



# Statistical Methods for Processing Neuroimaging Data from Two Different Sites with a Down Syndrome Population Application

Davneet S. Minhas<sup>1</sup>, Zixi Yang<sup>1</sup>, John Muschelli<sup>2</sup>, Charles M. Laymon<sup>1</sup>, Joseph M. Mettenberg<sup>1</sup>, Matthew D. Zammit<sup>3</sup>, Sterling Johnson<sup>3</sup>, Chester A. Mathis<sup>1</sup>, Ann D. Cohen<sup>1</sup>, Benjamin L. Handen<sup>1</sup>, William E. Klunk<sup>1</sup>, Ciprian M. Crainiceanu<sup>3</sup>, Bradley T. Christian<sup>3</sup>, and Dana L. Tudorascu<sup>1</sup>(✉)

<sup>1</sup> University of Pittsburgh, Pittsburgh, PA 15213, USA  
dlt30@pitt.edu

<sup>2</sup> Johns Hopkins University, Baltimore, MD 21205, USA

<sup>3</sup> University of Wisconsin-Madison, Madison, WI 53705, USA

**Abstract.** Harmonization of magnetic resonance imaging (MRI) and positron emission tomography (PET) scans from multi-scanner and multi-site studies presents a challenging problem. We applied the Removal of Artificial Voxel Effect by Linear regression (RAVEL) method to normalize T1-MRI intensities collected on two different scanners across two different sites as part of the Neurodegeneration in Aging Down syndrome (NiAD) study. The effects on FreeSurfer regional cortical thickness and volume outcome measures, in addition to FreeSurfer-based regional quantification of amyloid PET standardized uptake value ratio (SUVR) outcomes, were evaluated. A neuroradiologist visually assessed the accuracy of FreeSurfer hippocampus segmentations with and without the application of RAVEL. Quantitative results demonstrated that the application of RAVEL intensity normalization prior to running FreeSurfer significantly impacted both FreeSurfer volume and cortical thickness outcome measures. Visual assessment demonstrated that the application of RAVEL significantly improved FreeSurfer hippocampal segmentation accuracy. The RAVEL intensity normalization had little impact on PET SUVR measures.

**Keywords:** Harmonization · MRI · PET

## 1 Introduction

Positron emission tomography (PET) and structural magnetic resonance imaging (MRI) are two neuroimaging modalities frequently used in studies of Alzheimer's Disease (AD). [11C] Pittsburgh Compound B (PiB) PET standardized uptake value ratio (SUVR) and T1-weighted MRI-based volumetric measures, such as hippocampal volume and entorhinal cortical thickness, play a crucial role in studying the progression of AD in the elderly (Jack et al. 2017; Schwarz et al. 2016; Villemagne et al. 2011), autosomal dominant AD mutation carriers (Yau et al. 2015), and Down Syndrome (DS) populations (Lao et al. 2017).

Adults with DS are at high risk to reveal AD pathology in part due to the triplication of chromosome 21 encoding the amyloid precursor protein. Adults with DS are typically affected by AD pathology by their 30s, demonstrate AD-associated structural and functional brain changes in midlife, and have a 70–80% chance of clinical dementia by their 60s (Lao et al. 2017; Tudorascu et al. 2019).

Multi-site multimodal neuroimaging studies are becoming increasingly popular in studying neurodegenerative diseases, but there is a lack of statistical methods for combining data acquired across multiple sites and scanners. The use of different scanner types for data acquisition can introduce a significant amount of variability in the image processing, which propagates into the statistical analysis pipeline. Therefore, data harmonization should be considered a key element in the analysis of data from multi-site studies.

Previous research in addressing the multi-site confound and data harmonization has focused on either characterizing differences in derived imaging measures obtained across different scanners (Tummala et al. 2016) or on harmonizing derived measures using statistical models that account for scanner variability (Fortin et al. 2018; Johnson, Li and Rabinovic 2007). In both cases the focus has been on post-image processing derived measures.

Heinen and colleagues (Heinen et al. 2016) have characterized the robustness of measures from 1.5T and 3T MRI on the same subjects, using multiple automated techniques and evaluating accuracy using manual segmentations, and show that all automated methods demonstrate variability in accuracy depending on the segmented compartment/tissue type (Heinen et al. 2016). Pfefferbaum et al. characterized agreement between volume measures acquired at 1.5T and 3T using two techniques (FreeSurfer and SRI24) (Pfefferbaum et al. 2012). Their results demonstrated varying degrees of agreement depending on region and analysis method. Notably, the globus pallidus showed particularly poor agreement with either method. It has also been shown that including acquisition site as a confounding variable does not improve prediction accuracy of a cognitive outcome from imaging data (Rao, Monteiro and Mourao-Miranda 2017).

The focus of our study was to investigate the impact of a T1-weighted MRI intensity normalization method, Removal of Artificial Voxel Effect by Linear regression (RAVEL) (Fortin et al. 2016), on MRI- and PET-based outcome measures. The purpose of RAVEL is to remove unwanted variation from and standardize MRI voxel intensity values. RAVEL consists of two main steps: (1) factor analysis of control voxels within cerebrospinal fluid (CSF) space, where voxel intensities are known to be unassociated with disease status and clinical covariates, to estimate the principal directions of unwanted technical variability; and (2) linear regression at the whole brain level, including both covariates of biological interest and direction of unwanted variation.

We specifically focused on investigating MRI intensity normalization using RAVEL for improving neuroimaging outcomes derived from intensity normalized images. Four major steps were combined to create a novel characterization of the effects of RAVEL normalization on neuroimaging derived outcomes. The processes in place were: 1) RAVEL intensity normalization, 2) MRI segmentation using FreeSurfer v5.3 (FS) and derivation of cortical thickness and Region-of-Interest (ROI) volumes, 3)

PET SUVR quantification using ROIs derived from MRI images processed with and without RAVEL intensity normalization, and 4) neuroradiological evaluation to assess the accuracy of hippocampus segmentation derived with and without RAVEL intensity normalization. The effect of RAVEL normalization on MRI- and PET-based outcome measures was explicitly assessed. However, the harmonization of outcome measures was not, as the assessed cohort did not include same-subject data scanned across multiple scanners and/or sites, which would be necessary to explicitly and accurately assess harmonization.

## 2 Methods

### 2.1 Study Participants

Forty subjects with Down syndrome were selected from two sites participating in the Neurodegeneration in Aging Down Syndrome (NiAD) study, 20 subjects from the University of Pittsburgh (UPitt) and 20 from the University of Wisconsin (UWisc). The 20 subjects from each site were selected to span the age range of total study participants (32–54 y/o from UPitt, 30–54 y/o from UWisc), and the frequency of amyloid burden positivity as assessed by PiB PET (D. L. Tudorascu et al. 2018) was matched across sites. In all, 10 subjects with PiB scans indicative of amyloid deposition (PiB (+)) and 30 subjects without indication of deposition (PiB (–)) were used in this work. Amyloid classification was established as part of our existing standard analysis of the data and based on previously published methods from our group (Cohen et al. 2013).

### 2.2 Image Acquisition

**MRI.** Sagittally acquired structural T1-weighted MRIs were acquired at both sites. A Siemens MAGNETOM Prisma 3T scanner was employed at the UPitt (TE = 2.95 ms; TI = 900.0 ms; TR = 2300 ms; Weighting = PD; Flip Angle =  $11.0^\circ$ ; Pulse Sequence = GR/IR). A General Electric Discovery MR750 scanner was used at the UWisc (TE = 3.04 ms; TI = 400.0 ms; TR = 7.35 ms; Weighting = T1; Flip Angle =  $11.0^\circ$ ; Pulse Sequence = GR). MRIs from both sites had an in-plane pixel spacing of 1.05 mm and a sagittal thickness of 1.2 mm.

**PET.** PiB PET imaging was performed on a Siemens mCT Biograph PET/CT scanner (UPitt) and a Siemens EXACT HR + PET scanner (UWisc). For both sites, a nominal dose of 15 mCi PiB was delivered as a bolus injection over approximately 30 s, followed by a saline flush, and subjects were imaged over 50–70 min post-injection. PET images were reconstructed into four 5-min time frames using the manufacturers' software and included corrections for scatter, deadtime, random coincidences, and radioactive decay.

Reconstruction at the UPitt site was via ordered subset expectation maximization (OSEM) using 4 iterations of 12 subsets with no post filtering. Data required for attenuation and scatter correction were derived from a low-dose CT-scan, without contrast, acquired at the start of the scan session. Voxel size of the reconstructed PET image was 1.02 mm  $\times$  1.02 mm  $\times$  2.03 mm (axial).

PET images at the UWisc site were also produced via OSEM (4 iterations, 16 subsets) with no post filtering. Attenuation data were acquired in transmission scans performed using Ge-68/Ga-68 rotating rod sources. PET image voxel size was 2.57 mm × 2.57 mm × 2.43 mm (axial).

### 2.3 RAVEL Intensity Normalization

The application of RAVEL to a set of T1-weighted MRIs requires 3 conditions: (Condition 1) MRIs to be in a common anatomical space, (Condition 2) MRIs to be skull-stripped, and (Condition 3) a single control ROI indicating common CSF voxels across all images. Condition 2 is, in fact, not an explicit requirement of the RAVEL process but is for the WhiteStripe WM normalization step which precedes RAVEL [<https://cran.rproject.org/web/packages/WhiteStripe/index.html>]. The MRI preprocessing steps undertaken to meet these conditions are outlined below.

First, MRIs were processed through the Statistical Parametric Mapping v12 (SPM12) Segment tool (Ashburner and Friston 2005) in order to generate forward and inverse nonlinear registration transforms to SPM12-defined Montreal Neurological Institute (MNI) space and subject-specific GM, WM, and CSF tissue probability maps in both subject native space and MNI space. Incidentally, the application of SPM Segment tool is commonly used as part of the Centiloid process in amyloid PET imaging studies of Alzheimer's disease and aging in Down syndrome (Klunk et al. 2015; Tudorascu et al. 2018).

To satisfy Condition 1, native-space MRIs were warped to MNI space using subject-specific SPM12-defined forward transforms. In addition, MNI-space MRIs were bias corrected using the “N4BiasFieldCorrection” function implemented within the “ANTsR” library (Avants 2019), applied in RStudio version 1.2.1335.

To satisfy Condition 2, a single MNI-space intracranial mask was generated by summing the GM, WM, and CSF tissue prior probability maps provided within SPM12 (TPM.nii), thresholding the sum image at 0.1 and binarizing the result. To remove noncontiguous extracranial voxels from the MNI-space intracranial mask, the “cluster” function within FSL FMRIB Software Library v5.0 (Jenkinson et al. 2012) was used to identify and isolate the largest cluster. All MNI-space bias-corrected MRIs were then masked using the intracranial mask to generate skull-stripped MRIs.

To satisfy Condition 3, MNI-space CSF tissue probability maps were first thresholded at a value of 0.3 and binarized to generate subject-specific MNI-space CSF ROIs. The intersection of all MNI-space CSF ROIs was then calculated to generate a single CSF control ROI indicating all common CSF voxels across subjects.

RAVEL intensity normalization was subsequently performed using the “normalizeRAVEL” function within the “RAVEL” R library (<https://github.com/Jfortin1/RAVEL>) with MNI-space bias-corrected skull-stripped MRIs as “input.files” input, MNI-space intracranial mask as “brain.mask” input, and MNI-space CSF ROI as “control.mask” input. Default parameters were used. Resulting RAVEL-intensity normalized MRIs were also in MNI space with WM voxel intensities centered on a value of 0, such that GM and CSF voxel intensities were negative. This is a consequence of WhiteStripe normalization, which is run by default within the “normalizeRAVEL” function.

To remove negative voxel intensities from RAVEL-normalized images, the minimum voxel value across all images was added to each RAVEL-normalized MRI. The voxel intensity-shifted MNI-space RAVEL-normalized MRIs were subsequently spatially warped back to subject-specific native space using the SPM12 Segment-derived inverse nonlinear transforms.

## 2.4 MRI Processing

Native-space unnormalized and RAVEL-normalized T1-weighted MRIs were processed using FreeSurfer v5.3 (FS) to generate regional cortical thicknesses, volumes, and binary ROI images for sampling PiB PET images. Unnormalized MRIs were processed through FS using default parameters. RAVEL-normalized images were processed using the “-noskullstrip” flag given that skull-stripping had already been performed as part of the RAVEL normalization process.

## 2.5 PiB PET Processing

The multi-frame (5 min/frame) PiB PET images were visually inspected for frame-to-frame motion, and motion corrected when appropriate, using PMOD software version 3.711 (PMOD Technologies LLC, <https://www.pmod.com/>). Single-frame PET images were generated by averaging the time frames from 50–70 min post-injection. The single-frame PET images were subsequently registered to subject-specific unnormalized T1-weighted MRIs using the “Coregister: Estimate and Reslice” tool within SPM12. Previously described PiB-specific composite ROIs were used to sample single-frame PET images, generating regional PiB radioactivity concentration values (Tudorascu et al. 2018). Regional values were normalized by dividing by the FS-defined cerebellar GM radioactivity, resulting in regional SUVRs. PiB regional SUVR values were derived using both unnormalized MRI-derived FS ROIs and RAVEL-normalized MRI-derived FS ROIs. However, the PiB-PET images were not re-registered to RAVEL-normalized MRIs. The results of this process, for each subject, were two sets of PiB-PET SUVR measures based on FS ROIs: one set from the unnormalized MRI and one set from the RAVEL-normalized MRI. For the purposes of this work, we evaluated PET results using a striatal ROI, a region of early accumulation of amyloid deposition in DS (Tudorascu et al., 2019), and a global ROI that is the union of all our standard amyloid quantitation regions. Amyloid status (PiB (–) or PiB (+)) was not reevaluated as part of this work.

## 2.6 Statistical Evaluation

Descriptive statistics including means and standard deviations (SD) were computed for each cortical thickness and ROI volume as well as for the global PiB SUVR derived measures with and without the use of RAVEL. Group differences were determined using Cohen’s *d* effect sizes for each measure between PiB (–) and PiB (+) subject groups.

Intraclass correlation coefficients (ICCs) and their corresponding 95% confidence intervals (CI) were calculated to measure agreement between the derived measures from unnormalized MRI and RAVEL-normalized MRI-derived FS cortical thicknesses, volumes, and PiB SUVR values. FS volumes were normalized to FS-reported total intracranial volume (tICV) before statistical assessment to account for sex-specific differences in anatomical volumes (Schwarz et al. 2016).

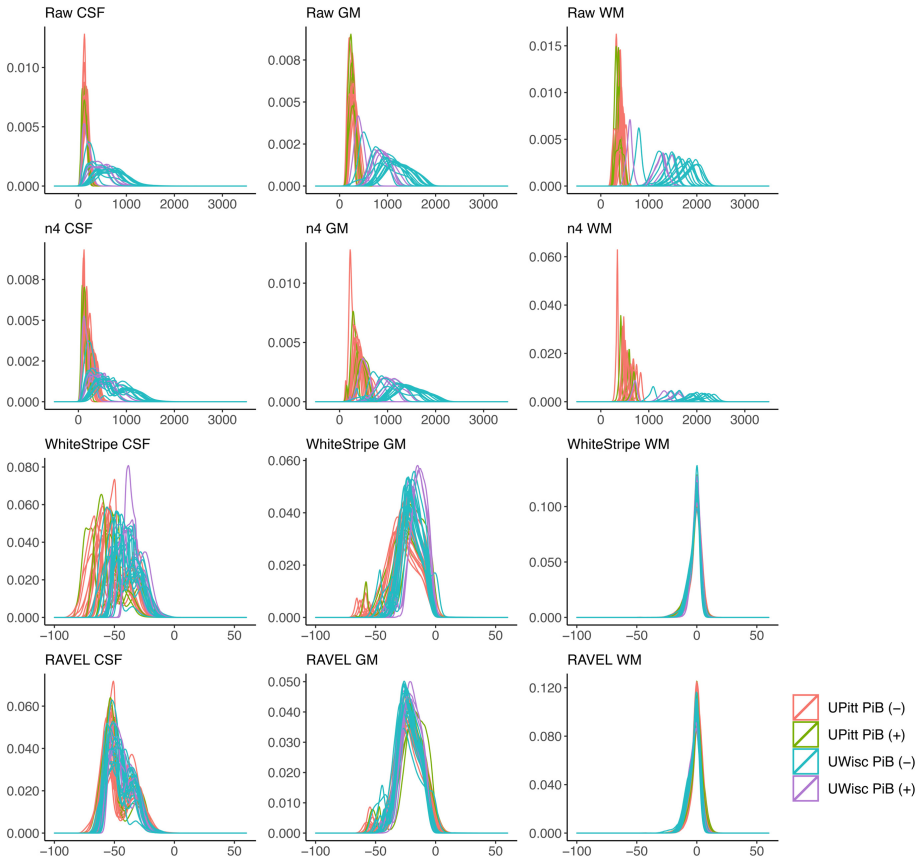
A neuroradiologist (JMM) visually rated unnormalized and RAVEL-normalized MRI-derived FS hippocampal masks on unnormalized MRIs using a four-point scale to rate the accuracy of the segmentations (1 = poor, 2 = some errors, 3 = good, 4 = excellent). The rater was blinded to subject information, including age, gender, amyloid status and to the origin and methodology from which hippocampal masks were derived. The Wilcoxon signed-rank test was used to assess pairwise differences in rater ranking between unnormalized MRI-derived FS hippocampal masks and RAVEL-normalized MRI-derived FS hippocampal masks.

### 3 Results

Density plots, smoothed versions of histograms, of MRI voxel intensities are shown by site and amyloid status at multiple preprocessing steps (Raw, after applying n4 correction, after applying White Stripe and, last, after RAVEL) in Fig. 1. It can be observed that after RAVEL intensity normalization the histograms of intensities have a much greater overlap, reinforcing that RAVEL normalization significantly reduces MRI voxel intensity variability.

Descriptive statistics and PiB(-) to PiB(+) effect sizes for volumes, cortical thicknesses, and SUVR are presented in Table 1. The derived PET global SUVR and striatum SUVR measures were almost identical with or without the use of RAVEL in the preprocessing stream. The estimated regional brain volumes obtained using RAVEL were larger for left and right hippocampus in both PiB(-) and PiB (+) groups. For assessing variability, standard deviations were larger for the PiB (+) group, SD = 0.67 for left hippocampus and SD = 0.52 for right hippocampus when RAVEL was used compared to SD = 0.57 for left hippocampus and SD = 0.50 for right hippocampus when RAVEL was not used. In the PiB (-) group, the SD was smaller and almost identical regardless of the use of RAVEL.

However, the cortical thickness values were much more variable, with no clear pattern observed. For example, the means and SDs for the right entorhinal values are almost identical across methods, while left entorhinal values diverged (3.77 mm mean with SD = 0.40 for PiB (-) when RAVEL was not used and 3.69 mm mean with SD = 0.39 when RAVEL was used). Similarly, in the PiB (+) group, the values were 3.37 mm, SD = 0.44 without RAVEL and 3.49 mm, SD = 0.42 with RAVEL. The computed Cohen's d effect sizes between PiB (+) and PiB (-) were 0.95 without RAVEL and 0.45 with RAVEL.



**Fig. 1.** Density plots of MRI voxel intensities by site (University of Pittsburgh [UPitt] and University of Wisconsin [UWisc]), amyloid classification status (PiB (-) and PiB (+)), and tissue type (cerebrospinal fluid [CSF], grey matter [GM], and white matter [WM]) for multiple steps of the RAVEL intensity normalization process (unnormalized [Raw], bias corrected [n4], WhiteStripe-normalized [WhiteStripe], and RAVEL-normalized [RAVEL]). Results shown for RAVEL are prior to the constant intensity shift described in the text.

ICCs are shown for SUVR, volumes, and cortical thickness between outcomes obtained with and without RAVEL in Table 2. The ICCs show very good agreement between PET SUVR measures, with values greater than 0.9. The ICC values were lower for volumes, with a range of ICC = 0.365 (95% CI: 0.066, 0.604) in the right amygdala up to ICC = 0.830 (95% CI: 0.703, 0.906) for the right hippocampus. The ICCs for the cortical thicknesses were much lower, specifically for middle temporal left (0.228 (95% CI: -.083, 0.5)) and middle temporal right (0.33 (95% CI: 0.028, 0.579)). Areas with highest ICC values for the cortical thickness were left fusiform (0.851 (95% CI: 0.589, 0.864)) and right fusiform (0.871 (95% CI: 0.77, 0.929)).

**Table 1.** Descriptive statistics for MRI- and PET- based outcome measures stratified by amyloid status (PiB (-) and PiB (+)) and MRI intensity normalization method (unnormalized [Raw] and RAVEL-normalized [RAVEL]), and Cohen's d effect sizes between PiB (-) and PiB (+) groups.

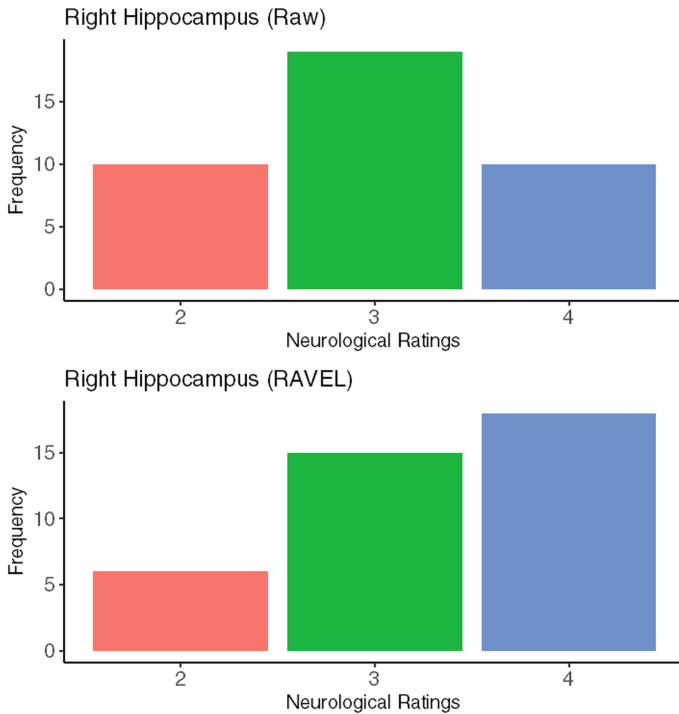
Region	Mean (SD)				Effect size	
	PiB (-)		PiB (+)		Raw	RAVEL
	N = 30		N = 10			
	Raw	RAVEL	Raw	RAVEL		
<b>PiB SUVR</b>						
Global	1.10 (0.07)	1.11 (0.07)	1.78 (0.40)	1.80 (0.39)	3.19	3.26
Striatum	1.28 (0.18)	1.31 (0.19)	2.26 (0.37)	2.30 (0.38)	3.86	3.80
<b>Volume (ICV normalized)</b>						
Left Hippo-campus	3.00 (0.32)	3.22 (0.33)	2.24 (0.57)	2.38 (0.67)	1.83	1.83
Right Hippo-campus	3.02 (0.31)	3.16 (0.30)	2.27 (0.50)	2.47 (0.52)	1.97	1.81
Left Amygdala	1.45 (0.18)	1.32 (0.25)	1.07 (0.32)	0.96 (0.26)	1.64	1.34
Right Amygdala	1.56 (0.26)	1.32 (0.28)	1.13 (0.32)	1.06 (0.21)	1.51	0.93
<b>Cortical thickness (mm)</b>						
Left Entorhinal	3.77 (0.40)	3.68 (0.39)	3.37 (0.44)	3.49 (0.42)	0.95	0.45
Right Entorhinal	3.93 (0.37)	3.94 (0.38)	3.55 (0.48)	3.55 (0.48)	0.91	0.90
Left Fusiform	2.82 (0.20)	2.87 (0.18)	2.65 (0.21)	2.71 (0.16)	0.81	0.90
Right Fusiform	2.95 (0.19)	2.94 (0.18)	2.78 (0.27)	2.82 (0.26)	0.77	0.53
Left Inferior Temporal	2.79 (0.27)	2.93 (0.20)	2.65 (0.22)	2.82 (0.14)	0.50	0.59
Right Inferior Temporal	2.86 (0.26)	2.92 (0.21)	2.75 (0.20)	2.82 (0.19)	0.41	0.47
Left Middle Temporal	2.86 (0.26)	3.09 (0.20)	2.75 (0.21)	2.94 (0.21)	0.43	0.71
Right Middle Temporal	2.80 (0.27)	3.01 (0.25)	2.70 (0.27)	2.86 (0.28)	0.37	0.58

Neuroradiological ratings revealed that RAVEL intensity normalization significantly improved right hippocampal segmentation accuracy (Wilcoxon W-value = 69, p-value = 0.02) as shown in Fig. 2. When RAVEL was used, the right hippocampal segmentation was excellent for 18 subjects as compared to only 10 when RAVEL was not used. A single case is illustrated in Fig. 3, showing the hippocampus segmentation by method with arrows pointing to the areas that were incorrectly segmented without RAVEL normalization. The arrows point to voxels that were incorrectly classified as hippocampus when RAVEL was not implemented (in red).

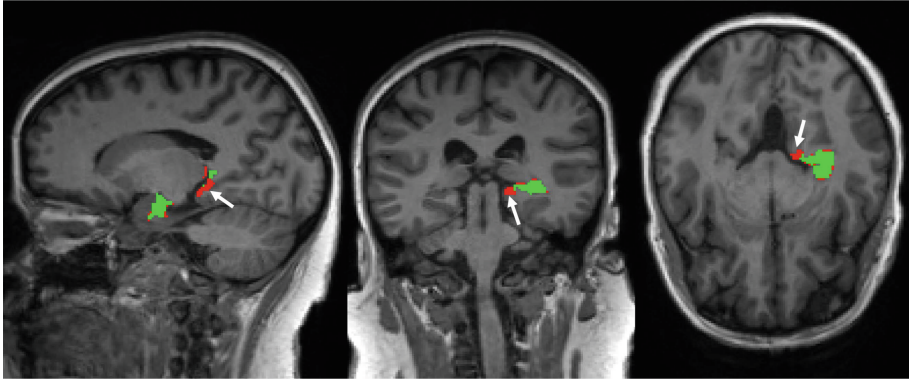


**Table 2.** Intraclass correlation coefficients (ICC) between unnormalized and RAVEL-intensity normalized FreeSurfer-based regions of interest outcome measures for both sites (University of Pittsburgh and University of Wisconsin) combined.

Region	ICC (95% CI)
<b>PiB SUVR</b>	
Global	0.998 (0.996, 0.999)
Striatum	0.995 (0.992, 0.998)
<b>Volume</b>	
Left Hippocampus	0.754 (0.583, 0.862)
Right Hippocampus	0.830 (0.703, 0.906)
Left Amygdala	0.616 (0.382, 0.776)
Right Amygdala	0.365 (0.066, 0.604)
<b>Cortical thickness</b>	
Left Entorhinal	0.683 (0.477, 0.819)
Right Entorhinal	0.758 (0.589, 0.864)
Left Fusiform	0.851 (0.738, 0.918)
Right Fusiform	0.871 (0.770, 0.929)
Left Inferior Temporal	0.474 (0.197, 0.682)
Right Inferior Temporal	0.647 (0.425, 0.796)
Left Middle Temporal	0.228 (-0.083, 0.500)
Right Middle Temporal	0.330 (0.028, 0.579)



**Fig. 2.** Visual ratings of FreeSurfer-based right hippocampus segmentations based on unnormalized (Raw) and RAVEL-intensity normalized (RAVEL) MRIs. A four-point rating scale was used, such that 1 = poor, 2 = some errors, 3 = good, and 4 = excellent. (Color figure online)



**Fig. 3.** Sagittal, coronal, and axial slices of a single subject MRI with right hippocampal segmentations overlaid generated without the use of RAVEL (red) and with RAVEL (green). Arrows indicate extraneous non-hippocampus voxels classified as hippocampus prior to RAVEL intensity normalization. (Color figure online)

## 4 Discussion

In this study we show that using the RAVEL intensity normalization technique in a multi-site imaging study of neurodegeneration in Down syndrome leads to increased accuracy in hippocampal segmentation, an area of high significance for characterizing Alzheimer's disease pathology. However, it remains unclear if RAVEL intensity normalization removes site- and scanner- specific variability in derived structural MRI outcomes, such as volume and cortical thickness.

In terms of PET SUVR measures, observed differences were small between the values obtained when RAVEL was used compared to when it was not used (Table 2). This is likely due to the differences in spatial resolution of PET imaging ( $\sim 5$  mm) relative to MRI ( $\sim 1$  mm), such that voxel-level differences in MRI-based ROI segmentation may have little impact on the related PET ROI outcome measure.

Unlike the results for the PET SUVRs, the MRI-measured cortical thickness and regional volume data exhibited much larger variability between the two preprocessing methods (with and without RAVEL). RAVEL clearly impacted volume and cortical thickness measures as exhibited by low levels of ICC agreement ( $<0.5$  ICC) in the right amygdala volume and left inferior temporal, left middle temporal, and right middle temporal cortical thicknesses. It is, however, unclear whether the application of RAVEL resulted in superior cortical thickness measures given the seemingly random pattern of change in effect size between PiB (-) and PiB (+) subjects with and without RAVEL.

PiB (+) subjects exhibiting amyloid deposition are further along the spectrum of AD pathophysiology than PiB (-) subjects and thus may demonstrate an increased level of cortical atrophy, or smaller volumes and cortical thicknesses, than PiB (-) subjects (Jack et al. 2013). Therefore, it is expected that a method which improves MRI-based regional segmentation accuracy would increase the effect sizes of volumes

and cortical thicknesses between PiB (−) and PiB (+) groups. Five of the eight cortical thickness regions presented in Table 1 had improved effect sizes with the application of RAVEL, but the greatest difference in cortical thickness effect size was observed in the left entorhinal region with no normalization outperforming RAVEL-intensity normalization. For cortical thickness measurements, the features of the cortical ribbon present challenges in processing that are dependent on size, shape and location within the particular ribbon.

In conclusion, applying RAVEL normalization during preprocessing of image data acquired at different sites using different scanners reduced variability in the distribution of voxel intensities and improved the accuracy of hippocampal segmentation but not the discriminatory power of quantitative FS-based outcome measures. The primary purpose of this work was to assess the effects of RAVEL intensity normalization on MRI- and PET-derived outcome measures. However, the efficacy of RAVEL in removing site- and scanner- specific variability from volume and cortical thickness measures was not assessed. Future work will address this by applying RAVEL on same-subject data acquired on multiple scanners to more accurately assess RAVEL's effect on harmonization of imaging outcome measures than was possible with this cohort. The effect of each preprocessing step (n4 bias correction, skull-stripping, and WhiteStripe intensity normalization) on the accuracy and harmonization of MRI- and PET- based outcome measures of same-subject data across different scanners will also be more thoroughly characterized.

RAVEL can be easily incorporated into the preprocessing stream, and the code is available online (<https://github.com/Jfortin1/RAVEL>) (Fortin et al. 2016).

**Acknowledgement.** This research investigation was funded by the National Institute of Aging (R01 AG031110 to B.L.H and B.T.C; U01AG051406 to B.L.H., W.E.K., and B.T.C., U01AG051406-05S1 to C.M.L and D.L.T. and R01 AG063752 to D.L.T.).

## References

- Ashburner, J., Friston, K.J.: Unified segmentation. *Neuroimage* **26**(3), 839–851 (2005)
- Avants, B.B.: ANTsR: ANTs in R: Quantification Tools for Biomedical Images (2019)
- Cohen, A.D., Mowrey, W., Weissfeld, L.A., Aizenstein, H.J., McDade, E., Mountz, J.M., Nebes, R.D., Saxton, J.A., Snitz, B., De Kosky, S., Williamson, J., Lopez, O.L., Price, J.C., Mathis, C.A., Klunk, W.E.: Classification of amyloid-positivity in controls: comparison of visual read and quantitative approaches. *Neuroimage* **71**, 207–215 (2013). <https://doi.org/10.1016/j.neuroimage.2013.01.015>
- Fortin, J.-P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooperh, C., Favai, M., Mc Grath, P.J., Mc Innis, M., Phillips, M.L., Trivedi, M.H., Weissmangl, M.M., Shinohara, R.T.: Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* **167**, 104–120 (2018). <https://doi.org/10.1016/j.neuroimage.2017.11.024>
- Fortin, J.-P., Sweeney, E.M., Muschelli, J., Crainiceanu, C.M., Shinohara, R.T.: Removing inter-subject technical variability in magnetic resonance imaging studies. *NeuroImage* **132**, 198–212 (2016). <https://doi.org/10.1016/j.neuroimage.2016.02.036>

- Heinen, R., Bouvy, W.H., Mendrik, A.M., Viergever, M.A., Biessels, G.J., et al.: Robustness of automated methods for brain volume measurements across different MRI field strengths. *Plos one* **11**(10), e0165719 (2016). <https://doi.org/10.1371/journal.pone.0165719>
- Jack Jr., C.R., Knopman, D.S., Jagust, W.J., Petersen, R.C., Weiner, M.W., Aisen, P.S., Shaw, L. M., Vemuri, P., Wiste, H.J., Weigand, S.D., GLesnick, T., Pankratz, V.S., CDonohue, M., Trojanowski, J.Q.: Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **12**(2), 207–216 (2013). [https://doi.org/10.1016/s1474-4422\(12\)70291-0](https://doi.org/10.1016/s1474-4422(12)70291-0)
- Jack Jr., C.R., Wiste, H.J., Weigand, S.D., Therneau, T.M., Lowe, V.J., Knopman, D.S., Gunter, J.L., Senjem, M.L., Jones, D.T., Kantarci, K., Machulda, M.M., Mielke, R.O., Roberts, M.M., Vemuri, P., Reyes, D.A., Petersen, R.C.: Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement* **13**(3), 205–216 (2017). <https://doi.org/10.1016/j.jalz.2016.08.005>
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: *FSL*. *Neuroimage* **62**(2), 782–790 (2012)
- Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**(1), 118–127 (2007)
- Klunk, W.E., Koeppe, R.A., Price, J.C., Benzinger, T.L., Devous, M.D. Sr., Jagust, W.J., Johnson, K.A., Mathis, C.A., Minhas, D., Pontecorvo, M.J., Rowe, C.C., Skovronsky, D.M., Mintun, M.A.: The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers Dement*, **11**(1), 1–15, e11-14 (2015). <https://doi.org/10.1016/j.jalz.2014.07.003>
- Lao, P.J., Handen, B.L., Betthausen, T.J., Mihaila, I., Hartley, S.L., Cohen, A.D., Bulova, P.D., Tumuluru, R.V., Murali, D., Mathis, C.A., Barnhart, T.E., Stone, C.K., Price, J.C., Devenny, D.A., Maillick, M.R., Klunk, W.E., Johnson, S.C., Christian, B.T.: Longitudinal changes in amyloid positron emission tomography and volumetric magnetic resonance imaging in the nondemented Down syndrome population. *Alzheimers Dement (Amst)* **9**, 1–9 (2017). <https://doi.org/10.1016/j.dadm.2017.05.001>
- Pfefferbaum, A., Rohlfing, T., Rosenbloom, M.J., Sullivan, E.V.: Combining atlas-based parcellation of regional brain data acquired across scanners at 1.5 T and 3.0 T field strengths. *Neuroimage* **60**(2), 940–951 (2012)
- Rao, A., Monteiro, J.M., Mourao-Miranda, J.: Predictive modelling using neuroimaging data in the presence of confounds. *NeuroImage* **150**, 23–49 (2017). <https://doi.org/10.1016/j.neuroimage.2017.01.066>
- Schwarz, C.G., Gunter, J.L., Wiste, H.J., Przybelski, S.A., Weigand, S.D., Ward, C.P., Senjem, M.L., Vemuri, P., Murray, M.E., Dickson, D.W., Parisi, J.E., Kantarci, K., Weiner, M.W., Petersen, R.C., Jack, C.R.: A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage Clin.* **11**, 802–812 (2016). <https://doi.org/10.1016/j.nicl.2016.05.017>
- Tudorascu, D.L., Anderson, S.J., Minhas, D.S., Yu, Z., Comer, D., Lao, P., Hartley, S., Laymon, C.M., Snitz, B.E., Lopresti, B.J., Johnson, S., Price, J.C., Mathis, C.A., Aizenstein, H.J., Klunk, W.E., Handen, B.L., Christian, B.T., Cohen, A.D.: Comparison of longitudinal A $\beta$  in nondemented elderly and Down syndrome. *Neurobiol. Aging* **73**, 171–176 (2019). <https://doi.org/10.1016/j.neurobiolaging.2018.09.030>
- Tudorascu, D.L., Minhas, D.S., Lao, P.J., Betthausen, T.J., Yu, Z., Laymon, C.M., Lopresti, B.J., Mathis, C.A., Klunk, W.E., Handen, B.L., Christian, B.T., Cohen, A.D.: The use of Centiloids for applying [(11)C] PiB classification cutoffs across region-of-interest delineation methods. *Alzheimers Dement (Amst)* **10**, 332–339 (2018). <https://doi.org/10.1016/j.dadm.2018.03.006>

- Tummala, S., Chu, R., Khalid, F., Dupuy, S., Tauhid, S., Healy, B., Bakshi, R.: Cortical Thickness Measurements from 1.5 t vs. 3t MRI in Healthy Subjects and Patients with Multiple Sclerosis (P4. 179). In: AAN Enterprises (2016)
- Villemagne, V.L., Pike, K.E., Chetelat, G., Ellis, K.A., Mulligan, R.S., Bourgeat, P., Ackermann, U., Jones, G., Szoeke, C., Salvado, O., Martins, R., O'Keefe, G., Mathis, C.A., Klunk, W.E., Ames, D., Masters, C.L., Rowe, C.C.: Longitudinal assessment of Abeta and cognition in aging and Alzheimer disease. *Ann. Neurol.* **69**(1), 181–192 (2011). <https://doi.org/10.1002/ana.22248>
- Yau, W.W., Tudorascu, D.L., McDade, E.M., Ikonomic, S., James, J.A., Minhas, D., Mowrey, W., Sheu, L.K., Snitz, B.E., Weissfeld, L., Gianaros, P.J., Aizenstein, H.J., Price, J.C., Mathis, C.A., Lopez, O.L., Klunk, W.E.: Longitudinal assessment of neuroimaging and clinical markers in autosomal dominant Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* **14**(8), 804–813 (2015). [https://doi.org/10.1016/S1474-4422\(15\)00135-0](https://doi.org/10.1016/S1474-4422(15)00135-0)