
Measurement of Longitudinal β -Amyloid Change with ^{18}F -Florbetapir PET and Standardized Uptake Value Ratios

Susan M. Landau^{1,2}, Allison Fero², Suzanne L. Baker², Robert Koeppe³, Mark Mintun⁴, Kewei Chen⁵, Eric M. Reiman⁵, and William J. Jagust^{1,2}

¹Helen Wills Neuroscience Institute, University of California, Berkeley, California; ²Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; ³Radiology Department, University of Michigan, Ann Arbor, Michigan; ⁴Avid Radiopharmaceuticals, Inc., Philadelphia, Pennsylvania; and ⁵Banner Alzheimer's Institute, Phoenix, Arizona

The accurate measurement of β -amyloid ($\text{A}\beta$) change using amyloid PET imaging is important for Alzheimer disease research and clinical trials but poses several unique challenges. In particular, reference region measurement instability may lead to spurious changes in cortical regions of interest. To optimize our ability to measure ^{18}F -florbetapir longitudinal change, we evaluated several candidate regions of interest and their influence on cortical florbetapir change over a 2-y period in participants from the Alzheimer Disease Neuroimaging Initiative (ADNI). **Methods:** We examined the agreement in cortical florbetapir change detected using 6 candidate reference regions (cerebellar gray matter, whole cerebellum, brain stem/pons, eroded subcortical white matter [WM], and 2 additional combinations of these regions) in 520 ADNI subjects. We used concurrent cerebrospinal fluid $\text{A}\beta_{1-42}$ measurements to identify subgroups of ADNI subjects expected to remain stable over follow-up (stable $\text{A}\beta$ group; $n = 14$) and subjects expected to increase (increasing $\text{A}\beta$ group; $n = 91$). We then evaluated reference regions according to whether cortical change was minimal in the stable $\text{A}\beta$ group and cortical retention increased in the increasing $\text{A}\beta$ group. **Results:** There was poor agreement across reference regions in the amount of cortical change observed across all 520 ADNI subjects. Within the stable $\text{A}\beta$ group, however, cortical florbetapir change was 1%–2% across all reference regions, indicating high consistency. In the increasing $\text{A}\beta$ group, cortical increases were significant with all reference regions. Reference regions containing WM (as opposed to cerebellum or pons) enabled detection of cortical change that was more physiologically plausible and more likely to increase over time. **Conclusion:** Reference region selection has an important influence on the detection of florbetapir change. Compared with cerebellum or pons alone, reference regions that included subcortical WM resulted in change measurements that are more accurate. In addition, because use of WM-containing reference regions involves dividing out cortical signal contained in the reference region (via partial-volume effects), use of these WM-containing regions may result in more conservative estimates of actual change. Future analyses using different tracers, tracer-kinetic models, pipelines, and comparisons with other biomarkers will further optimize our ability to accurately measure $\text{A}\beta$ changes over time.

Key Words: amyloid; Alzheimer's disease; PET imaging

J Nucl Med 2015; 56:567–574

DOI: 10.2967/jnumed.114.148981

Emerging longitudinal amyloid PET studies suggest that there are several analytic challenges that make measurement of β -amyloid ($\text{A}\beta$) change different from cross-sectional $\text{A}\beta$ analyses. These challenges are further complicated by the fact that there is no gold standard for measurement of $\text{A}\beta$ change, because rates of change cannot be inferred from cross-sectional autopsy data. Because there is currently no consensus on how to address these methodologic challenges, existing longitudinal amyloid PET studies have used methods that were developed for cross-sectional datasets. However, recent data suggest that the amount of annual change in ^{11}C -Pittsburgh compound B (^{11}C -PiB) that we can detect in research studies (1) and clinical trials (2) is similar to test–retest error rates, which are approximately 5%–9% for ^{11}C -PiB standardized uptake value ratios (SUVRs) (3,4). The precise measurement of $\text{A}\beta$ change is therefore a critical concern for Alzheimer disease (AD) research that seeks to identify subtle or early changes in $\text{A}\beta$ and for clinical trials aimed at separating meaningful change from noise (2), particularly for individuals who are early in the course of disease. Early identification of change is an important problem because interventions that lower brain $\text{A}\beta$ deposition will be tested in early stages of disease, and robust outcome measures are required to assess efficacy.

Several factors contribute to variability of longitudinal amyloid PET measurements, including meaningful physiologic changes due to disease, the tracer used, consistency of acquisition conditions (e.g., patient positioning within the field of view, medication use), and tracer quantitation method (distribution volume ratio [DVR] vs. SUVR). A factor of particular concern is selection of an appropriate reference region that is expected to remain free of fibrillar amyloid over time. Whole cerebellum in particular has been used frequently for florbetapir PET normalization (5–7), due to the relative sparing of neuritic plaques in the cerebellum and an existing cerebellum-based positivity threshold derived from young normal subjects (8) and validated in an autopsy study (9). However, cerebellar or pons reference regions may introduce noise in longitudinal measurements. For example, the small size of the pons or the low position of the cerebellum in the field of view may introduce artifacts due to image truncation, alterations in scatter, attenuation correction, and counting rate and sensitivity (10,11).

Received Oct. 20, 2014; revision accepted Feb. 6, 2015.
For correspondence or reprints contact: Susan M. Landau, 118 Barker Hall MC #3190, University of California Berkeley, Berkeley, CA 94720-3190.
E-mail: slandau@berkeley.edu
Published online Mar. 5, 2015.
COPYRIGHT © 2015 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

A final concern is nonspecific tracer retention in white matter (WM), which has been observed for all amyloid PET imaging agents to varying degrees (12–15). Measurements of tracer retention in the cortex are contaminated by WM retention, so lack of stability in WM retention over time could lead to detection of spurious changes in the cortex. Inclusion of WM (such as centrum semiovale) in the reference region could perhaps weaken cortical signal of interest, because sub-cortical WM is itself affected by the combined effects of tracer retention in neighboring cortical voxels and partial-volume averaging.

However, recent work from our group and others using ^{11}C -PiB (16) and ^{18}F -florbetapir PET data from the Alzheimer Disease Neuroimaging Initiative (ADNI) has suggested that a WM reference region reduces variability in longitudinal cortical measurements (11,17,18) and also improves the power to detect A β -modifying treatment effects and cognitive change (10). A key challenge in this work is the lack of a gold standard for evaluating the accuracy of different reference regions. To address this challenge, here we used cerebrospinal fluid (CSF) A β_{1-42} measurements and diagnostic status to identify ADNI subjects whose longitudinal cortical florbetapir (SUVR) was unlikely to change (stable A β group) and those whose SUVRs were likely to increase (increasing A β group) over a 2-y follow-up period. We then evaluated several candidate reference regions, defined using each subject's native space structural image, based on the extent to which these groups showed the anticipated patterns of change.

MATERIALS AND METHODS

Study Design and Rationale

We calculated annual change in florbetapir cortical retention based on 2 scanning sessions at a 2-y interval using several reference regions with varying amounts of WM. To determine which reference regions were optimal for detecting longitudinal cortical change, we compared reference regions based on our assumptions that cortical A β change should be minimal in subjects expected to remain stable (stable A β group) and cortical A β should increase in a physiologically plausible way (or should not decrease) in subjects expected to increase (increasing A β group).

To define A β status, we used CSF A β_{1-42} measurements that were acquired concurrently with florbetapir scans at baseline and 2-y follow-up visits and analyzed by the ADNI Biomarker core laboratory with a CSF A β_{1-42} autopsy-validated positivity cutoff of 192 pg/mL, as described previously (19).

Participants

ADNI is a multisite longitudinal biomarker study that has enrolled more than 1,500 cognitively normal older individuals, people with early mild cognitive impairment (EMCI) or late MCI, and people with early AD (www.adni-info.org provides additional information). Our

complete dataset was made up of 520 ADNI participants who underwent a baseline structural MR imaging scan and baseline and follow-up florbetapir scans at a 2-y interval using a protocol described elsewhere (<http://adni-info.org>). All participants gave written informed consent that was approved by the Institutional Review Board of each participating institution.

Stable and Increasing A β Groups

The stable A β group ($n = 14$) was used as a proxy for a test–retest sample. It was defined as all subjects who were apolipoprotein E4 (ApoE4) noncarriers cognitively normal at baseline and follow-up, had lumbar punctures that were concurrent with (within 3 mo of) both baseline and follow-up florbetapir scans, and were CSF A β -negative at both of these visits. We further restricted the group to subjects who were not decreasing on CSF A β to reduce the possibility that these CSF A β -negative subjects had subthreshold but increasing cortical A β . Because a small CSF A β decrease may reflect measurement error rather than a meaningful decrease, we estimated CSF A β test–retest variability by calculating the amount of CSF A β increase in cognitively normal subjects, a change that likely reflects noise. We used this value of 3.4 pg/mL as an approximation of CSF A β test–retest variability and eliminated from our stable A β group any subject who decreased by more than 3.4 pg/mL from baseline to follow-up. This value is in line with recent test–retest comparisons performed by the ADNI biomarker core (20).

The increasing A β group (normal, $n = 37$; EMCI, $n = 54$; total $n = 91$) was defined as subjects who were diagnosed as cognitively normal or EMCI at baseline, had lumbar punctures that were concurrent with (within 3 mo of) the baseline florbetapir scan, and were CSF A β_{1-42} -positive at this baseline visit.

Characteristics of the stable and increasing A β groups are summarized in Table 1.

Florbetapir PET Image Processing

Florbetapir synthesis and image acquisition details are described in detail elsewhere ((7), <http://adni-info.org>). Briefly, florbetapir images consisted of 4×5 min frames acquired at 50–70 min after injection, which were realigned, averaged, resliced to a common voxel size (1.5 mm^3), and smoothed to a common resolution of 8 mm^3 in full width at half maximum.

Structural T1 images acquired concurrently with the baseline florbetapir images were used as a structural template to define cortical regions of interest and reference regions in native space for each subject using Freesurfer (version 4.5.0; surfer.nmr.mgh.harvard.edu) as described previously (7,21,22).

Baseline and follow-up florbetapir scans for each subject were coregistered to baseline structural MR scans, which were subsequently used to extract weighted cortical retention means (SUVRs) from gray matter within 4 large cortical regions of interest (frontal, cingulate, parietal, and temporal cortices) that were averaged to create an

TABLE 1
Stable and Increasing A β Group Characteristics

Characteristic	<i>n</i>	Age (y)	Education (y)	Female (%)	ApoE4 carriers (%)	Florbetapir scan interval (y)
Stable Aβ group						
Normal	14	75.2 (4.4)	17.0 (3.2)	36%	NA (0%)	2.0 (0.0)
Increasing Aβ group						
Normal	37	75.3 (6.3)	16.7 (2.3)	62%	41%	2.0 (0.7)
EMCI	54	73.1 (6.4)	15.7 (2.7)	41%	63%	2.0 (0.2)
Total	91	74.0 (6.4)	16.1 (2.6)	49%	54%	2.0 (0.1)

Data in parentheses are SDs.
NA = not applicable.

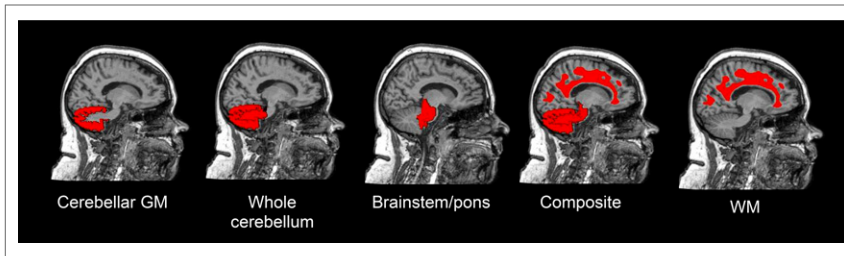


FIGURE 1. Candidate reference regions for florbetapir analysis are overlaid on an example subject's MR imaging scan.

SUVR as described in greater detail in recent studies (7,21) and online (23).

Candidate Reference Regions

Cortical SUVRs for each subject at each time point were generated by dividing the cortical summary average (described above) by each of the following candidate reference regions to create 5 different SUVRs (listed in order of increasing WM making up the reference region) (Fig. 1): Freesurfer-defined cerebellar gray matter, whole cerebellum, pons/brain stem, and eroded subcortical WM, which was created by smoothing a binarized Freesurfer-defined subcortical WM image to the 8 mm³ resolution of the florbetapir image and then thresholding this image at 0.70 to erode WM-defining voxels away from gray matter. The fifth region (composite reference region) was made by averaging means of the whole cerebellum, pons/brain stem, and eroded subcortical WM regions (17).

A sixth normalization strategy (WM ratio) used a 2-step normalization process in which, first, the cortical summary average (Ctx) is normalized to the whole cerebellum (WhCereb) to bring the SUVR units in line with cross-sectional florbetapir measures that are often scaled to whole cerebellum (and corresponding positivity threshold).

$Ctx_{V1} = Ctx_{V1}/WhCereb_{V1}$, $Ctx_{V2} = Ctx_{V2}/WhCereb_{V2}$, $WM_{V1} = WM_{V1}/WhCereb_{V1}$, and $WM_{V2} = WM_{V2}/WhCereb_{V2}$. V1 and V2 are visits 1 and 2, respectively.

$$SUVR\ V1 = \frac{Ctx_{V1}}{\frac{WM_{V1}}{WM_{V1}}}, \quad \text{Eq. 1}$$

$$SUVR\ V2 = \frac{Ctx_{V2}}{\frac{WM_{V2}}{WM_{V1}}}, \quad \text{Eq. 2}$$

$$\Delta SUVR = \frac{Ctx_{V2}}{\frac{WM_{V2}}{WM_{V1}}} - \frac{Ctx_{V1}}{\frac{WM_{V1}}{WM_{V1}}}, \quad \text{Eq. 3}$$

In Equation 1, the denominator cancels, so the initial step of normalizing to whole cerebellum is critical in its influence on the resulting value. A critical feature of this method is that the percentage change measurements are identical to those obtained with normalization using the WM region alone. This is because the second visit (Eq. 2) is scaled by a term representing change in WM (WM_{V2}/WM_{V1}), resulting in change measurements (Eq. 3) that are weighted by change in WM and therefore identical for the WM ratio method and the WM reference region alone.

Mean annual percentage SUVR change was calculated for each reference region, and reference regions were compared according to the following criteria: in the stable A β group SUVR change should be close to zero and for the increasing A β group SUVR change should be physiologically plausible (e.g., an annual increase of less than 20%), and the number of subjects with declining SUVRs should be minimal.

Statistical Methods

For the stable A β group, 1-sample *t* tests were used to compare mean annual percentage SUVR change with zero for each reference region, and $\alpha = 0.05$ was Bonferroni-corrected to adjust for the 6 comparisons, resulting in $\alpha = 0.008$.

For the increasing A β group, visit 1 to visit 2 increases were tested with paired *t* tests, and correlations were examined with linear regression. Minimum, maximum, and mean annual percentage SUVR change/SD and coefficient of variation (CV) were calculated for each reference region. The examination of SUVR percentage change values (as opposed to difference measurements) enabled the evaluation of different reference regions despite differing scales.

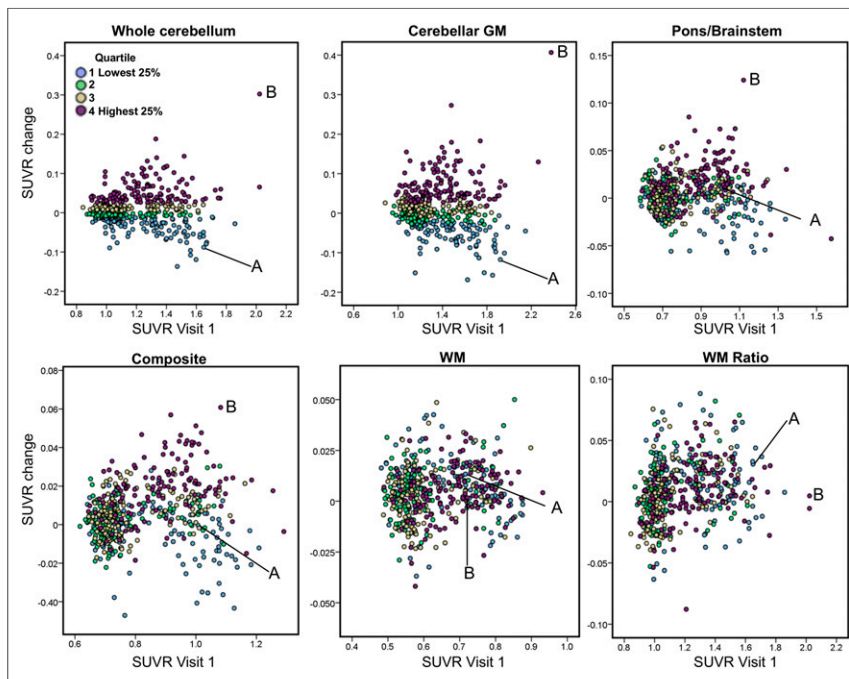


FIGURE 2. SUVR change is plotted against baseline SUVR using 6 candidate reference regions for all 520 subjects. Subjects were divided into quartiles based on SUVR change using whole cerebellum reference region (upper left; blue = lowest 25%, purple = highest 25%), and this quartile assignment was used for the other 5 reference regions. Subjects A and B show SUVR pattern that varies across reference regions (see "Results" section).

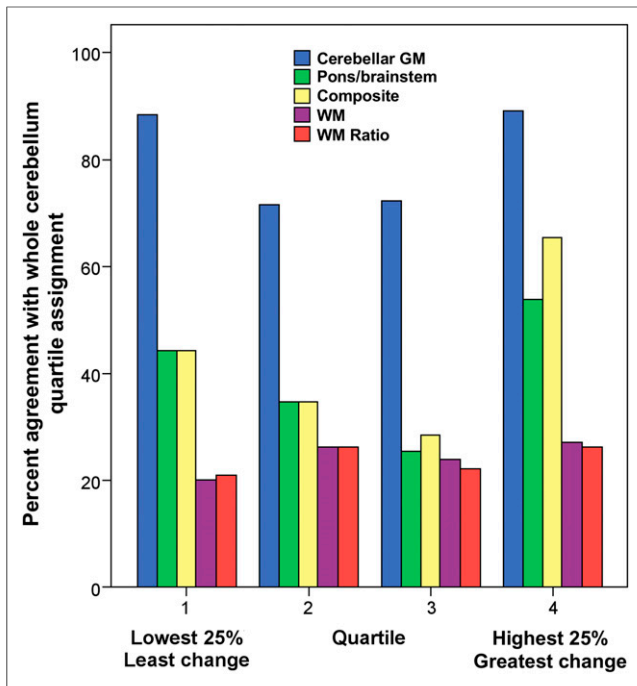


FIGURE 3. For each quartile defined by SUVR change using whole cerebellum reference (x-axis), percentage of subjects is shown who are assigned to same quartile defined by SUVR change using other reference regions.

RESULTS

Influence of Reference Region on Longitudinal Cortical SUVR Change

To determine the consistency in annual SUVR change that can be detected using different reference regions, we divided all 520 subjects into quartiles of SUVR change based on the whole cerebellum reference region (Fig. 2, upper left). This quartile assignment was used to define the color scale for the other candidate reference region plots, such that for all plots in Figure 2, blue represents subjects with the lowest 25% florbetapir change based on whole cerebellum normalization, green and yellow represent intermediate degrees of change, and purple represents subjects with the highest 25% change based on whole cerebellum normalization. The lack of preservation of quartile assignment is evident in the disordered positions of the blue, green, yellow, and purple points as the reference region changes. For a more quantitative examination of this agreement, we determined the likelihood that a subject assigned to a particular quartile based on whole cerebellum normalization was assigned to the same quartile using the other normalization regions (Fig. 3). Agreement was high between whole cerebellum quartile assignment and cerebellar GM quartile assignment (blue) because of the proximity of these regions to one another. However, other reference regions had agreement that was near chance (25%), particularly for the middle 2 quartiles, which are clustered around zero and contain a narrow range of values, increasing the likelihood that quartile assignment would not agree across reference regions.

We also examined more extreme cases of disagreement between reference regions and found that, for example, approximately 22% of subjects who were in the highest quartile (greatest change) using whole cerebellum were in the lowest quartile (least change) when using WM. Conversely, one third of subjects (34%) who were in the lowest quartile when using whole cerebellum were in

the highest quartile when using WM. An example subject (subject A) who demonstrated this pattern is labeled in Figure 2, in addition to the corresponding florbetapir images shown in raw form and normalized using several example reference regions (Fig. 4A). A second subject (subject B), who was a cognitively normal member of the increasing A β group, had SUVRs at both time points that were highly dependent on reference region (Fig. 2, example images shown in Fig. 4B). This subject was an outlier (highly positive at both visits) when using cerebellar GM or whole cerebellum due to unexplained extremely low tracer uptake in the cerebellum but was in the middle of the distribution when using a WM-containing reference region (Fig. 5). Examination of ADNI florbetapir scanning and reconstruction records do not indicate that any special concerns were noted for either scans for these subjects.

Change in Stable A β Group

Annual SUVR change and absolute SUVR change across reference regions for the stable A β group are summarized in Table 2. Annual mean percentage change ranged from $-1.2 \pm 1.7\%$ (WM and WM ratio) to $0.8 \pm 1.6\%$ (cerebellar GM). Annual absolute percentage change ranged from $1.0 \pm 0.9\%$ (composite) to $1.6 \pm 1.3\%$ (WM and WM ratio). No reference region resulted in an annual SUVR mean percentage change that was different from zero based on our statistical criteria, but the absolute change was different from zero for all reference regions, indicating significant test–retest error across all regions (1-sample *t* tests; $P \leq 0.001$).

Comparison of WM and WM Ratio Methods

The WM and WM ratio regions both resulted in identical percentage SUVR change measurements (see the “Candidate Reference Regions” section), so mean/SD percentage SUVR change is identical when using these 2 reference regions (Tables 2 and 3). However, the SUVR units for these 2 methods differ in that the WM reference region results in SUVRs ranging from approximately 0.5 to 1.0 (due to the relatively high florbetapir retention in WM relative to cortex) whereas the WM ratio results in SUVRs at each time point that are on the same scale as those obtained with whole cerebellum normalization, ranging from approximately 0.8 to 2.0.

Change in Increasing A β Group

Annual SUVR change and absolute change across reference regions for the increasing A β group are summarized in Table 3. Annual SUVRs increased across all reference regions ($P < 0.05$), although not all regions survived our Bonferroni-corrected threshold ($\alpha = 0.008$). Although the mean/SD annual percentage change is comparable across reference regions, the minimum and maximum values differ such that the composite reference region results in the narrowest range of change values (-3.4% to 5.6%) and cerebellar GM results in the broadest range of SUVR change (-10.4% to 18.5%). Subcortical WM-containing regions (composite, WM, WM ratio) resulted in low maximum annual increases of 6%–8%, compared with brain stem/pons (11%), whole cerebellum (15%), and cerebellar GM (19%). The CV is lowest for the composite region (1.18) and highest for cerebellar GM (3.43).

Plots of visit 1 versus visit 2 SUVRs of the increasing A β group for each reference region (Fig. 5) show that as the amount of WM making up the reference region increases, the proportion of decliners decreases. Specifically, 25% of the group decreases from visit 1 to 2 when using the cerebellar GM reference region, whereas 11% of the group decreases when using WM, WM ratio, and composite reference regions.

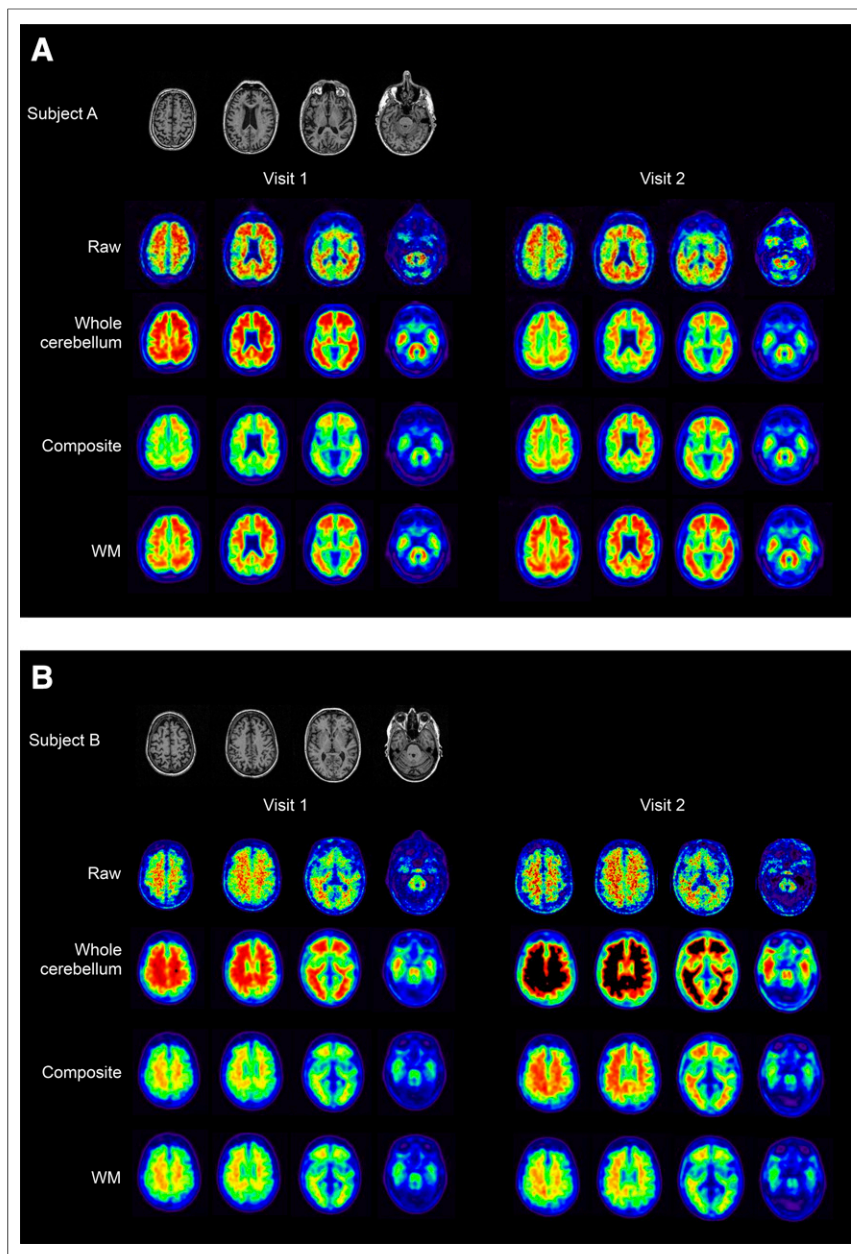


FIGURE 4. Sample baseline MR imaging and florbetapir scans at each visit are shown for 2 example subjects (subjects A and B) who had a reference region–dependent pattern of cortical retention (see “Results” section; Figs. 2 and 5). The top row (raw) of PET scans is averaged, coregistered frames that have not been intensity normalized. Subsequent rows show coregistered, averaged frames at uniform voxel size and resolution, normalized using reference region shown. Scale for color display selected for whole cerebellum was arbitrary but consistent across visits. Color scale for other reference regions was selected by transforming upper threshold for whole cerebellum to units consistent with other regions. This was performed using linear regression equations resulting from correlation between whole cerebellum–normalized SUVRs and other reference regions at baseline across all 520 subjects.

DISCUSSION

The cerebellum is frequently used as a reference region in florbetapir PET studies, primarily because of a validated whole cerebellum–based positivity threshold (8,9), but it may not be adequate for longitudinal analyses (10,11,17). Here, we examined the influence of 6 candidate reference regions on annual florbetapir cortical change in a large multicenter sample over a 2-y follow-

up period. We addressed the absence of a gold standard for assessing longitudinal A β measurement accuracy using concurrent CSF A β measurements and diagnostic information from a large ADNI sample to define groups expected to remain stable and to increase on florbetapir. Our study had several main findings. First, agreement in the amount of cortical change detected using different reference regions was poor, indicating that reference region selection has a substantial effect on which subjects appear to increase and which appear to remain stable (or decrease). Second, in the stable A β group, reference regions performed similarly as all showed minimal SUVR change. Third, in the increasing A β group, use of a reference region containing subcortical WM (as opposed to cerebellum or pons alone) resulted in SUVR change that was more physiologically plausible and also consistent with the idea that A β is not likely to decrease over time in individuals who are A β -positive at baseline and at early disease stages (normal, EMCI). Finally, we introduced a new reference region approach (WM ratio) with SUVR units that are consistent with whole cerebellum normalization, permitting application of a whole cerebellum–based positivity threshold but with percentage change that is equivalent to that obtained with WM normalization. Together, these findings suggest that selection of an appropriate reference region plays a critical role in longitudinal analysis, particularly when examining change in A β -positive individuals, and that use of a reference region containing subcortical WM results in SUVR change measurements that appear to be the most accurate.

On the basis of recent work showing lower rates of amyloid PET change in ApoE4– normal subjects (24) and agreement between CSF A β_{1-42} and florbetapir positive/negative status of 83%–91% ($\kappa = 72$) in a subset of these subjects (25), we expected that the stable A β group would show minimal cortical A β change and could therefore serve as a test–retest proxy for optimizing reference region selection. Our findings across reference regions showed that absolute percentage annual SUVR change was approximately 1%–2%

across all reference regions, suggesting that test–retest variability was adequate and comparable across all regions. Further examination revealed that SUVRs decreased when using regions including subcortical WM, and greater decreases roughly corresponded to greater amounts of WM in the reference region. Assuming that cortical change is indeed minimal in this group, this finding suggests that WM tracer retention may increase

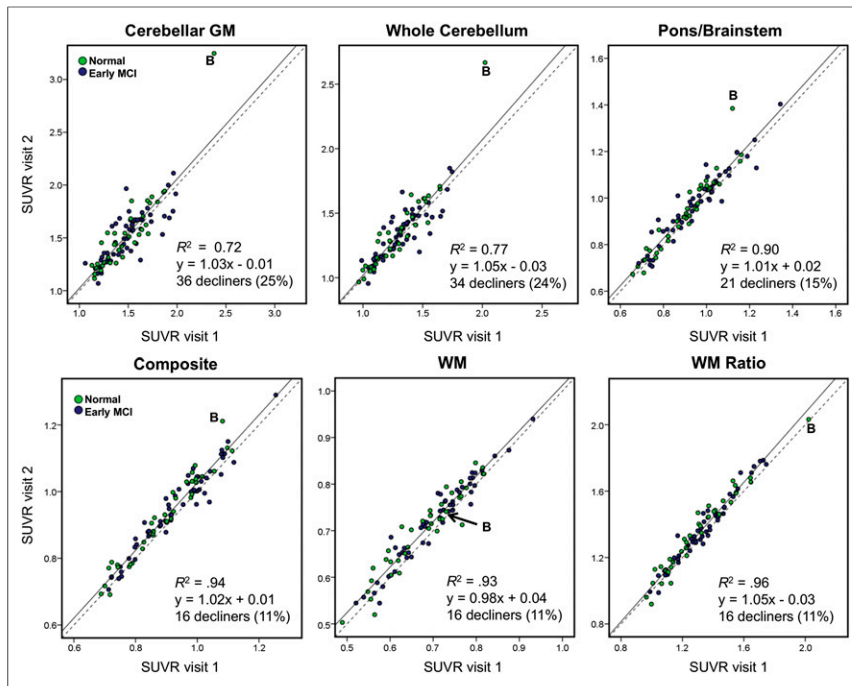


FIGURE 5. Relationship is shown between SUVRs at visit 1 vs. visit 2 in $A\beta$ -increasing group calculated using 6 different reference regions. R^2 values, linear regression equations, and number (and percentage) of decliners are shown for plots, which are arranged in order of increasing WM making up reference region. Example subject (subject B) who showed a discrepant pattern across reference regions is shown (Fig. 4B).

slightly over time, so use of a reference region that includes WM may result in a more conservative estimate of actual cortical increase.

We selected subjects likely to increase based on evidence that cognitively normal subjects who are ^{11}C -PiB-positive at baseline were likely to show ^{11}C -PiB increases over time (26). The increasing $A\beta$ group included cognitively normal and EMCI subjects but not subjects with greater cognitive impairment (e.g., late MCI and AD) because $A\beta$ may decrease after reaching a peak of accumulation in some patients (24), perhaps reflecting atrophy at later disease stages or a maximum saturation of $A\beta$ pathology in the cortex. We found that use of a reference region that includes subcortical WM resulted in SUVR change measurements that were less likely to decrease, compared with the cerebellum or pons alone. Use of subcortical WM-containing reference regions also resulted in maximum annual increases of 6%–8%, whereas pons and cerebellar regions resulted in increases greater than 10%.

Although the upper limit of annual amyloid accumulation is unknown, current models of the 20- to 40-y $A\beta$ accumulation time course (1,24) suggest that increases greater than 10%/y are unlikely. We note that these models are derived from longitudinal ^{11}C -PiB data (using a cerebellar reference), which also may be vulnerable to some of the reference region-based variability presented here (16). Finally, although the association between visit 1 and visit 2 was higher for reference regions containing WM ($R^2 = 0.93$ – 0.69) than for cerebellar regions ($R^2 = 0.72$ – 0.77 ; Fig. 5), the relevance of this is unclear because variability in amyloid accumulation rate is unknown.

Overall, the composite reference region (an average of the means of the whole cerebellum, brain stem/pons, and eroded WM regions) resulted in annual SUVR change with the lowest CV for the increasing $A\beta$ group. The composite region also resulted in the lowest annual mean absolute percentage SUVR difference for the stable $A\beta$ group (though all regions were comparable). The relatively large number of voxels may account for some of the SUVR stability when using the composite region. For example, inclusion of

WM could reduce the influence of low or high cerebellar retention due to changes in subject placement in the scanner or poor definition of the brain stem/pons due to differences in movement across scans. A similar argument can be made for the WM reference region alone, because it contains a larger number of voxels than cerebellum or pons and because the voxels are in the same axial plane as the cortical target regions, perhaps making it less vulnerable to scatter- or attenuation-correction errors that may occur at one or both visits. The large size of regions that include WM is also a potential drawback in that WM shares signal with cortical regions of interest, despite substantial erosion of WM away from the cortex. This means that when the cortical mean is divided by the WM-containing reference region, some of the cortical signal is lost because some cortical signal is common to both regions. This slight loss of signal provides further support for our observation that reference regions that include WM result in a conservative estimate of change.

TABLE 2
Comparison of Cortical Change Across All Reference Regions in Stable $A\beta$ Group

Reference region	Annual mean % change	SD	Annual mean absolute value % change	SD
Cerebellar gray matter	0.8%	1.6%	1.4%	1.1%
Whole cerebellum	0.2%	1.4%	1.2%	0.8%
Pons	−0.5%	1.9%	1.6%	1.0%
Composite	−0.6%	1.2%	1.0%	0.9%
WM	−1.2%	1.7%	1.6%	1.3%
WM ratio	−1.2%	1.7%	1.6%	1.3%

TABLE 3
Increasing A β Group Change

Reference region	Normal (n = 37)					EMCI (n = 54)					Total (n = 91)					t (increase)	p (increase)
	Min	Max	Annual mean % change	SD	CV	Min	Max	Annual mean % change	SD	CV	Min	Max	Annual mean % change	SD	CV		
Cerebellar GM	-4.8%	17.1%	1.8%	4.5%	2.55	-10.4%	18.5%	1.1%	4.9%	4.41	-10.4%	18.5%	1.4%	4.7%	3.43	2.4	0.017
WC	-3.7%	15.0%	1.8%	3.7%	2.06	-9.3%	14.1%	1.1%	4.0%	3.85	-9.3%	15.0%	1.4%	3.9%	2.89	2.9	0.004
Pons	-2.2%	11.1%	1.7%	2.4%	1.40	-4.4%	8.2%	1.4%	2.5%	1.81	-4.4%	11.1%	1.5%	2.5%	1.62	5.7	<0.001
Composite	-1.8%	5.6%	1.7%	1.7%	0.99	-3.4%	4.5%	1.2%	1.7%	1.36	-3.4%	5.6%	1.4%	1.7%	1.18	7.8	<0.001
WM	-4.0%	6.8%	1.7%	2.2%	1.25	-2.5%	6.8%	1.3%	1.6%	1.28	-4.0%	6.8%	1.5%	1.9%	1.28	7.7	<0.001
WM ratio	-4.0%	6.8%	1.7%	2.2%	1.25	-2.5%	6.8%	1.3%	1.6%	1.28	-4.0%	6.8%	1.5%	1.9%	1.28	7.9	<0.001

Min = minimum; Max = maximum; t (increase) = t value representing visit 1 to visit 2 increase; p (increase) = P value representing significance of visit 1 to visit 2 increase; WC = whole cerebellum.

An unexpected finding was the amount of disagreement between reference regions; for example, a third of subjects who decreased the most (lowest quartile) using whole cerebellum normalization increased the most (highest quartile) using WM normalization, suggesting that unusual retention patterns are not uncommon (both cross-sectionally and longitudinally) and that use of multiple reference regions and visual examination of raw and normalized scans may be helpful in these cases. These cases also raise the possibility that similar issues (e.g., subject positioning within the field of view) may influence cross-sectional analyses as well. Although the whole cerebellum standard has been compared with other potential reference regions in cross-sectional analyses (7,27), and critically, in autopsy-validated data (9), this issue warrants further study.

This study has several limitations. First, as mentioned previously, the lack of a gold standard for evaluating A β change is a key concern but is unlikely to be resolved with current biomarker tools. As an alternative, we used CSF A β to help predict florbetapir change. Although agreement between CSF A β and florbetapir status is high, some amount of disagreement between the 2 (25) may have introduced some error into our group selection, but this was unlikely to bias our results in favor of a particular reference region. Second, we currently have only 2 scans available to estimate florbetapir change. The use of 3 scans would improve our ability to estimate rates of change more accurately but may be less clinically relevant than the current approach due to scan cost. Third, our findings are specific to florbetapir SUVR measurements and may not generalize to other methods of PET analysis (i.e., DVR measurements) or other PET tracers. There are other contributions to variability in change measurements (e.g., image acquisition and image analysis methods) that are beyond the scope of this study. Additionally, percentage change SUVR measurements (as opposed to SUVR differences) may deviate from statistical normality and therefore warrant the use of nonparametric statistical tests because they are a ratio of 2 numbers that are themselves ratios. Although the key findings of this study were the comparison of reference regions to one another and not the statistical significance of SUVR changes, our overall results were generally unchanged by the use of nonparametric statistics. Fourth, our image coregistration was performed using the 50- to 70-min SUVR image, which is dominated by regions with higher

tracer retention and could therefore result in coregistration errors. Although dynamic data are not available in ADNI, use of early frame image data may result in more accurate MR imaging coregistration. A final related issue is that our target and reference regions were anatomically defined and based on each subject's baseline structural MR imaging scan, introducing the possibility that cortical or WM atrophy may have occurred, which could result in either over- or underestimations of change, depending on whether atrophy occurred in the target or reference region (or both). Partial-volume correction of this data, although an important area of future investigation, would require structural MR images at each time point (which were not currently available for all subjects) and may increase noise, thus complicating interpretation of the already complex relationship between reference regions.

CONCLUSION

We have evaluated longitudinal florbetapir cortical change calculated using several candidate reference regions. Our findings suggest that reference region selection has a substantial influence on the change that can be detected. Although the whole cerebellum has been used frequently for cross-sectional analyses of florbetapir data, this study indicates that a reference region containing subcortical WM results in more longitudinal measurements that contain less noise than those obtained with the cerebellar or pons regions.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. Data collection and sharing for this project was funded by the ADNI (National Institutes of Health grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan

Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Susan M. Landau has previously consulted for Genentech, Avid Radiopharmaceuticals, Inc., Janssen Alzheimer Immunotherapy, and Biogen Idec. Allison Fero, Suzanne L. Baker, and Kewei Chen have nothing to declare. Eric M. Reiman has been a paid advisor to AstraZeneca, Cerespir, Eisai, Eli Lilly, Novartis, Sanofi, and Takeda. He and his colleague have received research support from Avid Radiopharmaceuticals, Genentech, and Novartis. Mark Mintun is an employee of Avid Radiopharmaceuticals, Inc. William J. Jagust collaborates with Avid Radiopharmaceuticals, Inc., through participation in the ADNI. He is currently a consultant to Genentech/Banner Alzheimer Institute and Synarc.

No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

Data used in preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgment_List.pdf.

REFERENCES

- Jack CR Jr, Wiste HJ, Lesnick TG, et al. Brain beta-amyloid load approaches a plateau. *Neurology*. 2013;80:890–896.
- Rinne JO, Brooks DJ, Rossor MN, et al. ¹¹C-PiB PET assessment of change in fibrillar amyloid-beta load in patients with Alzheimer's disease treated with bapineuzumab: a phase 2, double-blind, placebo-controlled, ascending-dose study. *Lancet Neurol*. 2010;9:363–372.
- Lopresti BJ, Klunk WE, Mathis CA, et al. Simplified quantification of Pittsburgh compound B amyloid imaging PET studies: a comparative analysis. *J Nucl Med*. 2005;46:1959–1972.
- Tolboom N, Yaquub M, Boellaard R, et al. Test-retest variability of quantitative [¹¹C]PiB studies in Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2009;36:1629–1638.
- Doraiswamy PM, Sperling RA, Coleman RE, et al. Amyloid-beta assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study. *Neurology*. 2012;79:1636–1644.
- Fleisher AS, Chen K, Liu X, et al. Apolipoprotein E epsilon4 and age effects on florbetapir positron emission tomography in healthy aging and Alzheimer disease. *Neurobiol Aging*. 2013;34:1–12.
- Landau SM, Breault C, Joshi AD, et al. Amyloid-beta imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. *J Nucl Med*. 2013;54:70–77.
- Joshi AD, Pontecorvo MJ, Clark CM, et al. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. *J Nucl Med*. 2012;53:378–384.
- Clark CM, Schneider JA, Bedell BJ, et al. Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA*. 2011;305:275–283.
- Chen K, Rontiva A, Thiyyagura P, et al. Improved power to characterize longitudinal amyloid-β PET changes and evaluate amyloid-modifying treatments using a cerebral white matter reference region. *J Nucl Med*. 2015;56:560–566.
- Matthews D, Marendic B, Andrews R, et al. Longitudinal amyloid measurement for clinical trials: a new approach to overcome variability. Paper presented at: Human Amyloid Imaging; January 15–17, 2014; Miami, Florida.
- Vandenberghe R, Van Laere K, Ivanou A, et al. ¹⁸F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. *Ann Neurol*. 2010;68:319–329.
- Villemagne VL, Mulligan RS, Pejoska S, et al. Comparison of ¹¹C-PiB and ¹⁸F-florbetaben for Aβ imaging in ageing and Alzheimer's disease. *Eur J Nucl Med Mol Imaging*. 2012;39:983–989.
- Wong DF, Rosenberg PB, Zhou Y, et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand ¹⁸F-AV-45 (florbetapir [corrected] F 18). *J Nucl Med*. 2010;51:913–920.
- Forsberg A, Engler H, Almkvist O, et al. PET imaging of amyloid deposition in patients with mild cognitive impairment. *Neurobiol Aging*. 2008;29:1456–1465.
- Tryptsen V, DiBernardo A, Samtani M, Novak GP, Narayan VA, Raghavan N. Optimizing regions-of-interest composites for capturing treatment effects on brain amyloid in clinical trials. *J Alzheimers Dis*. 2015;43:809–821.
- Koeppel RA. Data analysis for amyloid PET imaging: longitudinal studies. Paper presented at: Human Amyloid Imaging; January 15–17, 2014; Miami, Florida.
- Brendel M, Hogenauer M, Delker A, et al. Improved longitudinal [¹⁸F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction. *Neuroimage*. 2015;108:450–459.
- Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009;65:403–413.
- Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta Neuropathol*. 2011;121:597–609.
- Landau SM, Marks SM, Mormino EC, et al. Association of lifetime cognitive engagement and low beta-amyloid deposition. *Arch Neurol*. 2012;69:623–629.
- Mormino EC, Kluth JT, Madison CM, et al. Episodic memory loss is related to hippocampal-mediated beta-amyloid deposition in elderly subjects. *Brain*. 2009;132:1310–1323.
- Landau S, Jagust W. Florbetapir processing methods. Alzheimer's Disease Neuroimaging Initiative website. http://adni.bitbucket.org/docs/UCBERKELEYAV45/ADNI_AV45_Methods_JagustLab_04.29.14.pdf.
- Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol*. 2013;12:357–367.
- Landau SM, Lu M, Joshi AD, et al. Comparing positron emission tomography imaging and cerebrospinal fluid measurements of beta-amyloid. *Ann Neurol*. 2013;74:826–836.
- Vlasko AG, Mintun MA, Xiong C, et al. Amyloid-beta plaque growth in cognitively normal adults: longitudinal [¹¹C]Pittsburgh compound B data. *Ann Neurol*. 2011;70:857–861.
- Landau SM, Thomas BA, Thurfjell L, et al. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. *Eur J Nucl Med Mol Imaging*. 2014;41:1398–1407.