuscript

IEEE Trans Med Imaging. Author manuscript; available in PMC 2020 March 01.

Published in final edited form as:

IEEE Trans Med Imaging. 2019 March ; 38(3): 697–709. doi:10.1109/TMI.2018.2870441.

Scan-time Corrections for 80-100 Minute Standardized Uptake Volume Ratios to Measure the ^{18}F-AV-1451 Tracer for Tau Imaging

Mark He,
Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC (markhe@live.unc.edu).

Suzanne L. Baker,
Lawrence Berkeley National Laboratory.

Vyoma D. Shah,
Helen Wills Neuroscience Institute at University of California, Berkeley.

Lawrence Berkeley National Laboratory.

Samuel N. Lockhart,
Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC.

William J. Jagust
Helen Wills Neuroscience Institute at University of California, Berkeley.

Lawrence Berkeley National Laboratory.

Abstract

The ^{18}F-AV-1451 PET tracer binds to tau, an Alzheimer's Disease (AD) biomarker. The standardized uptake value ratio (SUVR) 80–100 min window is widely used to quantify tau binding, although ^{18}F-AV-1451 continues increasing relative to a reference region in regions with tau deposition. Left uncorrected, acquisition time inaccuracies can lead to errors from −4% to 6% in 20-min SUVR measurements in subjects with Alzheimer's Disease. In 40 subjects with scans from 75–115 min following ^{18}F-AV-1451 injection, we created 20-min reconstructions (4×5 min) of start-times ranging from 7585 min, as proxies of offset scans and calculated the mean in regions of interest (ROIs). We developed a Segmented Least Squares (SLS) method to obtain error-minimizing weighting coefficients for ^{18}F-AV-1451 ROIs that best predict SUVR 80–100 from weighted means of SUVRs from offset start-times. We compared residual errors of our SLS method to those in (1) uncorrected offset 20-min-SUVRs, (2) the mean of 5-min frames within the 80–100 window, and (3) a least-squares interpolation method. We evaluated errors induced by start-time offset on SUVRs for each method. SLS, which corrected using least-squares coefficients of 5-min components, consistently reduced errors across all offset start-times. Effect size analysis for simulated clinical longitudinal ^{18}F-AV-1451 drug trials showed that uncorrected 20-min offset SUVRs would require up to 20% more participants to detect treatment effects compared to using SLS. Correction of SUVR scantime errors by SLS minimizes errors compared to other correction methods and may be extended to other scanners and tracers.

Index Terms—

Tau; PET Imaging; SUVR; Scan-time error; Alzheimer's Disease

I. INTRODUCTION

A. SUVR 80–100for Tau PET Scans

Tau and beta-amyloid proteins are key biomarkers of Alzheimer's Disease (AD). Positron emission tomography (PET) tracers have enabled *in vivo* measurement of amyloid for over a decade. New tracers allow us to measure tau *in vivo* as well, facilitating the study of tau accumulation in aging and dementia. Standardized uptake value ratio (SUVR) from 80–100 min has become the standard method for quantification of the tau tracer ^{18}F-AV-1451 [3,13,14,17,19,20], also known as *flortaucipir* [23], due to the need for a short scan time and the high correlation with Simplified Reference Tissue Models and Logan Graphical Analysis, which require longer scan times. Long scans are uncomfortable or unfeasible for aging and dementia populations, and not realistic for larger scale studies.

Recent work demonstrated that ^{18}F-AV-1451 SUVR 80–100 correlated with region-of-interest (ROI) non-displaceable Binding Potential and Distribution Volume Ratio imputed from reference tissue models with R2 between 0.95 and 0.99 [3]; however, SUVR values continued to increase in areas of tau deposition out to 2.5 hours, adding errors to the measurement if scan start-time is not precise. Our goal was to develop a statistical method for increasing the validity and reliability of ^{18}F-AV-1451 tau PET signal in a clinical trials setting, motivated by the need for efficient, correct measurement of tau using ^{18}F-AV-1451 SUVR 80–100. Any such advancement in the ability to reliably quantify tau accumulation has the potential for accelerating clinical trials, maximizing the benefit-cost ratio, and bridging imaging results to larger clinical applications.

B. Offset Scan Times in SUVR 80–100

Several challenges come with SUVRs in general, particularly with shortened SUVR 80–100 min scan times. Scan start-times may be offset due to material constraints and human error. Operator error and measurement error are common in PET scanning sites [10]. Scan-time errors for oncological PET studies starting at 80 min are on average two minutes early and have an uncertainty of up to 16 minutes [26]. Pilot studies with SUVR 80–100 only start at 80-min 60% of the time [22].

The simplest way to handle PET SUVR scan-time offsets is to ignore them. Previous research [4], for example, has shown that scan-time errors of up to two min for ^{18}F-CLT (Cerenkov Luminescence) PET SUVRs do not need correction. However, in tracers that are slow to reach steady state like ^{18}F-AV-1451, these assumptions may not be accurate. Some [22] have argued that correction by interpolation is needed for ^{18}F-AV-1451. Because ^{18}F-AV-1451 continues to increase relative to the reference region in brain regions with high uptake, slightly offset scan start-times result in inaccuracies in SUVR measurement.

C. Existing Methods to Correct for Scan-time Error

We first highlight some general treatments of PET SUVRs for AD-related biomarkers. Several authors have characterized SUVRs as log-normal and justified usage of log transformation [24, 25]. Moreover, existing studies only consider SUVRs greater than or equal to 1.3 to reflect ^{18}F-AV-1451 binding to tau deposition [22]. In later sections we describe a method that will incorporate a log-transform for SUVRs ≥ 1.3 but not apply any transforms for SUVRs ≤ 1.3.

There is some discussion regarding treatment of PET scanning errors for non-AD related studies. Van der Hoff [26] proposed a correction method based on tracer kinetics for oncological PET. However, the method requires detailed knowledge of the individual scans. We propose a correction method specifically for SUVRs from 80–100 min binned in 5-minute intervals for tau tracer ^{18}F-AV-1451 that requires prior minimal assumptions or information pertaining to scanning conditions. For the rest of the study, we focus on scans with 4 × 5 min frames. Though many scanners may have more advanced capabilities, such as list mode, acquiring data in 5 minute frames is standard for multi-site studies like ADNI. Furthermore, we make the practical assumption that accessing the most advanced settings of scanners are outside the skillset of most PET technologists.

We now discuss specific existing methods to treat SUVRs in the form of 4×5 min frames. Initial studies [22] found that SUVR 80–100 min are prone to error. One proposal corrects offset start-time error by using an interpolation method to predict where the 80–100 min mean SUVR should be relative to offset 5-min SUVR frames. The method calculates SUVR regression slope with respect to the four time points, then estimates SUVR 80–100 based on this slope and the offset mean. This method, applicable to voxels and ROIs, is used for SUVRs ≥ 1.3 (lower SUVRs, treated as ‘non-binding’, are omitted). We refer to this method as ‘interpolation’.

Another simple way to correct for offset SUVRs in the form of 4×5 min frames is by averaging the 5-min frames that fall within 80–100 min. Using this correction, a 75–95-min scan would omit the first 5-min frame from 75–80 min. We refer to this method as ‘mean**’. The mean** method is an approach that would minimize potential biases due to differences in acquisition time, but would decrease the signal to noise of the SUVR due to shorter imaging time adding variability.

In addition to the interpolation and the mean** methods, we posit that the standard procedure to is to ignore the offset starttime and leave the data uncorrected, averaging across all 4 frames; we refer to this procedure as the ‘mean’ method.

We assessed the three methods described above using error comparisons, and also explored a new method of weighted means using segmented least squares (SLS). This study aimed to (1) detect and quantify expected error produced by scans starting from between 75–85 min, (2) develop and optimize the SLS method on a training dataset, (3) compare SLS to existing methods in rectifying these errors in a test set, and (4) model the benefit of using the different corrections on statistical power in a simulated drug trial. We compared this approach to the interpolation, mean**, and mean methods.

II. Methods

A. Description of Subjects

Our sample contained a total of 40 subjects (table 1), including 14 healthy normal controls (NC), 9 AD patients, and the rest with mild cognitive impairment (MCI), frontaltemporal dementia (FTD), Parkinson's Disease (PD), and Corticobasal Syndrome (CBS). We applied a training/test methodology across this sample to reliably assess our approach and to ensure that it was robust across a range of data. We split the sample into a 'training set' and a 'test set' of similar demographic characteristics in diagnosis, age (within diagnosis), and gender, to ensure methodological validity and power across multiple sources of data. Nineteen subjects (11M/8F) were in the test set: 7 NC (3M/4F, mean age 78±4 years), 4 AD (2M/2F, mean age 68±10 years), 3 MCI, 3 FTD, 1 PD, and 1 CBS. Twenty-one subjects (10M/11F) were assigned to the training set: 7 NC (2M/5F, mean age 77±7 years), 5 AD (1 early onset AD, 1M/4F, mean age 60±6 years), 3 MCI, 2 FTD, and 4 PD.

B. Data Acquisition

Subjects were injected with approximately 370 MBq ^{18}F-AV-1451 and scanned from 75–115 min post-injection on a Siemens Biograph Truepoint 6 PET/CT. We included subjects who moved minimally, judged by sufficient PET/CT overlay, for the 40-min scan. Data were reconstructed using 11 start-times from 75–85 min. List-mode PET data were binned according to start-time into 4×5 min frames and reconstructed using ordered subset expectation maximization smoothed with a 4mm3 Gaussian kernel. These scans were reconstructed with iterative OSEM 2D with 4 iterations and 21 subsets with zoom=2, matrix size 336×336×109, voxel size 1.0182 × 1.0182 × 2.027 mm [8]. Scans were corrected with measure attenuation correction and used model-based Compton scatter correction [28, 29]. We also reconstructed the test group using (i) filtered back-projection with Gaussian 4mm, (ii) iterative reconstruction with Gaussian 6mm, and (iii) iterative reconstruction with Gaussian 0mm in order to test how applicable corrections are to different resolutions and reconstruction types.

Magnetic resonance imaging (MRI) was performed for each subject, using T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequences. These scans were used to define regions of interest (ROIs) for PET data analysis. Control subjects received MPRAGE on a 1.5T Siemens Avanto MRI (TR=2110 ms, TE=3.58 ms, 1×1×1mm voxel size) while non-control subjects received MPRAGE on a 3T Siemens Magnetom Trio (TR=2300 ms, TE=2.98 ms, 1×1×1mm voxel size). Native-space MRIs were segmented into ROIs by FreeSurfer v5.3 (http://surfer.nmr.mgh.harvard.edu/).

Within each offset start-time, the 4×5 min frames were realigned using SPM12 (http://www.fil.ion.ucl.ac.uk/spm), then averaged to create a 20-min SUVR. PET scans were coregistered to MPRAGE, then SUVRs normalized by cerebellar gray were calculated for 68 cortical ROIs for each subject and each offset start-time. For this study we compared these normalized cortical ROIs across subjects, treating them as a representative sample of tau-accumulating regions.

C. Segmented Least Squares (SLS) Model

Due to the slow ^{18}F-AV-1451 dissociation rate, we hypothesized different behavior in ROIs with different amounts of ^{18}F-AV-1451 accumulation. Data exploration supports this proposition. For example, tracer dynamics are different for the right middle-temporal ROI between an NC and an AD subject: the AD ROI increases for all offset-time points in a non-linear manner, while the NC ROI stays stable (Fig. 1).

We quantified the precise extent to which the error was related to amount of time offset. First, we calculated offset scan-time errors for a range of starting times between 75–85 min, reconstructed using 4×5 min frames. Using these reconstructions, we were then able to propose an alternative method of correction that outperforms the uncorrected SUVR and the interpolation method.

Fig. 2 depicts distributions of low SUVR values (blue lines, normally distributed), high SUVR values (pink lines, not normally distributed) and log of high SUVR values (red lines, closer to normally distributed than pink lines). Existing literature supports this finding: prior studies have addressed dichotomizing SUVRs to high and low values [13, 27] as well as fitting SUVRs to a lognormal distribution [24, 25]. Thie et al. 2000 [25] qualify the distributional structure of SUVRs due to its being modeled as products of rate constants, which can lead to lognormality. Biologically significant SUVRs are modeled as a product of biological factors such as tissue accumulation rate, blood clearance rate, and the Gjedde-Patlak accumulation rate [25]. Since high SUVR values yield signal, they are subject to the above assumptions. Lower SUVRs reflect less tracer binding and are often regarded as errors [21]. We therefore conclude that low SUVR values can be modeled as Gaussian.

Based on the above justifications, we hypothesized that the full sample of ROI SUVRs may be best described with an ordinary least squares model for low values and a log-log linear model for high values. Fig. 2 shows us that subjects exhibit different distributions of SUVRs, and these concentrations seem to be differentiable by a numerical cutoff at around 1.3. In equation (2) below, we try a range of cutoff values from 1 to 3 as these values large span the density curves in Fig. 2. We systematically compared the above with normal and lognormal models for all SUVRs and found that the normal/lognormal dichotomization model consistently yielded the smallest sum-of-squared errors (Table A.4).

We found that approximate empirical distributions of the ROI-wise SUVR values supported our hypothesis (Fig. 2). A cutoff at 1.3 (which we will further explain in the following section) separates the pool of training SUVR ROIs into an approximately normal distribution for low SUVRs and approximately log-normal distribution for high SUVRs (Fig. 2) at 80–100 min for both NC and AD subjects. Log-transformed SUVRs above 1.3 (for both total subjects and AD subset) approximate normal distributions. Distributions for all 20-min SUVRs (offset and 80–100 min) are approximately log-normal for high values and normal for low values. Such a practice has been described in the literature [24, 25].

To empirically justify the above claims, we used the Kolmogorov-Smirnov (KS) test to evaluate normality for of the high and low SUVRs from 80–100 min. For total and AD subjects, bootstrapped p-values for low SUVR values did not yield a significant departure

from simulated data adhering to a normal distribution with the same characteristics (.12 for total, .26 for AD-only). However, for high SUVR values, the test yielded p-values low enough to conclude a significant departure from normality (.001 for total, .02 for AD-only). When the high SUVR values are log-transformed, however, p-values are .03 for total SUVRs and .12 for AD-only subjects. These results lend evidence to the claim that high SUVRs are log-normally distributed and low SUVRs are normally distributed. This observation is reflected in the shape of the distribution curve in Fig. 2 (top left). That log-transformed high SUVRs are somewhat discrepant from a normally simulated dataset is superseded by goodness-of-fit in the lognormal linear model (Table A.4).

We combined the high and low models into a single segmented regression system that optimizes parameters with respect to log-likelihood from both equations. This method is often used in data that experiences structural change [2, 12] and previous literature as well as exploratory evidence (Fig. 1) has demonstrated inherently different underlying statistical distributions for high and low SUVRs.

Assuming that offset SUVRs are treated as if they were ‘uncorrected’ and simply averaged (mean method), Fig. 3 (Table A.1) shows the percentage errors (range or errors: −4 to 6%) of offset SUVRs of 68 Cortical ROIs for all subjects in the training set. When scan-time starts at 75 min the average percentage difference is −.14±1.24% but for scans where start-times occur after 81 min, the difference nears +1% for scans starting at 85 min, the error is . 89±1.48% on average.

The difference is even larger when isolated to AD subjects only: errors average between +1.01±.88% and +1.99±1.49% in offset scans after 80 min and exceed +3% in top quartiles of later scans. Table A.1 shows that the errors from the mean method both are biased and increase in variance as scan-start-times move farther from 80 minutes. These errors are sizeable given that existing longitudinal studies for the accumulation of tau [22] estimate the expected difference in tau accumulation for AD patients as 0.08 (4.5%) and MCI patients as 0.03 (2.5%) within a span of 18 months. A 5-minute scan start-time offset would on average induce errors that are up to half of the desired accumulation effect.

The SLS model splits the means of the 4 five-min-frame regressor variables (equivalent to 20-min SUVR) into *high* and *low* groups and fits a linear model to the low SUVRs and a weighted (by log-transformed SUVR 80–100) log-log model to high SUVRs. The model is written as follows:

$$\log\left(y_{i,h}\right) = \log\left(Y^{T}_{i,h,t}\right)A_t + e_{i.h}$$
$$y_{i,l} = Y^{T}_{i,l,t}B_t + e_{i.l} \quad (1)$$

In the above equations $y_{i,j}$ represents high (h) and low (l) mean SUVRs from 80–100, and $Y_{i,j,t} = [1\ y_{1,i,j,t}\ y_{2,i,j,t}\ y_{3,i,j,t}\ y_{4,i,j,t}]^T$ represents the vector of five-min SUVR frames and

intercept (represented by 1) for i number of ROIs at scan-starttime t = 75, 76, .., 85 min, with $j=1$ or h. $A_t=[A_{t,0}\ A_{t,1}\ A_{t,2}\ A_{t,3}\ A_{t,4}]^T$ and $B_t=[B_{t,0}\ B_{t,1}\ B_{t,2}\ B_{t,3}\ B_{t,4}]^T$ are the high- and low-SUVR coefficients specific to scan-time t. Errors for high and low SUVRs are represented by $e_{i,h}$ and $e_{i,j}$. The inner product of vector log($Y_{i,h,t}$) and A_t is computed and added to error $e_{i,h}$ to obtain high SUVR estimates, and the inner product of vector $Y_{i,l,t}$ and A_t is computed and added to error $e_{i,l}$ to obtain low SUVR estimates.

Each $Y_{i,h,t}$ is each five-min frame for SUVRs whose means are greater than a certain cutoff C, and $Y_{i,l,t}$ represents SUVR frames with means less than C. $Y_{i,h,t}$ and $y_{i,h}$ are log-normally distributed and $Y_{i,l,t}$ and $y_{i,l}$ are normally distributed. We fit two linear models to the high and low SUVRs to obtain the optimal weights A_t and B_t across t = 75, 76, …, 85 min and the optimal cutoff C

$$\{A_t, B_t, C\} = \underset{\{A,B,C\}}{\arg\min} \sum_{i}^{n_h} \left(y_{i,h} - \exp\left(\log\left(Y_{i,h,t}^T\right)A_t\right)\right)^2 + \sum_{i}^{n_l} \left(y_{i,l} - Y_{i,t,t}^T B_t\right)^2, \quad (2)$$

where $y_{i,l} \le C < y_{i,h}$.

The method for deriving optimal weighting parameters is:

1. Increment the cutoff point *(C)* by a small threshold ranging from SUVRs of 1 to 3,
2. Calculate optimal weights *At* and *Bt* to minimize log likelihood for low- and high-SUVR ROIs for every possible combination of high and low SUVR groups divided by cutoff,
3. Average sum-of-squared errors (SSE) for high and low groups across start-times from 75 to 85 min and select the SUVR threshold that yields lowest SSEs for SLS model,
4. For cortical regions across the 21 training subjects, split high/low SUVRs at the optimal cutoff with lowest squared errors, then derive the high and low sets of training-set parameters and fit these upon the test set.

The average residual errors across offset start-times reached a minimum of 1.3 in the training set (Fig. 4). The optimally-trained coefficients A and B were applied to ROIs in the test set and generalizable to any other ROI 4×5 min SUVR frames for ^{18}F-AV-1451, normalized by cerebellar gray (Table 2).

The optimal weights for each offset time (Table 2) represent the crux of our quantitative results and are applicable to future scans for ^{18}F-AV-1451. The table is split between *low* (SUVR below 1.3, represented by B_t) and *high* (SUVR above 1.3, represented by A_t). $A_{t,0}$ (for SUVRs>1.3) and $B_{t,0}$ represent intercepts (or base values), and each subsequent term

represents the weight the 5-min offset SUVR is given to be factored into the modeled SUVR 80–100 value. A correction for a ROI whose mean SUVR is below 1.3, scanned at 85 min, for example, will be to multiply the first frame (85–90 min) by .34, the second (90–95 min) by .28, the third (95–100 min) by .30, and the fourth (100–105 min) by .06. The weighted SUVRs are summed and then added to the intercept term of 0.02 to produce the corrected SUVR 80–100 value. If the mean SUVR is higher than 1.3, then the 5-min SUVRs are log-transformed, weighted by {.35, .33, .27, .03}, summed (in this case the intercept term is 0), and exponentiated.

We compared SLS to the three methods (mean, mean**, and interpolation) mentioned above to determine the best method for handling SUVRs that start at slightly offset scan-times. All 4 methods of correction were performed on subjects in the test set and across all offset start-times. We also verified our derived weighting parameters on test subjects reconstructed using: (i) filtered back-projection 4mm, (ii) iterated Gaussian 6mm, and (iii) iterated Gaussian 0mm. We fit ROIs from test set ROIs to these reconstruction parameters to see how well our model generalized to other scanning scenarios.

Means and standard deviations of errors between the 4 correction methods were compared and empirically sampled and used to simulate scan-time errors for clinical trials. We compared the mean percentage errors of the SLS model with the mean, mean**, and interpolation models for all subjects and AD-only subjects. Results are discussed in section III.

D. Monte Carlo Simulation of Tau Clinical Trials

We validated our method modeling a drug trial using Monte Carlo simulations. In the absence of tau-based therapeutic trials, simulations were modeled after a longitudinal tau studies [23, 9] and amyloid (another common pathology related to AD) therapeutic trials [18]. The simulations were built from several assumptions based on empirical data and previous studies. We describe the simulation assumptions made for (1) ROI values across subjects, (2) longitudinal errors for each scan, unrelated to offset scan start-time, (3) frequency of start-time offsets and associated errors, and (4) simulation for effects of treatment.

The baseline simulations (prior to simulated treatment and follow-up) have identical parameters for treatment and placebo groups. The treatment and placebo groups are both subject to scan-time and longitudinal errors. Each subject yields a measurement for some ROI that is distributed in the same way as the collection of training set cortical ROIs:

$$y_{i,S_1}(t) = y_{i,true} + N_{i,S_1}(t) + N_{i,long}\,. \tag{3}$$

In the above equation, $y_{i,t_{true}}$ represents 'true' SUVR 80–100 values generated from their empirical distributions as in (Fig. 2), $y_{i,S_1}(t)$ is the SUVR 80–100 at a given ROI in the baseline longitudinal scan measured at time t, $y_{i,\ true}$ is simulated SUVR 80–100 with no errors from other sources. $N_{i,\ long}$ represents longitudinal errors present in the first and second scans and is normally distributed with mean 0 and standard deviation of 0.04.

'True' SUVR values are simulated using the 'ks' package from CRAN to generate distributions from empirical kernel densities [7]. Longitudinal errors represent variations in SUVR induced by small differences in setting arising from different scanning sessions representative of other factors not accounted for by scan-time errors and variability surrounding treatment and growth effects. Longitudinal errors could include image errors, errors in corrections, patient motion, differences in patient placement and differences in PET/MRI coregistration, and is estimated based on results from trials of Bapineuzemab, an anti-amyloid agent. The error magnitude in our simulations is roughly similar to longitudinal errors in amyloid trials [16, 18].

The focus of the simulation was to evaluate the effect of the scan-time error $N_{i,S_1}(t) \cdot N_{i,S_1}(t)$ is the error if the scan was offset by 1 to 5 minutes and is discussed at length in section 1.B. Each scan $y_{i,true}$ starting at baseline time S_1 is offset by time t.

1) ROI values across subjects—We simulated ROI SUVR values for subjects in drug trials by using information from our training set. Because the distributions were qualitatively different (Fig. 3), we separately simulated high-SUVR and low-SUVR ROIs. SUVR 80–100 was generated from a mixture of normally-distributed variables for SUVRs under 1.3, and log-normally distributed SUVRs over 1.3. Our analysis in section II.C as well as other authors who discuss the lognormality of SUVRs over 1.3 justify this decision [24,25]. The rate of high and low SUVR incidence and the parameters of the normal and log-normal distributions were based on the SUVR densities from training set subjects. Though our preliminary findings suggest that low-valued SUVRs exhibit a normal distribution, we also simulated low values with lognormal data as an alternative characterization (Table A.5, A.6). While the advantage is not quite as stark, SLS is still preferable across a range of simulation parameters.

2) Longitudinal Errors—Longitudinal errors for this simulation were based on florbetapir scans for amyloid obtained in the Alzheimer's Disease Neuroimaging Initiative (ADNI), as well as longitudinal studies done on ^{18}F-AV-1451 [23, 9]. Longitudinal errors represent errors independent of offset scantime errors, such as patient motion, registration errors, and image noise errors. The ^{18}F-AV-1451 study [23] cites overall uncertainties of around 0.08 in SUVRs of AD subjects and 0.12 in SUVRs of MCI subjects, but do not differentiate these uncertainties, though the aggregated uncertainties roughly match those of our simulations. We primarily use florbetapir data to simulate longitudinal errors as described in [18]: between baseline and follow-up scans of florbetapir temporal ROI SUVRs, longitudinal correlations were very high at 0.97, but some longitudinal errors existed. Using principal components analysis, we found that percent error due to longitudinal errors was approximately 3% in both scans, assuming relatively equal variation

in first and second scans. Therefore, we added 3% longitudinal errors (approximately a standard deviation of 0.04) for first and second scans for each simulated subject.

3) Scan-time Errors: The focus of the simulation is to assess the impact of offset-scan times, which we refer to as scan-time errors. $N_{i,S_1}(t)$ represents errors associated with empirical scan-time errors for the four methods (SLS, mean, mean**, and interpolation), for $t = 75,76, \ldots, 85$ min. This assumption is based on empirical evidence that many scan sites start early or late by a few minutes. [22] showed in a small sample that scans intended to start at 80 min can actually begin ≥ 1 min offset (e.g. 79, 81, 77, or 84 min) nearly 40% of the time.

We used two schemes to simulate scan-time error. In the first scenario, we postulated that most sites mirror the starttime precision rate in [22] of 60% for 80-min scans, 10% for 79 and 81 min scans, 5% for 2-min offsets, 2.5% for 3-min offsets, 1.5% for 4-min offsets, and 1% for 5-min offsets. This assumption may be more conservative than a realistic scanning scenario, so we supplemented this assumption with a more liberal estimation of start-time precision rate, with a 30% probability that scan acquisition starts at 80 min, 12.5% for 79 and 81, 7.5% for 77, 78, 82 and 83, 5% for 76 and 84, and 2.5% for 75 and 85 min.

Errors associated with each start-time was added to each scan for every model, drawn from empirical kernel densities of each error distribution for every scan-time. Like in the simulations of 'true' SUVR values, we used the 'ks' package to generate simulated distributions from empirical kernel densities of scan-time errors [7].

4) Simulation of Treatment and Growth: The ^{18}F-AV-1451 study [23] show that the growth rate for AD patients are 0.08 ± 0.08 (approximately 4.5% growth with mean SUVR of 1.75) for AD patients and 0.03± 0.10 (approximately 2.5% growth with median SUVR of 1.25) for MCI patients over a span of 18 months. The other longitudinal ^{18}F-AV-1451 study showed an annual increase of 3% (0.0460.073 depending on ROI) in ADs and posited 25% annual reduction in the accumulation of tau in their hypothetical model of drug trials for tau reduction [9].

Longitudinal tau PET studies were somewhat empirically similar in growth behavior compared to amyloid, which was treated in the Bapineuzumab trial. We posit that treatment effects may also be similar across the two biomarkers. In Bapineuzumab drug trials [18], treatment group 11C-PiB SUVRs increased by an average of 0.001±0.021, while placebo group SUVRs increased by 0.102±0.026 (25% error) after 78 weeks. The treatment effect was −0.101±0.034, corresponding to approximately 4.5% of the baseline SUVR, and treatment errors had variance of roughly a third of treatment size.

We simulate the longitudinal change (growth) of tau assuming that the treatment effect is equal in mean but with differing variations: that is, the treatment effectively inoculates the growth effect. This assumption is motivated by empirical effects in amyloid clinical trials [18]. Effect E is informed by the ^{18}F-AV-1451 growth rates in the ^{18}F-AV-1451 study [9]. Growth rate $G_I(E)$ for SUVRs under both treatment and placebo regimes is normally distributed with mean E and standard deviation of .25*E, corresponding to 25% error in the

Bapineuzumab trials. We assumed that reduction and growth rates were approximately equivalent, as in [18]: our simulated treatment $T_i(E)$ attenuated ^{18}F-AV-1451 SUVRs at varying rates normally distributed with means -E with 33% error variance. Randomized scan-time errors $N_{i,S_2}(t)$ associated with scan-time offsets t are included in addition to treatment and growth effects, and the simulation model is represented as follows:

$$y^{treat}{}_{i,S_2}(t,E) = y_{i,true} + G_i(E) + T_i(E) + N_{i,S_2}(t) + N_{i,long},$$
$$y^{placebo}{}_{i,S_2}(t,E) = y_{i,true} + G_i(E) + N_{i,S_2}(t) + N_{i,long}.$$

(4)

$y^{placebo}{}_{i,S_2}(t,E)$ and $y^{treat}{}_{i,S_2}(t,E)$ represent SUVRs (with errors added) of second longitudinal scans from 80–100 min, respectively designating the simulated placebo and treatment groups for tau reduction. The final two terms represent longitudinal and scan-time errors at second longitudinal scan S_2, generated in the same way as detailed in section II.D. Differences between the treatment and placebo SUVRs were evaluated to compute effect size.

5) Effect Size Comparison: Effect size for longitudinal differences between treatment and control groups was calculated using Cohen's d, a mean difference measure normalized by pooled standard deviation. In the following equation, we consider the test statistic, Cohen's d, which depends on the means of paired differences between treatment regime observations $y^{treat}_{S_2}$ under the second scan and y_{S_1} under the first scan S_1, and between paired differences between placebo regime observations $y^{placebo}_{S_2}$ and y_{S_1}. The effect of the treatment is captured by the average of the paired differences $\overline{\left(y^{treat}_{S_2} - y_{S_1}\right)} - \overline{\left(y^{placebo}_{S_2} - y_{S_1}\right)}$. Each simulated y_{S_1} corresponds exactly to a simulated y_{S_2}, so we use a pooled standard deviation s to normalize derive Cohen's d.

$$d = \frac{\overline{\left(y^{treat}_{S_2} - y_{S_1}\right)} - \overline{\left(y^{placebo}_{S_2} - y_{S_1}\right)}}{s}$$
$$s = \sqrt{\frac{(n_{treat}-1)s^2_{\Delta treat} + (n_{placebo}-1)s^2_{\Delta placebo}}{n_{treat}+n_{placebo}-2}}.$$

(5)

In the first equation, $y^{treat}{}_{S_i}$ and $y^{placebo}{}_{S_i}$ are SUVRs from baseline and second scans S_1 and S_2, n_{treat} and $n_{placebo}$ are fixed at 50 samples, s is the pooled normalized standard deviation for the difference of the first and second trials across treatment and control groups. $s^2_{\Delta treat}$ is the sample standard deviation of pretrial-posttrial differences $\left(y^{treat}_{S_2} - y_{S_1}\right)$ for the treatment group and $s^2_{\Delta placebo}$ is the sample standard deviation for the placebo group $\left(y^{placebo}_{S_2} - y_{S_1}\right)$.

We calculated Cohen's d for treatment and control groups and then derived the minimum required sample size (independent of how many samples we simulate with) for significant results with fixed power. Cohen's d is inversely related to power, significance, and sample size. When power and significance are fixed, a larger Cohen's d corresponds to a smaller required sample size and vice-versa. We used the 'pwr' package on CRAN to determine the number of *required participants* to ensure that statistical power is fixed at 0.9 (as in [18]) and significance of $p < 0.0001$ [5, 15].

We calculated a simulated effect size value by calculating Cohen's d between 100 generated samples in the treatment and placebo groups (number of samples is arbitrary and unrelated to *minimum sample size* for significance converted from effect size). We replicated the simulation 10,000 times and averaged Cohen's d to produce a robust effect size estimate. Finally, we converted the average Cohen's d to the minimum sample size required for significant results with statistical power. We ran the simulation on the total set of training ROIs as well as AD subjects' ROIs only. Simulation of AD-only ROIs may be more useful from a clinical perspective as AD subjects' treatments may provide more informative results for clinical studies of tau accumulation.

III. RESULTS

A. Percent Bias Comparison in Test Set

We fit the model described in Section II.D to test set ROIs under the scanning and reconstruction parameters described in Section II.B. When scans start early or late, the 20-min SUVR carries a small percentage error compared to the true 80–100 min SUVR dependent on how distant the offset start-time is from the 80-min mark.

Density plots of percentage errors (with and without corrections) for iterative reconstructions with 4mm smoothing show reductions in variance and bias in the SLS method compared to mean, mean**, and interpolation methods (Fig. 5). Standard deviations of errors are smallest for SLS compared to other methods. For scans starting after 80 min, mean, mean**, and interpolation methods yield percent errors up to 5% (Fig. A.2). SLS produces the smallest errors compared to other methods, especially for later scans starting at 81–85 min. SLS is the only method whose errors are consistently centered around zero in late start-times for AD subjects (Fig. 5b). SLS is also the only method whose high and low SUVRs have the lowest bias and variance.

When fitting training set parameters upon the total ROIs in the test set, the mean**, interpolation, and SLS methods all yielded smaller errors compared to the raw mean. In ROIs of AD subjects for scans starting at 85 min, for example, offset errors induced slightly higher mean percentage biases (0.9±1.4%) from interpolation than from SLS (0.7±1.3%). In scans with start-times from 85 min, differences in percent bias in AD subjects average 2. 1 ±0.1 % between the SLS method and mean, 1.2±0.1% between SLS and mean**, and 0.2±0.1% between SLS and interpolation (Fig. A.2, Fig. 5).

B. Alternative Reconstruction Parameters

While it is not feasible to determine whether SLS parameters are entirely dependent on the scanner itself because we do not have access to data acquired on other scanners, we can approximate differing scanning conditions with other reconstruction conditions. We assessed the impact of the SLS method on other reconstruction parameters in order to gauge potential impacts in other scanning scenarios. We explored these factors by using different reconstruction types and amounts of smoothing to reflect different resolutions. The SLS parameters generated with 4mm Iterative Smoothing were used to fit data from filtered backprojection smoothed at 4mm and Iterative OSEM 2D smoothed at 0mm and 6 mm in addition to examine the generalizability of our method. The SLS parameters (from Table 2) are fitted onto test set ROIs with different reconstruction parameters. With the exception of the biases across all correction methods in earlier scan-offsets, the SLS method outperforms all the other methods. The test-set errors resulting from fitting SLS parameters show general agreement between different reconstruction parameters but with some discrepancies at larger scan-offset times. Because the SLS errors do not differ significantly between different reconstruction parameters (table 3, table A.1), we found overall agreement in the conclusions drawn from the discussion of SLS results in section III.A. Furthermore, since the SLS parameters can be extended to other reconstruction settings, we infer that they may be extrapolated to scanners that cannot acquire data in listmode and therefore cannot be rebinned with a precise starttime of 80 min.

The larger the deviation from starting the scan at 80 min, the better the SLS method corrected for SUVR differences even across reconstruction parameters. Other than for 79 and 81 min, the SLS method is just as robust for FBP 4mm and Iterative OSEM 2D smoothed at 0mm and 6 mm (table 3). For scans only off by 1 min (starting at 79 and 81 min), SLS produces slightly higher mean absolute error by a factor of 0.01–0.08% compared to other methods in all alternative reconstruction scenarios. SLS does not out-perform other methods at short start-times under other. We posit that SLS out performs the interpolation method because when tau is present, it is increasing non-linearly. However, it is likely that at small scan time offsets, the increase would be appear linear. Therefore, effects due to reconstruction differences dominate non-linear weighting effects when scan-time offsets are small but when they are large, the SLS weighting parameters correct for the underlying non-linearity induced by scan-time errors. As such, it is advisable to forgo the SLS method for another method that assumes linear errors when offsets are very small under alternative scanning parameters. But the SLS method holds for larger deviations in scan time.

C. Effect Size Comparisons for Simulations

We further analyzed effects of SLS correction applied to reconstructions with Iterative OSEM 2D smoothed at 4mm using effect size analysis. When testing differences between the four methods using Monte Carlo simulations, we found better results for the SLS method over a range of adjusted treatment values (table 4). Treatments with average values ranging from 0.01 to 0.05 require an inversely proportional number of participants: for a small treatment of 0.01 an effective study would require up to 2000 subjects, but if the drug treatment effect were 0.05, the study would only require approximately 100 subjects.

When considering simulations over both the total set of available ROIs and those restricted to AD subjects (table 4), the SLS method consistently required fewer subjects than others. The mean method consistently yields between 10% to 20% more subjects required than the SLS method. The interpolation method yields better results than either of the mean methods but is outperformed consistently by the SLS method. Though the difference is at times very small, we will see in the following section that such a small difference actually corresponds to a notable reduction in required participants in a simulated study.

Under the assumptions of the Jack et al. 2018 [9] and Southekal et al. 2017 [23], tau should have a growth rate of between 0.01 to 0.02. Assuming that a hypothetical drug effectively inoculates growth, as in Bapineuzumab studies [18], drug trial simulations consistently privilege the SLS method. If SUVRs grew by 0.02 (and treatments reduced SUVRs by 0.02), under any scanning-error assumptions, the SLS method would require 70–100 fewer subjects compared to the uncorrected mean, 50–70 subjects compared to mean**, and 18–40 compared to interpolation. If, instead, treatments were larger (matched to the scale of the Bapineuzumab trial relative to baseline SUVRs of 2.2 for PiB-PET), equivalent treatment sizes would be approximately 0.05 for all training subjects. In this case, the SLS reduced around 15–20 subjects compared to the uncorrected mean, 5–7 subjects compared to mean**, and 3–4 subjects compared to the interpolation method.

Because the effect size studies are simulated, we do not expect them to reflect the exact number of required participants. However, the SLS method consistently yields reduced sample size compared to other methods. Furthermore, differences are amplified when treatment effects become smaller. A pilot drug will draw larger differences between SLS and the other three methods if treatment size is low. The advantages of the SLS method appears to span across ‘conservative’ and ‘liberal’ scan-time error assumptions. See table 4 for sample sizes under the full range of treatment values. The required sample size will also increase dramatically if longitudinal errors are increased to reflect increases in errors due to factors such as patient motion or image noise, but the SLS method is still consistently preferable to others. Increases in longitudinal errors scales the differences in required sample size multiplicatively, so the method makes a larger difference even if scan-time offsets are a small source of error compared to others.

IV. Discussion and Conclusion

A. Discussion and Implications

Though the SUVR 80–100 min window is widely used for ^{18}F-AV-1451, slow washout of this tracer coupled with small scan start-time offsets produces errors. Left uncorrected, errors in scans are large: scans from 85–105 min cause 2 ± 1.5% distortion from the 'true' 80–100 SUVR in some time-frames (Fig. 3, Table A.2), which are roughly half the size of empirical growth rates in longitudinal studies [9, 23]. By correcting for SUVR estimates using a series of optimized weights drawn from segmented linear modeling, the estimates become more precise as errors' skew and variance are minimized. Clinical simulations show that the SLS method allows potential treatments to reduce up to 20% of subjects needed to achieve fixed 90% power and significance of $p<0.0001$.

The weighted mean method of correction we piloted in this study yields immediate and ramified implications. The SLS method may be applied to clinical trials for pharmaceutical research involving SUVR 80–100 for ^{18}F-AV-1451, and could become more widely used as ^{18}F-AV-1451 gains more traction in tau imaging. The coefficients derived in the SLS model are based on normalized SUVR values across subjects and provide information about SUVR behavior in relation to each other in a given sample of ROIs. As such, the optimized correction weights allow for different SUVR-scales across different baseline scanning conditions and reconstruction parameters.

1) Immediate Impacts—Our method is generalizable enough to be applied to ROI-wise ^{18}F-AV-1451 SUVRs from 80–100 min in any site, assuming initial reference region is cerebellar gray (scans can then be renormalized to a different reference region after weighting factors are applied). A clinical technician can directly apply the weighting parameters in Table 3 to offset time-scans for ^{18}F-AV-1451 SUVRs. This research results in quantifiable information for any multicenter study using ^{18}F-AV-1451 such as ADNI and the 4 Repeat Tauopathy Neuroimaging Initiative (4RTNI). Moreover, the SLS method produces better results than the mean, mean**, and interpolation methods on data reconstructed with different parameters from the test dataset (filtered backprojection smoothed with 4mm, and iterative OSEM 2D smoothed with 0mm and 6mm) in addition to the iterative OSEM 2D smoothed with 4mm reconstruction parameters. These checks allow us to say that our method is robust for a variety of resolutions and reconstruction methods. We describe limitations in section IV.B.

2) Impact for Clinical Trials—Though uncorrected means may be adjusted with simpler methods, such as the mean** method that omits 5-min-frames outside the 80–100-min range or with the interpolation method used in [22], we have shown that over a range of treatment and growth simulations, SLS is always the best possible method to rectify scan-time errors. Assuming tau growth rates are similar to the ^{18}F-AV-1451 longitudinal study [9], the SLS method can eliminate up to 20% more required participants than if SUVRs are not corrected.

We have demonstrated with Monte Carlo Simulations that the SLS method can empirically reduce time and resources necessary in acquiring research participants for pharmaceutical

companies and scanning sites to detect effectiveness in early phase studies. When assuming similar scanner settings as Gaussian 4mm, the SLS method is consistently the most efficient in simulations, making the method logistically the most preferable. We did not run simulations for different reconstruction settings corresponding to the errors in table 3, but given that the errors are much smaller under the SLS method (except for 1-min offsets), we infer that the SLS method will compare favorably to other methods. If corrections are forgone for 1-min offset, the SLS method will be even more advantageous.

Practically, eliminating required participants by decreasing offset-scantime errors is considerably impactful for a clinical trial. As a perhaps generous example, in comparing the SLS method to interpolation at an effect size of .01, reducing 195 subjects (1923 vs. 2118) at an estimate of $2,500 per PET scan could save nearly $500,000 in scan costs in a clinical trial, as well as expose many fewer people to ionizing radiation. Though they are simulations and as such an inexact measure, that they consistently reveal a reduction in required subjects attests to the effectiveness of the correction method.

Clinical trials for drugs are very sensitive to time offset errors in scans because it is crucial that biomarkers at very low concentrations are measured accurately. Indeed, Jack et al. (2018) [9] demonstrated that the increase in tau as measured by ^{18}F-AV-1451 is very small (annual increase of 0.046–0.073 depending on ROI, compared to 0.004 for normal controls). It is especially important to procure accurate scans in subjects at the early stages of the disease with small amounts of tau, as treatment is only possible before proliferation take place. Any removal of distortionary information will be necessary in this case to facilitate the development of drugs to treat patients in such a stage of Alzheimer's Disease. Therefore, correction of offset scan errors, however small, is a very important factor to the successful development of these drugs.

Our simulations of effect size rest on many assumptions regarding drug trials, and as such may be regarded as liberal estimates for reductions in required participants. However, even if the effects of scan-time correction were more modest than stated, even eliminating a few required participants will be of some significance in facilitating drug development. In a Phase 3 Alzheimer's Disease trial, for example, research participants require many more expenditures than just PET scan costs. Although millions of people are diagnosed with AD, the number of eligible participants for a drug trial can be impacted by factors such as access to scans and participation in competing studies, so it is imperative to be as cost efficient as possible. Indeed, reducing costs not only streamlines pharmaceutical development, but also affects drug companies' willingness to undertake clinical trials [21]. Furthermore, these cost-saving impacts will only increase as the effect size becomes smaller, which is the trend in AD clinical trials.

3) Methodological Impacts—PET tracer studies routinely differentiate high and low-valued concentrations for SUVRs and other quantifications [13, 27, 30]. ^{18}F-AV-1451 dichotomization is used in recent tau-PET imaging studies [3, 19, 22]. Most studies take the threshold arbitrarily or use young healthy control SUVRs as benchmarks. [27] was among the first to systematically discuss dichotomization of PiB-PET SUVRs using methods such as k-means clustering and Gaussian mixture modeling. Our work demonstrates another way

of partitioning high/low SUVR data using a grid-search method to minimize residuals in fitting offset 4×5 min frames to SUVR 80–100 min and may lead the way other similar approaches.

Other sites can further refine the SLS method with respect to ^{18}F-AV-1451 by replicating the method piloted in this paper with more samples or with longer scan-times. Furthermore, SLS weighting parameters may be derived for other radiotracers in any kind of PET imaging (not constrained to Alzheimer's Disease research) to create tracer-specific correction factors if there are any non-linear relationships between time and tracer washout. Usage of the SLS method in other tracer studies would provide clinical impacts similar to those detailed in the previous section.

B. Limitations and Future Research

Despite the experimental and simulation design factors that we take into account, several limitations of this study should be addressed in future work. Some limitations result from the fact that our data is sourced from only one site at Lawrence Berkeley National Laboratory, as dynamic multi-site ^{18}F-AV-1451 scans are not yet widely used. Multi-site investigations will yield more variation in scans and may provide a more generalized picture of scan-time error. LBNL scans are 40-min scans acquired from 75–115 min, so some characteristics may be slightly different from 20-min scans. Also, the lack of range in scanning times may be an issue, as time-errors only span from 75–85 min due to the reconstruction method. Rerunning the analysis on the suite of differing reconstruction parameters shows robustness in our method across differing initializations, but biases may exist in scans limited to a single site.

Other than two longitudinal ^{18}F-AV-1451 studies [23, 9], there are limited empirical longitudinal data for tau tracers. Additional studies will increase our understanding of tau behavior across time. Longitudinal ^{18}F-AV-1451 scans will let us better examine factors such as growth rate of tau and longitudinal errors and enable more accurate simulations.

Some limitations arise from not having more training/test subjects. 21 training and 19 test subjects allows for a large sample of pooled ROIs but does not provide enough data for robust analyses stratifying across ROIs. Though exploratory data analysis has shown that ROI data from multiple subjects can be evaluated from a single mixture distribution, some subject or ROI-specific stratifications may induce small biases in analysis. More detailed information could be gleaned from a voxel-wise study instead of ROI analysis.

In section III.B we compared results of SLS across different reconstruction parameters. In table 4 we observed that small scanning-time offsets (scans starting from 79 or 81 min) may be overcorrected by SLS across different reconstruction parameters, as the mean, mean**, and interpolation methods work slightly better. These overcorrections hint at small biases induced by different reconstruction scenarios. To avoid overcorrection, a technician may ignore errors or employ alternate methods for scans that start at 79 or 81 min. However, the SLS method has been shown to be more effective in correcting errors more than other methods for any scans off by more than 3 min, even across different reconstruction scenarios. We assume the SLS method is consistently effective across scanning scenarios

because it produces errors that are not statistically significant across distinctly different reconstruction parameters.

We encourage other researchers to test and refine the validity of the SLS weights specifically on ^{18}F-AV-1451 by replicating this study on other scanners under longer times or differing parameters, as well as the applicability of the SLS method on other PET radiotracers by replicating this study on those tracers.

Analysis of other sources of errors associated with PET scans may also constitute a further research direction. A model for clinical trials that incorporates scan-time errors in addition to more detailed treatments of co-registration errors, patient motion, image noise, and other sources of error may be useful in precisely analyzing the interacting effects of different types of error associated with PET scans. Such a study may be useful in producing more accurate effect size simulations for more efficient clinical trials.

C. Conclusion

The method piloted in this study may serve as a useful diagnosis for correction of scan-time errors. The SLS method corrects for scan start-time errors by accounting for the nonlinear rate of increase in high-tau ROIs in ^{18}F-AV-1451, which interpolation and averaging methods do not do. Though other errors such as patient motion, image noise, off-target binding and registration errors may account for greater overall error in using SUVRs to quantify tau, scan-time offsets are the only errors that lend themselves to the type of systematic analysis that we have conducted in this study. This study has analyzed the effect of scan-time offsets and simulated required sample sizes of while accounting for other possible sources of error such as longitudinal errors.

Replications of the SLS method may be run for other PET tracers to determine optimal weights for 5-min averages. We have demonstrated that a SLS weighted-mean approach outperforms a raw-mean or interpolation correction to make scans more accurate for SUVR 80–100 min for ^{18}F-AV-1451, but this methodology should be applicable for any tracer with slow washout. We have demonstrated that our analysis is robust for differing reconstruction parameters, and as such, replicable in a variety of differing initial conditions.

In the best case, the margin of error reduced by the SLS method can save up to 20% of required participants in hypothetical clinical trials for tau—facilitating conservation of time and resources. However, even if the effects of scan-time correction were more modest than stated, the cost and resources saved as well as the increased accuracy in scanning would still be a net positive. Any measure to conserve limited resources would be impactful for increasing the feasibility of drug development, especially when changes in biomarkers are small and highly varying. The novel method presented in this study can spur research and methods to further refine more accurate PET quantification for Alzheimer's-related neuroimaging and other radiotracer imaging studies as well.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the U.S. National Institute of Health Grant R01AG034570. Avid Radiopharmaceuticals enabled use of the [18F] AV-1451 tracer, but did not provide direct funding and were not involved in data analysis or interpretation.

References

[1]. Adler D (2005). vioplot: Violin plot. R package version 0.2. http://wsopuppenkiste.wiso.uni-goettingen.de/~dadler

[2]. Bai J, Perron P (2003). "Computation and Analysis of Multiple Structural Change Models." Journal of Applied Econometrics. 18 1–22.

[3]. Baker S, Price J, Lockhart S, He M, Huesman R, Schonhaut D, Faria J, Rabinovici G, Jagust W (2017). "Reference tissue-based kinetic evaluation of ^{18}F-AV-1451 in aging and dementia." J. Nuclear Med. 58(2): 332–338. $^{-}$

[4]. Bouchet F, Geworski L, Knoop B, Ferrer L, Barriolo-Riedinger A, Millardet C, Fourcade M, Martineau A, Belly-Poinsignon A, Djournessi F, Tendero K, Keros L, Montoya F, Mesleard C, Martin A, Lacoeuille F, Couturier O (2013). Calibration Test of PET Scanners in a Multi-Centre Clinical Trial on Breast Cancer Therapy Monitoring Using 18F-CLT. PLOS One. 8(3).

[5]. Champely S (2016). pwr: Basic Functions for Power Analysis. R package version 1.2–0. https://CRAN.R-project.org/package=pwr

[6]. Chien. DT, Szardenings A, Bahri S, Walsh J, Mu F, Xia C, Shankie W, Lerner A, Su M, Elizarov A, Kolb H (2014). "Early clinical PET imaging results with the novel PHF-tau radioligand [F18]-T808." J Alzheimer's Dis 38(1): 171–184. [PubMed: 23948934]

[7]. Duong T (2016). ks: Kernel Smoothing. R package version 1.10.4. https://CRAN.R-project.org/package=ks

[8]. Hudson HM, Larkin R (1994). Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans Med Imaging. 13(4): 601–609 [PubMed: 18218538]

[9]. Jack C, Wiste H, Schwarz C, Lowe V, Senjem M, Vemuri P, Weigand S, Therneau T, Knopman D, Gunter J, Jones D, Graff-Radford J, Kantarci K, Roberts R, Mielke M, Machulda M, Peterson R. Longitudinal tau PET in ageing and Alzheimer's disease. Brain. Awy059 [Epub ahead of print].

[10]. Kinahan P, Fletcher J (2010). PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy. Semin Ultrasound CT MR. 31(6): 496–505. [PubMed: 21147377]

[11]. Liu J, Wu S, Zidek J (1997). "On Segmented Multivariate Regression". Statistica Sinica._7: 497–525.

[12]. MacNamee RL, Yee S, Price J, Klunk W, Rosario B, Weissfeld L, Ziolko S, Berginc M, Lopresti B, DeKosky SS, Mathis C (2009). "Consideration of Optimal Time Window for Pittsburgh Compound B PET Summed Uptake Measurements". J Nuclear Med 50 (3): 348–255.

[13]. Ossenkoepple R, Schonhaut DR, Baker SL, O'Neil JP, Janabi M, Ghosh PM, Santos M, Miller ZA, Bettcher BM Gorno-Tempini ML, Miller BL, Jagust WJ,Rabinovici GD (2015) Tau, amyloid, and hypometabolism in a patient with posterior cortical atrophy. Ann Neurol. 7,(3) 338–42.

[14]. Pontecorvo M, Devous M, Joshi A, Lu M, Siderowf A, Arora A, Mintun M (2015). Relationships between 18F-AV-1451 (aka 18F-T807) PET Tau Binding and Amyloid Burden in Cognitively Normal Subjects and Patients with Cognitive Impairments Suspected of Alzheimer's Disease. Journal of Nuclear Medicine, 5ö(supplement 3), 245–245.

[15]. R Core Team (2013). R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria URL https://www.r-project.org/

[16]. Sahathevan R Linden T, Villamagne V, Churilov L, Ly J, Rowe C, Donnan G, Brodtmann A (2016). Positron Emission Tomographic Imaging in Stroke: Cross-Sectional and Follow-Up Assessment of Amyloid in Ischemic Stroke. Stroke; 113 (9).

[17]. Shcherbinin S, Devous MD, Schwarz AJ, Joshi AD, Navitsky M, & Mintun MA (2015). Region-dependent kinetics of the Tau PET tracer [18 F]-AV-1451 (T807). Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 11(7), 113.

[18]. Salloway S, Sperling R, Fox N, Blennow K, Klunk W, Raskind M, Sabbagh M, Honig L, Porsteinsson A, Ferris S, Reichert M, Ketter N, Nejadnik B, Guenzler V, Miloslavsky M, Wang D, Lu Y, Lull J, Tudor I., Liu E, Grundman M, Yuen E, Black R, Brashear H (2014). Two Phase 3 Trials of Bapineuzumab in Mild-to-Moderate Alzheimer's Disease. N. Engl. J. Med; 370 (4).

[19]. Schöll M, Lockhart S, Schonhaut D, O'Neil J, Janabi M, Ossenkoppele R, Baker S, Vogel J, Faria J, Schwimmer H, Rabinovici G, Jagust W (2016). "PET Imaging of Tau Deposition in the Aging Human Brain". Neuron 89: 971–982. [PubMed: 26938442]

[20]. Schwartz A, Yu P, Miller B, Shcherbinin S, Dickson J, Navitsky M, Joshi A, Devous M, Mintun M (2016). "Regional Profiles of the candidate tau PET ligand [18]F-AV-1451 recapitulate key features of Braak histopathological stages". Brain. 139; 1539–1550. [PubMed: 26936940]

[21]. Sertkaya A, Wong H, Jessup A, Beleche T (2016). "Key Cost Drivers of pharmaceutical clinical trials in the United States". Clinical Trials. 13(2): 117–126. [PubMed: 26908540]

[22]. Southekal S. Devous M. Navitsky M. Kennedy I. Joshi A. Mintun M. Correction for Acquisition Time Discrepancies in SUVR Analyses of ^{18}F-AV-1451 Tau Images. Poster presented at Human Amyloid Imaging Conference; Miami. USA. 2016. 2016

[23]. Southekal S Devous M, Navitsky M, Kennedy I, Navitsky M, Lu M, Joshi A, Mintun M (2017). "Flortaucipir F 18 Quantitation using a Parametric Estimate of Reference Signal Intensity (PERSI)". J. Nucl. Med; 117: [epub ahead of print].

[24]. Thie J (2004). Understanding the Standardized Uptake Value, Its Methods, and Implications for Usage. J. Nucl. Med; 45: 1431–1434. [PubMed: 15347707]

[25]. Thie. J, Hubner KF, Smith G (2000). The Diagnostic Utility of the Lognormal Behavior of PET Standardized Uptake Values in Tumors. J. Nucl. Med; 41: 1664–1672. [PubMed: 11037996]

[26]. Van den Hoff J, Lougovski A, Schramm G, Maus J, Oehme L, Petr J, Beuthien-Baumann B, Kotzerke J, Hofheinz F (2014). Correction of scan time dependence of standard uptake values in oncological PET. Eur. J. Nucl. Med. Mo. Imaging 4:18.

[27]. Villeneueve S, Rabinovici G, Cohn-Sheehy B, Madison C, Ayakta N, Ghosh P, La Joie R, Arthur-Bentil S, Vogel J, Marks S, Lehmann M, Rosen H, Olichey J, Boxer A, Miller B, Borys E, Jin L, Huang E, Grinberg L, DeCarli C, Seeley V, Jagust W. (2015). Existing Pittsburgh Compound-B positron emission tomography thresholds are too high: statistical and pathological evaluation. Brain 138: 2020–2033. [PubMed: 25953778]

[28]. Watson CC "New, faster, image-based scatter correction for 3D PET," IEEE Trans. Nucl. Sci, vol. 47, pp. 1587–1594, 2000.

[29]. Watson CC, Casey ME, Michel C, and Bendriem B, "Advances in scatter correction for 3D PET/CT," in IEEE Nucl. Sci. Conf. Rec, 2004, pp. 3008–3012

[30]. Weston P, Paterson R, Modat M, Burgos N, Cardoso M, Magdalinou N, Lehmann M, Dickson J, Barnes A, Bomanji J, Kayani I, Cash D, Ourselin S, Toombs J, Lunn. M,Mummery C, Warren J, Rossor M, Fox N, Zetterburg H, Schott J (2013). Using Florbetapir positron emission tomography to explore cerebrospinal fluid cut points and gray zones in small sample sizes. Alzheimers Dement. 1(4): 440–446.

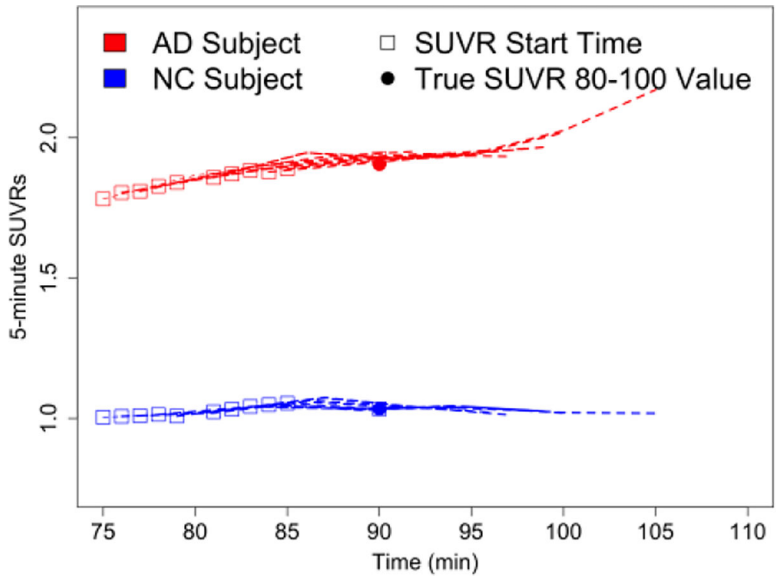


Fig. 1.

4 × 5 min offset SUVRs beginning with 75 min to 85 min scans and increasing start-times by 1 min for *right middle-temporal* region in a subject with Alzheimer's Disease (AD) and a normal control (NC) subject. The solid circle represents the *true* value and the squares represent the start-times of the scans. 5-minutes frames are increasing consistently for the AD subject but stable for the NC subject. Fig A.1 is another version of this figure that does not overlay the time-frames.

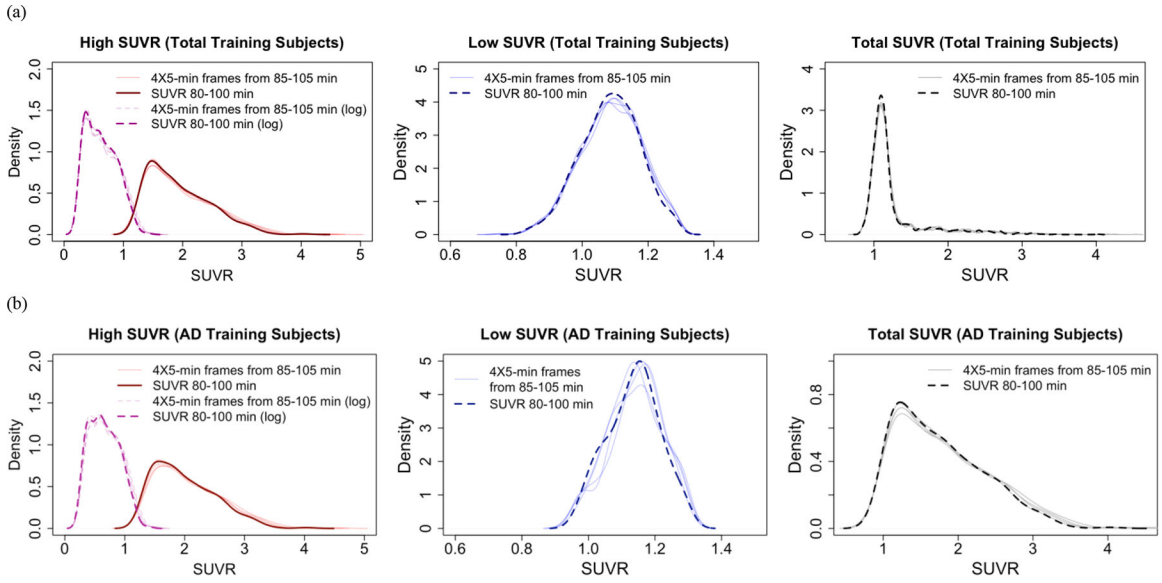


Fig. 2.

Distributions for SUVRs of cortical ROIs for total (top row) available subjects and for AD subjects (bottom). Density estimates for high-SUVR (>1.3) ROIs (left) are accompanied by log-transformed values, which are approximately normal. Low SUVR (≤1.3) ROIs (middle) are approximately normal. Kolmogorov-Smirnov tests were conducted on high and low values and further justifies the hypothesis that the low SUVR values and log-transformed high SUVR values are approximately Gaussian.

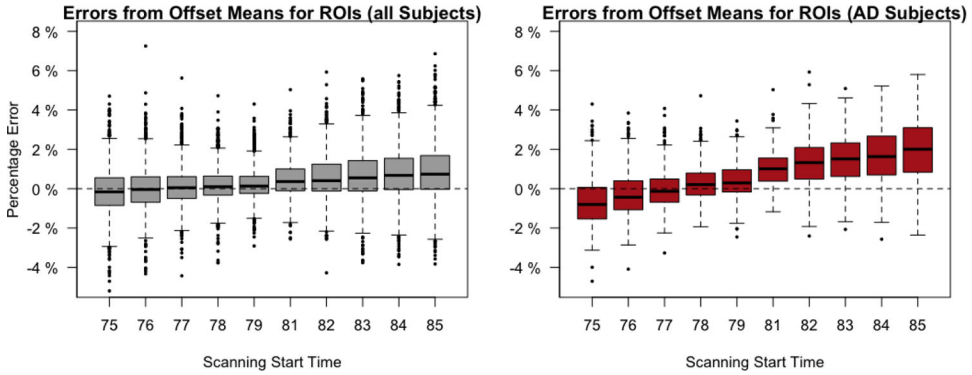


Fig 3:

Boxplots of percentage errors of uncorrected 20-min SUVRs from 75–85 min for total (68) ROIs in all (19) subjects in the test set and for ROIs in AD subjects in the test set. Errors are larger in AD subjects, reaching a mean of 2% in later scans. At a start-time of 85 min, the mean error reaches 1.99±1.49% in with a lower quartile of .84% and an upper quartile of 3.10%. SUVRs starting from time 80 represent the 'true' SUVR values. The errors depart somewhat from zero in all ROIs and noticeably from zero in AD-restricted ROIs. A table version of the plot may be found in the supplement as Table A.2.

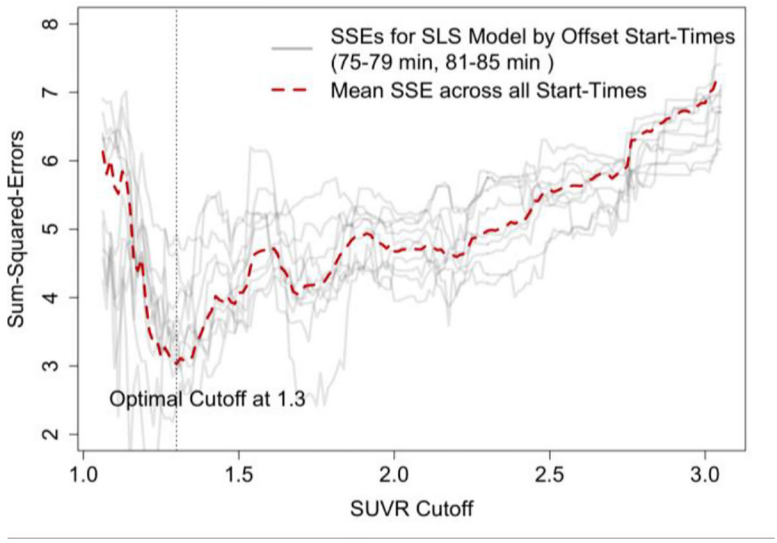
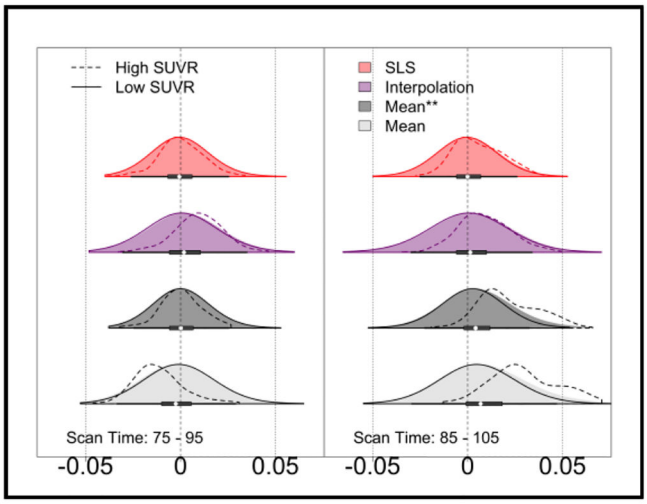


Fig. 4.

Residual sum-of-squares for the errors $e_{,h}$ and $e_{i,l}$ in (1). Errors are minimized when the cutoff point C is 1.3.

(a)



(b)

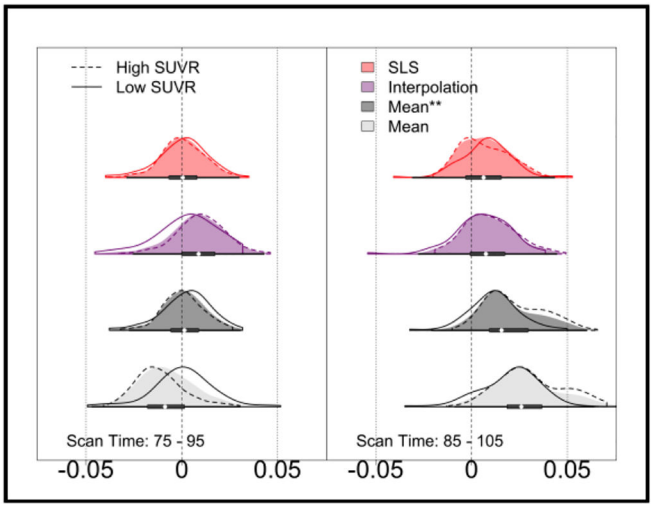


Fig. 5.

Density plots comparing errors between mean, mean**, SLS, and interpolation methods for total ROIs (a) and ROIs in AD subjects (b) for scanning times starting at 75 min and 85 min. The SLS method (red) consistently yields the smallest variance and least skew. The advantages of SLS are especially pronounced in later scans for AD subjects. A detailed version of this figure can be found in the supplement in Fig A.2.

Table 1:

SUBJECT CHARACTERISTICS

Subject characteristics for training and test sets. The groups were split into randomly generated groups until a similar distribution of diagnoses, sex, and age within diagnoses were assigned.

	Diagnosis	N	Age	Sex (M/F)
Training Set	Alzheimer's Disease (AD)	5	58 ± 5	¼
	Normal Control (NC)	7	77 ± 7	¾
	Mild Cognitive Impairment (MCI)	3	72 ± 7	2/1
	Frontal-temporal Dementia (FID)	2	57 ± 16	2/0
	Parkinson's Disease (PD)/ Corticobasal Syndrome (CBS)	4	69 ± 3	0/4
	Total	21	67 ± 12	10/11
Test Set	Alzheimer's Disease (AD)	4	68 ± 10	2/2
	Normal Control (NC)	7	78 ± 4	¾
	Mild Cognitive Impairment (MCI)	3	69 ± 3	3/0
	Frontal-temporal Dementia (FTD)	3	63 ± 17	2/1
	Parkinson's Disease (PD) / Corticobasal Syndrome (CBS)	2	73 ± 3	1/1
	Total	19	72 ± 9	11/8

Table 2:
OPTIMAL WEIGHTS FOR EACH OFFSET-TIME

Optimal coefficients for high (top) and low (bottom) SUVRs for four 5-min frames of offset SUVRs. Weights are applicable across sites and scanners

Log-Transformed Weights (High SUVRs>1.3)					
Time (min)	$A_{t,0}$	$A_{t,1}$	$A_{t,2}$	$A_{t,3}$	$A_{t,4}$
75	−.01	−.07	.31	.38	.37
76	−.01	−.03	.30	.36	.38
77	−.01	.03	.24	.40	.32
78	−.01	.07	.28	.35	.30
79	.00	.15	.25	.30	.29
81	.00	.23	.28	.28	.20
82	.00	.23	.39	.26	.11
83	.00	.29	.36	.26	.08
84	.00	.31	.34	.27	.07
85	.00	.35	.33	.27	.03
Weights (Low SUVRs ≤1.3)					
Time (min)	$B_{t,0}$	$B_{t,1}$	$B_{t,2}$	$B_{t,3}$	$B_{t,4}$
75	.01	.05	.35	.29	.31
76	.01	.12	.3	.29	.28
77	0	.17	.26	.29	.27
78	.01	.2	.27	.27	.25
79	.01	.24	.26	.25	.23
81	.02	.27	.26	.23	.22
82	.02	.28	.29	.23	.18
83	.02	.31	.27	.26	.13
84	.02	.32	.29	.26	.10
85	.02	.34	.28	.30	.06

Table 3:
PERCENTAGE ERRORS FOR ALTERNATE RECONSTRUCTION PARAMETERS

Comparison of mean percentage absolute error for four correction methods for differing reconstruction parameters: FBP (top), 0mm iterative smoothing (middle), and 6mm iterative smoothing (bottom) for test set subjects.

(a) Filtered Backprojection (FBP)				
Time (min)	**SLS**	**Mean**	**Mean****	**Interpolation**
75	0.83 ±1.1%	1.2 ±1.51%	0.88 ±1.43%	1.03 ±1.13%
76	0.77 ±1.00%	1.02 ±1.3%	0.78 ±1.24%	0.90 ±1.04%
77	0.69 ±0.90%	0.84 ±1.09%	0.73 ±1.03%	0.76 ±1.01%
78	0.61 ±0.79%	0.68 ±0.89%	0.68 ±0.82%	0.58 ±0.89%
79	0.49 ±0.60%	0.46 ±0.60%	0.46 ±0.58%	0.43 ±0.60%
81	0.46 ±0.58%	0.43 ±0.56%	0.43 ±0.57%	0.42 ±0.56%
82	0.55 ±0.75%	0.67 ±0.84%	0.67 ±0.82%	0.59 ±0.84%
83	0.62 ±0.80%	0.85 ±1.05%	0.67 ±0.97%	0.71 ±0.93%
84	0.69 ±0.91%	1.04 ±1.28%	0.74 ±1.19%	0.87 ±1.00%
85	0.70 ±0.95%	1.24 ±1.49%	0.81 ±1.32%	0.96 ±1.07%
(b) Iterative Smoothing (0 mm)				
Time (min)	**SLS**	**Mean**	**Mean****	**Interpolation**
75	1.01 ±1.38%	1.30 ±1.72%	1.03 ±1.80%	1.33 ±1.39%
76	0.87 ±1.15%	1.11 ±1.46%	0.89 ±1.44%	1.05 ±1.20%
77	0.80 ±1.07%	0.92 ±1.24%	0.88 ±1.26%	0.93 ±1.19%
78	0.69 ±0.92%	0.74 ±1.01%	0.74 ±1.01%	0.73 ±1.01%
79	0.58 ±0.77%	0.56 ±0.76%	0.56 ±0.77%	0.57 ±0.76%
81	0.58 ±0.75%	0.56 ±0.75%	0.56 ±0.75%	0.54 ±0.75%
82	0.69 ±0.91%	0.75 ±0.97%	0.75 ±0.98%	0.71 ±0.97%
83	0.77 ±0.99%	0.94 ±1.19%	0.83 ±1.14%	0.86 ±1.13%
84	0.85 ±1.10%	1.16 ±1.47%	0.89 ±1.39%	1.03 ±1.20%
85	0.84 ±1.12%	1.31 ±1.67%	0.93 ±1.56%	1.17 ±1.27%
(c) Iterative Smoothing (6 mm)				
Time (min)	**SLS**	**Mean**	**Mean****	**Interpolation**
75	0.65 ±0.80%	0.99 ±1.22%	0.68 ±1.03%	0.77 ±0.84%
76	0.64 ±0.78%	0.85 ±1.06%	0.62 ±0.93%	0.68 ±0.79%
77	0.56 ±0.68%	0.67 ±0.85%	0.55 ±0.77%	0.57 ±0.74%
78	0.50 ±0.59%	0.54 ±0.68%	0.54 ±0.61%	0.46 ±0.68%
79	0.42 ±0.47%	0.38 ±0.47%	0.38 ±0.46%	0.34 ±0.47%
81	0.40 ±0.50%	0.36 ±0.46%	0.36 ±0.47%	0.35 ±0.46%
82	0.48 ±0.62%	0.56 ±0.69%	0.56 ±0.64%	0.48 ±0.69%
83	0.53 ±0.67%	0.72 ±0.86%	0.55 ±0.74%	0.57 ±0.72%
84	0.59 ±0.76%	0.89 ±1.05%	0.60 ±0.91%	0.67 ±0.79%
85	0.58 ±0.75%	0.96 ±1.15%	0.63 ±1.01%	0.76 ±0.83%

Table 4:
REQUIRED SAMPLE SIZE FOR SIGNIFICANCE IN SIMULATED TAU TRIALS

Minimum required participants to achieve statistical power of .9 and significance of $p<.0001$ for tau clinical trials. Values are converted from an ensemble mean Cohen's d over 10,000 simulated clinical trials. SLS required fewer participants in all simulation schemes.

(a) Liberal Scan-time Errors for Total ROIs

Treatment	SLS	Mean	Mean**	Interpolation
−0.01	*1923*	*2276*	*2074*	*2118*
−0.015	*865*	*1013*	*963*	*879*
−0.02	*496*	*570*	*511*	*532*
−0.025	*321*	*362*	*340*	*331*
−0.03	*232*	*283*	*253*	*239*
−0.035	*175*	*208*	*187*	*181*
−0.04	*138*	*163*	*149*	*141*
−0.045	*112*	*130*	*120*	*115*
−0.05	*93*	*108*	*99*	*96*

(b) Liberal Scan-time Errors for ROIs in AD Subjects

Treatment	SLS	Mean	Mean**	Interpolation
−0.01	*1859*	*2358*	*2147*	*1964*
−0.015	*847*	*1104*	*972*	*956*
−0.02	*509*	*602*	*559*	*544*
−0.025	*337*	*396*	*353*	*347*
−0.03	*241*	*278*	*255*	*245*
−0.035	*178*	*209*	*195*	*183*
−0.04	*140*	*164*	*152*	*143*
−0.045	*114*	*132*	*122*	*117*
−0.05	*93*	*110*	*100*	*97*

(c) Conservative Scan-time Errors for Total ROIs

Treatment	SLS	Mean	Mean**	Interpolation
−0.01	*1908*	*2027*	*1988*	*2023*
−0.015	*789*	*875*	*895*	*841*
−0.02	*460*	*514*	*521*	*503*
−0.025	*300*	*339*	*324*	*308*
−0.03	*223*	*243*	*236*	*231*
−0.035	*165*	*182*	*177*	*171*
−0.04	*129*	*141*	*137*	*134*
−0.045	*106*	*115*	*111*	*108*
−0.05	*88*	*96*	*93*	*91*

(d) Conservative Scan-time Errors for ROIs in AD Subjects

Treatment	SLS	Mean	Mean**	Interpolation
−0.01	*1902*	*2163*	*1958*	*1920*
−0.015	*860*	*894*	*885*	*878*

(a) Liberal Scan-time Errors for Total ROIs				
Treatment	**SLS**	**Mean**	**Mean****	**Interpolation**
−0.02	*475*	*530*	*496*	*493*
−0.025	*309*	*338*	*323*	*319*
−0.03	*222*	*253*	*244*	*233*
−0.035	*167*	*186*	*180*	*171*
−0.04	*131*	*145*	*141*	*134*
−0.045	*106*	*117*	*114*	*108*
−0.05	*88*	*98*	*94*	*91*