**Supplementary Material**

**A Comparison of Partial Volume Correction Techniques for Measuring Change in Serial Amyloid PET SUVR**

 In this document we provide several supplementary figures and tables related to the main analyses in the primary document, followed by several related supplementary analyses that may be of interest to some readers.



Supplementary Figure 1. Flowchart of the Mayo cross-sectional SUVR processing pipeline.



Supplementary Figure 2. Flowchart of the Mayo Longitudinal SUVR processing pipeline.

Supplementary Table 1. (Tabular form of Figure 3 in the primary document). Difference between the annual rate of increase in PiB PET SUVR in clinically impaired versus that in clinically unimpaired subjects, when using each combination of measurement pipeline, PVC type, and reference region. Slopes for each group were assessed using a linear mixed-effects model of log-transformed SUVR values with separate fixed effect slopes and intercepts by impairment status. All methods showing differences > zero were considered equally plausible (we assume that amyloid should increase faster in impaired subjects, but the exact ground-truth difference is unknown). We plot difference in annualized SUVR change (y axis) as a percentage, rather than unscaled, to allow for comparison across methods and reference regions that have differing SUVR unit scales. Methods using GTM PVC produced group-wise differences that were larger but with much wider confidence intervals.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method**  | **Reference** | **PVC** | **Slope Difference (% SUVR) (95% CI)** |
| Mayo\_CrossSec | Cerebellar Crus | None | 0.283 (-0.274, 0.840) |
| Mayo\_CrossSec | Cerebellar Crus | 2-Comp. | 0.479 (-0.100, 1.058) |
| Mayo\_CrossSec | Cerebellar Crus | 3-Comp. | 0.498 (-0.173, 1.169) |
| Mayo\_CrossSec | Cerebellar Crus | GTM | 0.711 (-0.251, 1.672) |
| Mayo\_CrossSec | Cerebellar GM | None | 0.264 (-0.267, 0.796) |
| Mayo\_CrossSec | Cerebellar GM | 2-Comp. | 0.447 (-0.109, 1.003) |
| Mayo\_CrossSec | Cerebellar GM | 3-Comp. | 0.429 (-0.224, 1.083) |
| Mayo\_CrossSec | Cerebellar GM | GTM | 0.518 (-0.464, 1.500) |
| Mayo\_CrossSec | Whole Cerebellum | None | 0.338 (-0.154, 0.831) |
| Mayo\_CrossSec | Whole Cerebellum | 2-Comp. | 0.524 (0.004, 1.044) |
| Mayo\_CrossSec | Whole Cerebellum | GTM | 0.772 (-0.168, 1.712) |
| Mayo\_CrossSec | Pons | None | 0.470 (-0.069, 1.008) |
| Mayo\_CrossSec | Pons | 2-Comp. | 0.705 (0.134, 1.276) |
| Mayo\_CrossSec | Pons | GTM | 0.972 (-0.056, 2.001) |
| Mayo\_CrossSec | Supratentorial WM | None | 0.382 (-0.004, 0.768) |
| Mayo\_CrossSec | Supratentorial WM | 2-Comp. | 0.507 (0.088, 0.926) |
| Mayo\_CrossSec | Supratentorial WM | GTM | 1.667 (0.441, 2.892) |
| Mayo\_Longitudinal | Cerebellar Crus | None | 0.257 (-0.271, 0.784) |
| Mayo\_Longitudinal | Cerebellar Crus | 2-Comp. | 0.255 (-0.285, 0.795) |
| Mayo\_Longitudinal | Cerebellar Crus | 3-Comp. | 0.388 (-0.262, 1.038) |
| Mayo\_Longitudinal | Cerebellar Crus | GTM | 0.341 (-0.573, 1.254) |
| Mayo\_Longitudinal | Cerebellar GM | None | 0.206 (-0.298, 0.709) |
| Mayo\_Longitudinal | Cerebellar GM | 2-Comp. | 0.207 (-0.310, 0.724) |
| Mayo\_Longitudinal | Cerebellar GM | 3-Comp. | 0.327 (-0.308, 0.962) |
| Mayo\_Longitudinal | Cerebellar GM | GTM | 0.364 (-0.579, 1.306) |
| Mayo\_Longitudinal | Whole Cerebellum | None | 0.247 (-0.226, 0.719) |
| Mayo\_Longitudinal | Whole Cerebellum | 2-Comp. | 0.247 (-0.242, 0.735) |
| Mayo\_Longitudinal | Whole Cerebellum | GTM | 0.578 (-0.320, 1.476) |
| Mayo\_Longitudinal | Pons | None | 0.336 (-0.150, 0.821) |
| Mayo\_Longitudinal | Pons | 2-Comp. | 0.351 (-0.155, 0.857) |
| Mayo\_Longitudinal | Pons | GTM | 0.619 (-0.322, 1.559) |
| Mayo\_Longitudinal | Supratentorial WM | None | 0.261 (-0.016, 0.538) |
| Mayo\_Longitudinal | Supratentorial WM | 2-Comp. | 0.275 (-0.026, 0.577) |
| Mayo\_Longitudinal | Supratentorial WM | GTM | 1.719 (0.496, 2.941) |
| FreeSurfer\_CrossSec | Cerebellar GM | None | 0.262 (-0.250, 0.773) |
| FreeSurfer\_CrossSec | Cerebellar GM | GTM | 0.947 (-0.108, 2.003) |
| FreeSurfer\_CrossSec | Whole Cerebellum | None | 0.466 (-0.020, 0.951) |
| FreeSurfer\_CrossSec | Whole Cerebellum | GTM | 1.190 (0.066, 2.313) |
| FreeSurfer\_CrossSec | Pons | None | 0.409 (-0.142, 0.959) |
| FreeSurfer\_CrossSec | Pons | GTM | 1.022 (-0.132, 2.175) |
| FreeSurfer\_CrossSec | Supratentorial WM | None | 0.317 (-0.041, 0.675) |
| FreeSurfer\_CrossSec | Supratentorial WM | GTM | 1.307 (0.051, 2.563) |
| FreeSurfer\_Longitudinal | Cerebellar GM | None | 0.256 (-0.259, 0.771) |
| FreeSurfer\_Longitudinal | Cerebellar GM | GTM | 0.857 (-0.159, 1.872) |
| FreeSurfer\_Longitudinal | Whole Cerebellum | None | 0.438 (-0.045, 0.920) |
| FreeSurfer\_Longitudinal | Whole Cerebellum | GTM | 1.119 (0.041, 2.196) |
| FreeSurfer\_Longitudinal | Pons | None | 0.388 (-0.154, 0.929) |
| FreeSurfer\_Longitudinal | Pons | GTM | 0.945 (-0.172, 2.063) |
| FreeSurfer\_Longitudinal | Supratentorial WM | None | 0.362 (0.002, 0.722) |
| FreeSurfer\_Longitudinal | Supratentorial WM | GTM | * 1. .166, 2.632)
 |

# SUPPLEMENTARY ANALYSIS: ADDITIONAL PVC METHOD VARIANTS

## Introduction

 In this supplementary follow-up analysis, we repeated our main analysis while also including two additional variations of partial volume correction. These additional, less-common variations are modifications of the existing methods that each lie conceptually in-between the existing 3-compartment PVC and GTM. These are included in order to better understand the sources of GTM’s instability discovered in the primary analysis. In GTM PVC paradigms with parcellations similar to our analyses, the regions that most influence the stability of the cortical regions are those for supratentorial WM and those for non-brain tissue such as cerebrospinal fluid skull, dura, etc., because these regions are relatively large and border a large number of cortical regions of interest. The aim of this supplementary analysis is to compare variants that examine GTM’s estimates of signal in supratentorial WM and of non-brain tissue, respectively, to determine whether unstable estimation of each of these classes can explain GTM’s relative instability.

## Materials and Methods

 In this section we describe the two additional PVC variants included in this supplementary analysis. Both methods are not available via FreeSurfer, and thus are implemented only in the Mayo pipelines. All data and analyses (i.e., all data points on all plots that mirror the main analyses) are otherwise identical to as in the primary document.

### *Geometric Transfer Matrix, with zero-valued non-brain regions (GTM\_ZeroNonBrain)*

 In this method, we performed standard GTM but omitted all non-brain regions (cerebrospinal fluid, brain, skull, dura, etc.) by forcing the model to assume that the signal in these regions is zero. This assumption is unusual when taken in the context of GTM methods, but it matches the assumptions made by two-compartment and three-compartment voxel-based PVC, which also do not account for any spill-in of signal from non-brain regions. We include this variant in order to assess the longitudinal stability of GTM’s estimation of signal in non-brain regions by comparing it to standard GTM PVC. Specifically, if this modified GTM method performed significantly better than standard GTM, then we would be able to conclude that unstable estimation of signal in non-brain regions is a significant contributor to the instability of standard GTM.

### *Three-Compartment PVC with GTM WM Initialization (3-Comp., GTM WM Init.)*

 In this hybrid method we first ran the GTM\_ZeroNonBrain variant of GTM above, solely to obtain its estimated, corrected mean value for supratentorial WM. This estimated value for supratentorial WM was then used for standard three-compartment (Müller-Gärtner-style) PVC, replacing the standard step that derives this value from an atlas-derived mean of the centrum semiovale region. This hybrid is conceptually similar to what has been called Modified Müller-Gärtner (MMG) [1,2], but MMG does not necessarily use the ZeroNonBrain variant of GTM for the initial WM estimate, so we will not specifically call it MMG here. We included this method to assess the longitudinal stability of GTM’s estimation of the supratentorial WM signal by comparing it to standard three-compartment PVC with atlas-based WM estimation. Specifically, if the hybrid method performed better/worse than standard three-compartment, then we would be able to conclude that GTM’s estimation of supratentorial WM was more/less stable than the standard atlas-based WM estimate.

## Results

 We show the results of our supplementary analysis in Supplementary Figure 3. This figure is identical to Figure 3 in the main publication, other than the inclusions of the two additional PVC variant methods for the Mayo pipelines.



Supplementary Figure 3. Coefficient of variation (CV) in PiB PET SUVR when using each combination of measurement pipeline, PVC, and reference region. CV was estimated from a linear mixed-effects model of log-transformed SUVR values using 3 timepoints of PiB PET scans (n=278 subjects) with corresponding MRI.

Discussion

 Three-compartment PVC using the WM estimate from GTM trended toward slightly better stability than standard three-compartment PVC (but worse than two-compartment or no PVC), but all of these differences were not significant (Supplementary Figure 3). This finding suggests that GTM’s estimate of signal in supratentorial WM is at least not significantly worse than the standard atlas-based mean of the centrum semiovale, and thus that GTM’s estimate of WM is likely not a significant source of its instability relative to voxel-based methods.

 The relative performance of the GTM\_ZeroNonBrain variant was very mixed versus other PVC methods (Supplementary Figure 3). It performed significantly better than standard GTM when using the cerebellar crus, cerebellar GM, whole cerebellum, and pons reference regions for SUVR. However, it performed significantly worse than standard GTM when using a supratentorial WM reference region. In the (Mayo) cross-sectional pipelines, this variant performed comparably to the voxel-based methods for the cerebellar crus and cerebellar GM reference regions, and for the whole cerebellum and pons reference regions it performed in-between the voxel-based methods and standard GTM. When using the longitudinal pipeline, this method provided the most stable SUVRs (even more stable than no PVC) among the cerebellar Crus or GM reference regions, and it was comparable with the voxel-based methods when using the whole cerebellum as a reference region. Across Supplementary Figure 3, the GTM\_ZeroNonBrain variant situationally provided both some of the best and the worst longitudinal SUVR stability. This variability particularly depended on which reference region was used, suggesting that non-brain regions in the GTM model may differentially affect the stability of its signal estimates for infratentorial versus supratentorial regions. One could hypothesize that removing non-brain regions from the GTM model improved the stability of cerebellar/pontine estimates while destabilizing those of supratentorial WM. However, this would be inconsistent with the fact that these same SWM estimates performed reasonably when used in the new 3-compartment PVC variant. Future work will be needed to properly understand this GTM variant’s very varied performance.

## Conclusions

 Together, the results of this supplementary analysis suggest that relatively more of the measurement instability in GTM PVC for PiB PET may lie in its estimation of signal in non-brain regions than in supratentorial WM regions. Under some situations, removing non-brain regions from the model removes the difference in stability between GTM PVC and voxel-based methods, but for others (particularly, when used with supratentorial WM reference regions), stability becomes significantly worse. It is possible that the greater intensity inhomogeneity of these relatively sprawling regions may challenge GTM’s assumptions of a homogeneous distribution within each region. Future work will be needed to more thoroughly examine alternative parcellation schema and other tweaks for GTM PVC to improve its longitudinal measurement stability. We further discuss alternative GTM parcellation schema in the primary document.

# SUPPLEMENTARY ANALYSIS: ALTERNATE GROUP DEFINITIONS

## Introduction

 The analysis of group-wise differences in the main publication was designed to avoid circularity by defining groups based on clinical diagnosis, rather than amyloid measures. However, it may be helpful to also compare relative differences among groups where we are more confident that differences should be significant. To this end, here we repeat our longitudinal group-differences analysis but instead define the groups by their amyloid level at the third scan.

## Materials and Methods

 In this analysis, subjects were considered amyloid-positive according to their SUVR at the third scan, as measured by the Mayo Cross-Sectional method with no PVC and a cerebellar crus reference region. We used a threshold of SUVR>1.42 with this specific variant because this is the variant for which the 1.42 threshold was defined [3]. We recognize that defining group based on only one measurement methods is a biased approach, but analogous thresholds have not been developed for the other methods, which have different SUVR measurement scales. Otherwise, all analysis methods from the main publication were repeated.

## Results

 We show the results of our supplementary analysis in Supplementary Figure 4 and Supplementary Table 2. This figure is otherwise identical to Figure 3 in the main publication, but the group definitions have been changed as described above.



Supplementary Figure 4. Difference between the annual rate of increase in PiB PET SUVR in subjects that were amyloid-positive at visit 3 versus subjects that were amyloid negative at visit 3, when using each combination of measurement pipeline, PVC type, and reference region. Slopes for each group were assessed using a linear mixed-effects model of log-transformed SUVR values with separate fixed effect slopes and intercepts for each group. All methods showing differences > zero were considered equally plausible (we assume that amyloid should increase faster in A+ subjects, but the exact ground-truth difference is unknown). We plot difference in annualized SUVR change (y axis) as a percentage, rather than unscaled, to allow for comparison across methods and reference regions that have differing SUVR unit scales.

Supplementary Table 2. Tabular form of Supplementary Figure 4.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Reference** | **PVC** | **Slope Difference (% SUVR) (95% CI)** |
| Mayo\_CrossSec | Cerebellar Crus | None | 2.387 (2.033, 2.741) |
| Mayo\_CrossSec | Cerebellar Crus | 2-Comp. | 2.617 (2.252, 2.982) |
| Mayo\_CrossSec | Cerebellar Crus | 3-Comp. | 3.185 (2.768, 3.601) |
| Mayo\_CrossSec | Cerebellar Crus | GTM | 4.329 (3.711, 4.946) |
| Mayo\_CrossSec | Cerebellar GM | None | 2.418 (2.087, 2.748) |
| Mayo\_CrossSec | Cerebellar GM | 2-Comp. | 2.634 (2.291, 2.978) |
| Mayo\_CrossSec | Cerebellar GM | 3-Comp. | 3.190 (2.788, 3.592) |
| Mayo\_CrossSec | Cerebellar GM | GTM | 4.425 (3.799, 5.052) |
| Mayo\_CrossSec | Whole Cerebellum | None | 2.513 (2.210, 2.815) |
| Mayo\_CrossSec | Whole Cerebellum | 2-Comp. | 2.738 (2.425, 3.050) |
| Mayo\_CrossSec | Whole Cerebellum | GTM | 4.680 (4.080, 5.280) |
| Mayo\_CrossSec | Pons | None | 2.886 (2.557, 3.214) |
| Mayo\_CrossSec | Pons | 2-Comp. | 3.165 (2.826, 3.504) |
| Mayo\_CrossSec | Pons | GTM | 5.124 (4.466, 5.783) |
| Mayo\_CrossSec | Supratentorial WM | None | 1.793 (1.534, 2.052) |
| Mayo\_CrossSec | Supratentorial WM | 2-Comp. | 1.980 (1.700, 2.260) |
| Mayo\_CrossSec | Supratentorial WM | GTM | 5.668 (4.874, 6.462) |
| Mayo\_Longitudinal | Cerebellar Crus | None | 2.222 (1.883, 2.562) |
| Mayo\_Longitudinal | Cerebellar Crus | 2-Comp. | 2.323 (1.979, 2.666) |
| Mayo\_Longitudinal | Cerebellar Crus | 3-Comp. | 3.069 (2.668, 3.471) |
| Mayo\_Longitudinal | Cerebellar Crus | GTM | 3.962 (3.367, 4.558) |
| Mayo\_Longitudinal | Cerebellar GM | None | 2.238 (1.923, 2.553) |
| Mayo\_Longitudinal | Cerebellar GM | 2-Comp. | 2.354 (2.036, 2.673) |
| Mayo\_Longitudinal | Cerebellar GM | 3-Comp. | 3.111 (2.727, 3.495) |
| Mayo\_Longitudinal | Cerebellar GM | GTM | 4.241 (3.642, 4.841) |
| Mayo\_Longitudinal | Whole Cerebellum | None | 2.322 (2.033, 2.611) |
| Mayo\_Longitudinal | Whole Cerebellum | 2-Comp. | 2.430 (2.135, 2.726) |
| Mayo\_Longitudinal | Whole Cerebellum | GTM | 4.465 (3.889, 5.042) |
| Mayo\_Longitudinal | Pons | None | 2.651 (2.359, 2.942) |
| Mayo\_Longitudinal | Pons | 2-Comp. | 2.770 (2.465, 3.075) |
| Mayo\_Longitudinal | Pons | GTM | 4.709 (4.096, 5.321) |
| Mayo\_Longitudinal | Supratentorial WM | None | 1.368 (1.182, 1.554) |
| Mayo\_Longitudinal | Supratentorial WM | 2-Comp. | 1.460 (1.253, 1.667) |
| Mayo\_Longitudinal | Supratentorial WM | GTM | 5.872 (5.089, 6.656) |
| FreeSurfer\_CrossSec | Cerebellar GM | None | 2.372 (2.064, 2.680) |
| FreeSurfer\_CrossSec | Cerebellar GM | GTM | 4.995 (4.311, 5.680) |
| FreeSurfer\_CrossSec | Whole Cerebellum | None | 2.576 (2.274, 2.879) |
| FreeSurfer\_CrossSec | Whole Cerebellum | GTM | 5.263 (4.499, 6.027) |
| FreeSurfer\_CrossSec | Pons | None | 2.773 (2.423, 3.122) |
| FreeSurfer\_CrossSec | Pons | GTM | 5.332 (4.554, 6.110) |
| FreeSurfer\_CrossSec | Supratentorial WM | None | 1.477 (1.225, 1.729) |
| FreeSurfer\_CrossSec | Supratentorial WM | GTM | 5.476 (4.611, 6.340) |
| FreeSurfer\_Longitudinal | Cerebellar GM | None | 2.370 (2.059, 2.681) |
| FreeSurfer\_Longitudinal | Cerebellar GM | GTM | 4.691 (4.048, 5.334) |
| FreeSurfer\_Longitudinal | Whole Cerebellum | None | 2.592 (2.299, 2.886) |
| FreeSurfer\_Longitudinal | Whole Cerebellum | GTM | 5.091 (4.379, 5.803) |
| FreeSurfer\_Longitudinal | Pons | None | 2.711 (2.366, 3.056) |
| FreeSurfer\_Longitudinal | Pons | GTM | 5.146 (4.411, 5.882) |
| FreeSurfer\_Longitudinal | Supratentorial WM | None | 1.535 (1.285, 1.784) |
| FreeSurfer\_Longitudinal | Supratentorial WM | GTM | 5.493 (4.672, 6.315) |

## Discussion and Conclusions

 The comparative ordering of PVC methods in Supplementary Figure 4 is largely consistent with the ordering in Figure 3 in the primary document, where groups were defined by clinical status. As expected, the differences between groups were overall larger here than when using clinically-defined groupings, because these groups were directly defined by thresholding amyloid PET SUVR. As a result, all methods detected group-differences that were significantly greater than zero. GTM-based methods detected group-differences that were larger than the other methods, and these cross-method differences were mostly significant. However, as in the primary analysis, because the ground-truth difference between these groups is unknown, we consider all methods with differences significantly greater than zero to be equally valid. Additionally, because of the inherent circularity and bias present in this analysis, we recommend caution when interpreting its results and refer the reader instead to the group-wise analysis in the primary document.

# SUPPLEMENTARY ANALYSIS: CROSS-SECTIONAL GROUP-WISE DIFFERENCES

# Introduction

 Our aim in this study was to compare the relative precision in longitudinal PiB PET SUVR measurements when using various popular forms of PVC. Although cross-sectional comparisons have been explored in previous literature [1,4–6], in this section we provide our own cross-sectional comparisons using our same analysis methods and dataset (baseline SUVR values and diagnoses only) for use in direct comparison with our longitudinal results.

## Materials and Methods

 We used measurements from only the baseline scans from the primary analysis dataset. Groups were defined using the clinical diagnosis at this baseline (binary, impaired versus unimpaired), as in our primary longitudinal group-differences analysis. For each SUVR variant (combination of measurement pipeline, reference region, and PVC), we computed mean SUVR values (across both groups) and scaled all SUVR as a percentage of these means, to allow for comparison across methods and reference regions that have differing SUVR unit scales. We used the *t.test* function in R to assess differences between the groups and provide 95% confidence intervals. Longitudinal pipeline variants were omitted, since they are not applicable to cross-sectional data.

## Results

 We show the results of our cross-sectional group difference comparisons in Supplementary Figure 5 and Supplementary Table 3. In most reference regions, two-compartment PVC trended toward larger group-differences than no PVC, but these differences were not significant. Where applicable, three-compartment PVC trended toward larger group-differences than two-compartment PVC, but these differences were also not significant. GTM PVC had larger group-differences than all other methods, in all reference regions. These group-differences were significantly larger than when using no-PVC or two-compartment PVC methods, for all reference regions. Differences between the Mayo and FreeSurfer pipelines, and across reference regions, were not significant.



Supplementary Figure 5. Difference between cross-sectional, baseline PiB PET SUVR values in clinically impaired versus clinically unimpaired subjects, when using each combination of measurement pipeline, PVC type, and reference region. We plot differences as a percentage, rather than unscaled, to allow for comparison across methods and reference regions that have differing SUVR unit scales. GTM-based methods generally showed larger differences than no- or voxel-based PVC.

Supplementary Table 3. Tabular form of Supplementary Figure 5.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Reference** | **PVC** | **Mean Difference (% SUVR) (95% CI)** |
| Mayo\_CrossSec | Cerebellar Crus | None | 26.835 (19.005, 34.664) |
| Mayo\_CrossSec | Cerebellar Crus | 2-Comp. | 28.229 (20.105, 36.352) |
| Mayo\_CrossSec | Cerebellar Crus | 3-Comp. | 33.939 (24.189, 43.689) |
| Mayo\_CrossSec | Cerebellar Crus | GTM | 53.022 (38.256, 67.787) |
| Mayo\_CrossSec | Cerebellar GM | None | 26.550 (18.990, 34.110) |
| Mayo\_CrossSec | Cerebellar GM | 2-Comp. | 27.706 (19.830, 35.581) |
| Mayo\_CrossSec | Cerebellar GM | 3-Comp. | 33.274 (23.753, 42.794) |
| Mayo\_CrossSec | Cerebellar GM | GTM | 54.449 (39.479, 69.419) |
| Mayo\_CrossSec | Whole Cerebellum | None | 27.485 (19.767, 35.203) |
| Mayo\_CrossSec | Whole Cerebellum | 2-Comp. | 28.455 (20.397, 36.513) |
| Mayo\_CrossSec | Whole Cerebellum | GTM | 54.707 (40.009, 69.405) |
| Mayo\_CrossSec | Pons | None | 31.722 (23.309, 40.135) |
| Mayo\_CrossSec | Pons | 2-Comp. | 32.417 (23.525, 41.309) |
| Mayo\_CrossSec | Pons | GTM | 58.622 (43.331, 73.913) |
| Mayo\_CrossSec | Supratentorial WM | None | 18.382 (13.613, 23.151) |
| Mayo\_CrossSec | Supratentorial WM | 2-Comp. | 18.441 (13.264, 23.619) |
| Mayo\_CrossSec | Supratentorial WM | GTM | 60.754 (44.889, 76.619) |
| FreeSurfer\_CrossSec | Cerebellar GM | None | 25.225 (18.127, 32.322) |
| FreeSurfer\_CrossSec | Cerebellar GM | GTM | 58.302 (42.664, 73.940) |
| FreeSurfer\_CrossSec | Whole Cerebellum | None | 28.443 (20.725, 36.162) |
| FreeSurfer\_CrossSec | Whole Cerebellum | GTM | 62.945 (46.691, 79.199) |
| FreeSurfer\_CrossSec | Pons | None | 30.500 (22.184, 38.816) |
| FreeSurfer\_CrossSec | Pons | GTM | 64.028 (47.374, 80.681) |
| FreeSurfer\_CrossSec | Supratentorial WM | None | 15.409 (11.327, 19.492) |
| FreeSurfer\_CrossSec | Supratentorial WM | GTM | 64.848 (48.296, 81.399) |

## Discussion

 Across all other methodological variations, GTM found significantly larger cross-sectional group-differences than the other methods, and two- and three-compartment PVC trended toward larger group-differences than no PVC. This result was consistent with our hypotheses because PVC typically increases the dynamic range of output SUVR values when subjects have differing levels of atrophy, and GTM typically does this to a much greater magnitude than two- or three-compartment. Consistent with our longitudinal analyses (Figure 3, primary document), GTM had larger measurement variability (wider confidence intervals) than other methods. However, also as expected, differences between the clinical groups were much larger cross-sectionally (≈20-60% SUVR) than longitudinally (≈0-2.5% SUVR). Consequently, unlike in our longitudinal analyses, GTM’s larger measurement variability did not prevent effect-size differences between it and other methods from being significant.

 In the (un-eroded) supratentorial WM reference region, the no- and two-compartment PVC methods performed poorly because these methods do not attempt to correct for cross-contamination between the GM target region and the adjacent subcortical WM that is part of the reference region. GTM does account for this cross-contamination, and thus had much larger group-differences. Typically, WM reference regions are only used after erosion to exclude subcortical WM, but we did not include an eroded WM reference region because this is not available in FreeSurfer (see main document).

## Conclusions

 GTM significantly outperformed no-PVC and voxel-based-PVC methods for differentiating between impaired and unimpaired subjects using cross-sectional PiB PET SUVR values. Even though GTM is not commonly performed for cross-sectional amyloid PET, these results suggest that its increased cross-sectional group separation may significantly outweigh its penalty in additional measurement noise. However, the mechanisms behind these larger group-differences are uncertain, because these PVC techniques have been shown *not* to increase correlations between ante-mortem PiB PET measurements and post-mortem histological assessments [6], so more research is needed to determine the effects of PVC in amyloid PET imaging. We also emphasize that this supplementary analysis is only exploratory and retrospective, provided for comparison with our primary longitudinal analysis, and a larger, more though study would be needed to confirm its findings.

# REFERENCES

[1] Thomas BA, Erlandsson K, Modat M, Thurfjell L, Vandenberghe R, Ourselin S, Hutton BF (2011) The importance of appropriate partial volume correction for PET quantification in Alzheimer’s disease. *Eur J Nucl Med Mol Imaging* **38**, 1104–1119.

[2] Quarantelli M, Berkouk K, Prinster A, Landeau B, Svarer C, Balkay L, Alfano B, Brunetti A, Baron J, Salvatore M (2004) Integrated software for the analysis of brain PET/SPECT studies with partial-volume-effect correction. *J Nucl Med* **45**, 192–201.

[3] Jack CRJ, Wiste HJ, Weigand SD, Therneau TM, Lowe VJ, Knopman DS, Gunter JL, Senjem ML, Jones DT, Kantarci K, Machulda MM, Mielke MM, Roberts RO, Vemuri P, Reyes D, Petersen RC (2017) Defining imaging biomarker cut points for brain aging and Alzheimer’s disease. *Alzheimers Dement* **13**, 205–216.

[4] Su Y, Blazey TM, Snyder AZ, Raichle ME, Marcus DS, Ances BM, Bateman RJ, Cairns NJ, Aldea P, Cash L, Christensen JJ, Friedrichsen K, Hornbeck RC, Farrar AM, Owen CJ, Mayeux R, Brickman AM, Klunk W, Price JC, Thompson PM, Ghetti B, Saykin AJ, Sperling RA, Johnson KA, Scho PR, Buckles V, Morris JC, Benzinger TLS, Dominantly Inherited Alzheimer Network (2015) Partial volume correction in quantitative amyloid imaging. *Neuroimage* **107**, 55–64.

[5] Rousset OG, Ma Y, Evans AC (1998) Correction for partial volume effects in PET: principle and validation. *J Nucl Med* **39**, 904–911.

[6] Minhas DS, Price JC, Laymon CM, Becker CR, Klunk WE, Tudorascu DL, Abrahamson EE, Hamilton RL, Kofler JK, Mathis CA, Lopez OL, Ikonomovic MD (2018) Impact of partial volume correction on the regional correspondence between in vivo [C-11]PiB PET and postmortem measures of Aβ load. *Neuroimage Clin* **19**, 182–189.