Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10 (2018) 332-339

Neuroimaging

The use of Centiloids for applying [^{11}C]PiB classification cutoffs across region-of-interest delineation methods

Dana L. Tudorascu[a,*], Davneet S. Minhas[b], Patrick J. Lao[c,d], Tobey J. Betthauser[c,d], Zheming Yu[b], Charles M. Laymon[b], Brian J. Lopresti[b], Chet A. Mathis[b], William E. Klunk[e], Benjamin L. Handen[e], Bradley T. Christian[c,d], Ann D. Cohen[e]

[a]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[b]Department of Radiology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[c]Department of Medical Physics, University of Wisconsin–Madison, School of Medicine, Madison, WI, USA
[d]Waisman Center, University of Wisconsin–Madison, School of Medicine, Madison, WI, USA
[e]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Abstract

Introduction: Centiloid standardization was developed to establish a quantitative outcome measure of amyloid burden that could accommodate the integration of different amyloid positron emission tomography radiotracers or different methods of quantifying the same tracer. The goal of this study was to examine the use of Centiloids for establishing amyloid classification cutoffs for differing region-of-interest (ROI) delineation schemes.

Methods: Using ROIs from hand-drawn delineation in native space as the gold standard, we compared standard uptake value ratios obtained from the 6 hand-drawn ROIs that determine amyloid-positivity classification with standard uptake value ratio obtained from 3 different automated techniques (FreeSurfer, Statistical Parametric Mapping, and superimposed hand-drawn ROIs in Pittsburgh Compound B template space). We tested between-methods reliability using repeated measures models and intraclass correlation coefficients.

Results: We found high reliability between the hand-drawn standard method and other methods for almost all the regions considered. However, small differences in standard uptake value ratio were found to lead to unreliable classifications when the hand-drawn native space-derived cutoffs were used across other ROI delineation methods.

Discussion: The use of Centiloid standardization greatly improved the agreement of Pittsburgh Compound B classification across methods and may serve as an alternative method for applying cutoffs across methodologically different outcomes.

Keywords: Amyloid; Centiloid standardization; Down syndrome; ROI delineation in PET studies; ROI cutoff PET

1. Introduction

Amyloid positron emission tomography (PET) data are typically quantified using regions of interest (ROIs) delineated on structural MRI images using manual, or more recently, automated methods [1–7]. However, ROI segmentation on magnetic resonance imaging (MRI) images can be challenging in populations with brain structure abnormalities, such as Alzheimer's disease (AD) or Down syndrome (DS), particularly when automated processing routines are employed [8]. Differences in ROI delineation could substantially affect statistical outcomes when quantifying [^{11}C]Pittsburgh Compound B (PiB) PET standardized uptake value ratio (SUVR), which play a crucial role in studying the progression of AD in the elderly [9,10], autosomal dominant AD mutation carriers [11], and

*Corresponding author. Tel.: (412) 246-5692; Fax: (412) 246-6873.
E-mail address: dlt30@pitt.edu

https://doi.org/10.1016/j.dadm.2018.03.006

DS populations [8]. Therefore, a standardized method of ROI delineation would be useful in characterizing these disease populations when using SUVR outcomes.

ROIs manually defined in native space are particularly robust, as trained manual raters can better account for structural abnormalities, poor signal-to-noise ratio, and motion artifacts in MRI data. However, manual ROI tracings are susceptible to individual variability, and the process is time consuming, especially for larger studies [12]. Often, the task of manual ROI tracing for an imaging study is shared by several analysts resulting in inter-rater differences between subjects for the same region in the same cohort. As cohort sizes have increased in amyloid PET imaging studies, automated ROI delineation techniques have become more popular [3,6,13], yet the relative performance of these automated techniques in an integrated standardized framework remains unexplored.

The Centiloid Project was developed to standardize quantitative amyloid imaging measures on a 0–100 scale, with this scale being anchored at zero by young controls and 100 by AD patients. One of the major goals in the development of the Centiloid scale was to facilitate direct comparison of results across different analysis methods and tracers [14]. The goal of this work was to examine the use of Centiloid standardization [14] on [^{11}C]PiB SUVR classification. This was accomplished using an existing DS population dataset [8] to provide a comparison between [^{11}C]PiB SUVR outcomes determined from hand-drawn in native space (HD_{NS}) ROI and 3 automated methods of ROI analysis. SUVR threshold values for amyloid positivity have been previously described by our group based on tracing of 6 cortical HD_{NS} ROIs associated with amyloid β deposition in AD [1,12].The hand-drawn native space method was compared with the following automated methods: FreeSurfer [15,16], Statistical Parametric Mapping (SPM) using the Wake Forest University PickAtlas [17] extraction, and a hand-drawn method in PiB template space. Although the Centiloid method specifies a standard cortical + striatum target region and a whole-cerebellum reference region for initial analysis, smaller ROIs are accommodated by either (1) generation of a parametric Centiloid image for sampling smaller ROIs or (2) linear regression [14]. Here, we apply the Centiloid standardization linear regression approach to the global and striatum ROIs to examine its impact on [^{11}C]PiB SUVR classification.

Of all the forms of AD, DS has one of the most homogeneous and best understood initiating events in the overproduction of amyloid β due to 3 copies of chromosome 21 and the *APP* gene present in this chromosome. Adults with DS are uniformly affected by AD pathology by their fourth decade [18–20]. Furthermore, the early striatal pattern of amyloid deposition in DS is similar to that in autosomal dominant AD mutation carriers [21]. Adults with DS in their seventh decade have a 70%–80% chance of developing clinical dementia [22,23]. DS can be viewed in relation to AD as one of amplified sensitivity to risk and protective factors that moderate the relationship between amyloid β, neurodegeneration, and clinical dementia. Thus, DS provides a unique opportunity to study AD.

2. Methods

2.1. Subjects

A total of 83 adults with confirmed DS were recruited as previously described [24]. Participants were assessed for dementia using the Dementia Scale for Down Syndrome [25]. Three individuals who received a cognitive cutoff score > 3 (indicating dementia) were removed from this analysis. Thus, 80 subjects underwent the image processing described in the following.

2.2. Data acquisition

For PET scans, [^{11}C]PiB scans were acquired on Siemens ECAT HR + PET scanners at both sites using a nominal dose of 15 mCi of radiotracer. Preprocessing of dynamic [^{11}C]PiB data was performed in AIR, version 3.0 [26]. Dynamic PET data were corrected for inter frame motion and averaged over 50–70 min after injection. Parametric SUVR images were generated using a cerebellar gray matter ROI. For MRI scans, T1-weighted MRIs were acquired on a 3.0 T GE SIGNA 750 at the University of Wisconsin-Madison site and on a 3.0 T Siemens Magnetom Trio at the University of Pittsburgh Medical Center site. The SIGNA 750 acquisition used high-resolution volumetric spoiled gradient sequence (TI/TE/TR = 450/3.2/8.2 ms, flip angle = 12°, slice thickness = 1 mm no gap, matrix size = 256 × 256 × 156), whereas the Magnetom Trio acquisition used a magnetization prepared rapid acquisition gradient echo sequence (TI/TE/TR = 900/2.98/2300 ms, flip angle = 9°, slice thickness = 1.2 mm, matrix size = 160 × 240 × 256).

2.2.1. ROIs hand-drawn in native space

HD_{NS} ROIs were generated as previously described [1,12]. MR images were manually skull-stripped and reoriented such that the axial image planes were parallel to the anterior-posterior commissure line. [^{11}C]PiB images were registered to skull-stripped MRIs using AIR, version 3.0, and MRI images were resliced to PET resolution.

Manual ROI tracing was performed on skull-stripped MR images in PET native space using ROI Tool software (Siemens Medical Systems, Knoxville, TN). HD_{NS} ROIs included the anterior cingulate gyrus (ACG), anterior ventral striatum (AVS), frontal cortex (FRC), lateral temporal cortex (LTC), parietal cortex (PAR), precuneus cortex (PRC), and cerebellar gray matter. A global region (GBL) was created by generating a voxel-weighted average of the 5 cortical ROIs and the striatal ROI.

2.2.2. ROIs hand-drawn in MNI space

Spatial normalization of standardized uptake value PET images was performed using a DS-specific PET template

created using a 2-pass method, as previously described by our group [24]. The utilization of this PiB-based template allows the inclusion of subjects without an accompanying MRI (e.g., due to the excessive motion artifact). Normalized images were visually inspected and qualitatively assessed on cortical outline and striatal placement. No images were removed because of poor spatial normalization via the DS-specific PET template.

Hand-drawn in MNI (Montreal Neurological Institute) (HD_{MNI}), ROIs were created in ITK-SNAP, version 3.4.0, on a subset of normalized T1 MRIs and combined into a single mask for each ROI to ensure a proper fit despite any differences that may have persisted after spatial normalization. ROI masks were closely inspected for each subject.

2.2.3. ROIs from SPM using PickAtlas and MNI space

PickAtlas and MNI (PA_{MNI}) ROIs were defined from the Talairach Daemon database in MNI space (described previously) provided in the Wake Forest University PickAtlas toolbox in SPM ($_{NS}$.fil.ion.ac.uk/spm/software/). The binary masks were dilated (4 mm Gaussian smoothing and subsequent thresholding at 0.3) to account for differences that may have persisted after spatial normalization and were closely inspected for each subject.

2.2.4. FreeSurfer ROIs in native space

Preprocessed [^{11}C]PiB images were registered and resliced to native-space MRI images using PMOD, version 3.709. FreeSurfer (5.1.0) was used to process native-space MRI images [15,16]. Grey matter–specific ACG, FRC, LTC, PAR, and PRC FreeSurfer ROIs were generated by combining standard FreeSurfer atlas neocortical regions [27]. The Imperial College London Clinical Imaging Centre atlas ventral striatum region [28] was transformed to native-space MRIs using the subject-specific FreeSurfer talairach.m3z transform to generate the native-space FreeSurfer AVS ROI. Cerebellar grey matter was used as reference region to calculate regional and global SUVR values.

FreeSurfer in native space (FS_{NS}) ROIs were manually inspected and edited, where appropriate. Nine subjects failed FreeSurfer processing and were not further included.

2.2.5. Centiloids (CL_{MNI})

The level-1 Centiloid replication analysis was first performed as prescribed [14]. A linear correlation between replication results and the original Centiloid values yielded an R^2 of 0.999.

After the prescribed Level-2 Centiloid method calibration procedure, 50–70 min PiB SUVR images from the DS cohort were registered to corresponding MR images using SPM, version 12 (SPM12). Subject MR images were normalized to MNI space using the unified segmentation method in SPM12 [29], and the resulting forward transformations were applied to registered [^{11}C]PiB PET images. Four subjects failed the segmentation and normalization process, in addition to the 9 subjects who failed FreeSurfer processing, thus excluded from subsequent analyses.

Normalized [^{11}C]PiB images were sampled with the Centiloid Cortex (CTX) ROI (which includes striatum) and whole-cerebellum reference ROI to generate the standard Centiloid CTX SUVR values. A Centiloid-based AVS ROI was also generated by masking out all contiguous voxels in the Centiloid CTX ROI outside the striatum and used to calculate a "nonstandard" Centiloid AVS SUVR. The Centiloid CTX ROI was not further broken into cortical regions, thus only CTX and AVS SUVR values are presented for the Centiloid method and subsequent conversions.

Centiloid calibration for each of the other ROI methods was performed as described in [14]. Briefly, linear regression was performed between each set of nonstandard ROI method GBL SUVRs and the standard set of Centiloid CTX SUVRs across all subjects, generating a slope and intercept between the nonstandard ROI method GBL SUVR and standard Centiloid CTX SUVR, as described by Equation 2.2.3.2a found in the study by Klunk et al. [14]. Subject nonstandard GBL SUVR values were then converted to standard Centiloid CTX SUVR values using the nonstandard ROI method–specific slope and intercept. Subsequently, the converted CTX SUVR values were converted to Centiloid units using Equation 1.3b in the study by Klunk et al. [14].

Method-specific AVS SUVR values were analogously converted to Centiloid units using the same nonstandard ROI method–specific slope and intercept [14], Section 2.2.2 and 3.3.

2.2.6. PiB(+) classification

The average age for the n = 67 adults was 37.4 (standard deviation = 7.19), and there were 33 female and 34 male subjects with 54 (81%) apolipoprotein E (APOE) 4 noncarriers and 10 (15%) APOE4 carriers.

Using methods previously published by our group, we classified subjects as PiB(+) and PiB(−) based on sparse k-means SUVR cutoffs determined from HD_{NS} SUVR values for each of the six composite ROIs and the GBL ROI [30]. These regional cutoff values are method specific and only directly applicable to the HD_{NS} data. However, we applied these cutoffs without conversion across each of the ROI delineation methods, purely for comparison among ROI delineation methods.

The HD_{NS} AVS and GBL SUVR cutoff values were converted to Centiloid SUVR units using the HD_{NS} SUVR-to-Centiloid SUVR regression described previously and then converted to Centiloid units using the standard Centiloid equation Equation 1.3b in the study by Klunk et al. [14]. Centiloid unit cutoffs were then applied to all AVS and GBL Centiloid unit values across all methods.

2.3. Statistical methods

Descriptive statistics were calculated for all measurements (Table 1). Basic frequencies were also performed to

Table 1
Descriptive statistics for each ROI

ROI	HD_{NS} (N = 67)	HD_{MNI} (N = 67)	PA_{MNI} (N = 67)	FS_{NS} (N = 67)	Centiloid SUVR (N = 67)
ACG	1.35 (0.27)	1.30 (0.28)	1.29 (0.27)	1.32 (0.22)	NA
FRC	1.27 (0.28)	1.21 (0.17)	1.19 (0.29)	1.16 (0.25)	NA
PAR	1.29 (0.24)	1.10 (0.21)	1.24 (0.30)	1.30 (0.26)	NA
PRC	1.38 (0.31)	1.45 (0.31)	1.30 (0.31)	1.14 (0.20)	NA
AVS	1.45 (0.51)	1.39 (0.45)	1.34 (0.39)	1.33 (0.38)	1.22 (0.46)
LTC	1.29 (0.23)	1.20 (0.15)	1.21 (0.21)	1.11 (0.18)	NA
GBL	1.31 (0.27)	1.27 (0.22)	1.23 (0.28)	1.17 (0.21)	1.15 (0.23)
AVS* (Centiloid)	23.22 (38.50)	24.49 (40.88)	21.67 (28.49)	29.20 (38.15)	19.68 (43.41)
GBL* (Centiloid)	13.19 (20.20)	13.19 (19.72)	13.19 (20.41)	13.19 (20.95)	13.19 (21.23)

Abbreviations: ROI, region of interest; SUVR, standard uptake value ratio; HD_{NS}, hand-drawn in native space; ACG, anterior cingulate gyrus; AVS, anterior ventral striatum; FRC, frontal cortex; LTC, lateral temporal cortex; PAR, parietal cortex; PRC, precuneus cortex; GBL, global region; HD_{MNI}, hand-drawn in MNI; PA_{MNI}, PickAtlas and MNI; FS_{NS}, FreeSurfer in native space; SD, standard deviation.

NOTE. Data are presented as SUVR mean (SD), except *Striatum and Global, which are presented as Centiloid mean (SD).

compute the percentage of subjects classified as PiB(+) if the hand-drawn prespecified cutoff was used (Table 2). To assess the differences in the ROI SUVR measurements between the four methods, a repeated measures analysis was performed with a fixed factor, ROI method, and a random subject effect to account for within-subject correlation. The Kenward-Rogers method [31] was used for computing the degrees of freedom. The following statistical model was fit for each ROI SUVR (y):

$$y_{ij} = \beta_0 + \beta_1 M_{ij1} + \beta_2 M_{ij2} + \beta_3 M_{ij3} + b_{i0} + \varepsilon_{ij}$$

where:

i. $\beta_0$represents the intercept;
ii. M_{ijk} (k = 1,2,3) is the dummy variable for method factor for the jth observation on the ith subject (i = 1,2,…,67; j = 1,2,3,4) and
   a. M_1 = 1 if method is FreeSurfer, 0 otherwise,
   b. M_2 = 1 if method is HD_{MNI}, 0 otherwise,
   c. M_3 = 1 if method is SPM, 0 otherwise;
iii. jth is the observation on the ith subject (i = 1,2,…,67; j = 1,2,3,4);
iv. b_{i0} is the subject-specific random effect ($b_{0i} \sim N(0, \sigma_0^2)$); and
v. $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ is the random error term.

Table 2
PiB amyloid positivity by ROI and method

ROI	HD_{NS}, N (%)	HD_{MNI}, N (%)	PA_{MNI}, N (%)	FS_{NS}, N (%)
ACG	11 (16.4)	12 (17.9)	11 (16.4)	8 (11.9)
FRC	12 (17.9)	2 (3.0)	11 (16.4)	5 (7.5)
PAR	9 (13.4)	2 (3.0)	12 (17.9)	13 (19.4)
PRC	14 (20.9)	15 (22.4)	13 (19.4)	4 (6.0)
AVS	20 (29.9)	17 (25.4)	15 (22.4)	14 (20.9)
LTC	12 (17.9)	7 (10.5)	11 (16.4)	5 (7.5)
GBL	12 (17.9)	10 (14.9)	12 (17.9)	5 (7.5)
AVS (Centiloid)	20 (29.9)	19 (28.4)	17 (25.4)	21 (31.3)
GBL (Centiloid)	12 (17.9)	13 (19.4)	13 (19.4)	14 (20.9)

Abbreviations: ROI, region of interest; HD_{NS}, hand-drawn in native space; ACG, anterior cingulate gyrus; AVS, anterior ventral striatum; FRC, frontal cortex; LTC, lateral temporal cortex; PAR, parietal cortex; PRC, precuneus cortex; GBL, global region; HD_{MNI}, hand-drawn in MNI; PA_{MNI}, PickAtlas and MNI; FS_{NS}, FreeSurfer in native space.

NOTE. Data are presented as n = PiB(+)/percentage (%).

The β coefficients (β_1, β_2, β_3) represent the difference in SUVR mean estimates between each method examined and the HD_{NS}; β_1 corresponds to the difference in means between HD_{NS} and FS_{NS}, β_2 corresponds to the difference in means between HD_{NS} and HD_{MNI} space, and β_3 corresponds to the difference in means between HD_{NS} and PA_{MNI}.

The statistical model presented previously was used for the analysis of each of the SUVR values for each ROI calculated by the original methods (i.e., before Centiloid conversion). When the Centiloid SUVR was added, for the AVS and GBL regions, another method along with its corresponding parameter were included in the model, $\beta_4 M_{ij4}$, where (β_4)represents the difference in the estimated means between HD_{NS} and Centiloid-translated SUVR measure. Thus, the model implemented for this scenario had the following equation:

$$y_{ij} = \beta_0 + \beta_1 M_{ij1} + \beta_2 M_{ij2} + \beta_3 M_{ij3} + \beta_4 M_{ij4} + b_{i0} + \varepsilon_{ij}$$

where the parameters are the same as described previously.

In addition, we also computed intraclass correlation coefficients (ICC) using a one-way random effects model between HD_{NS} and each of the other methods for each ROI SUVR [32,33]. The ICC was used to quantify the between-method reliability for each ROI. All statistical analyses were performed using SAS 9.3 (SAS Institute, Cary, NC). All statistical tests were two sided and considered significant if the associated P value $< .05$ (95% confidence interval [CI] for the estimated parameter differences excludes 0). No multiple comparison correction was performed because the ROIs were set a priori. However, for the comparisons performed among methods, Tukey-Kramer [34] adjusted 95% CIs are presented in Table 3.

Bland-Altman plots [35] were used to investigate agreement between the HD_{NS} Centiloids translation and the other ROI delineation methods after conversion to Centiloid units (Supplementary Figs. 1 and 2). Bland-Altman plots quantify

Table 3
Repeated measures analysis for each ROI

ROI	HD_{MNI} versus HD_{NS}	PA_{MNI} versus HD_{NS}	FS_{NS} versus HD_{NS}	Centiloid SUVR versus HD_{NS}
ACG	−0.053 (−0.074 to −0.031)	−0.060 (−0.081 to −0.039)	−0.031 (−0.052 to −0.010)	NA
FRC	−0.057 (−0.086 to −0.027)	−0.070 (−0.100 to −0.041)	−0.100 (−0.130 to −0.071)	NA
PAR	−0.186 (−0.216 to −0156)	−0.052 (−0.082 to −0.022)	0.011 (−0.019 to 0.041)	NA
PRC	0.069 (0.041 to 0.097)	−0.075 (−0.103 to −0.047)	−0.238 (−0.266 to −0.210)	NA
AVS	−0.056 (−0.089 to −0.024)	−0.105 (−0.138 to −0.073)	−0.115 (−0.147 to −0.082)	−0.228 (−0.260 to −0.195)
LTC	−0.087 (−0.110 to −0.065)	−0.085 (−0.108 to −0.063)	−0.176 (−0.199 to −0.154)	NA
GBL	−0.046 (−0.065 to −0.027)	−0.086 (−0.106 to −0.067)	−0.142 (−0.161 to −0.123)	−0.163 (−0.183 to −0.144)
AVS* (Centiloid)	1.268 (−1.764 to 4.300)	−1.547 (−4.579 to 1.486)	5.980 (2.948 to 9.012)	−3.536 (−6.569 to −0.504)
GBL* (Centiloid)	0 (−1.415 to 1.415)	0 (−1.415 to 1.415)	0 (−1.415 to 1.415)	0 (−1.415 to 1.415)

Abbreviations: ROI, region of interest; SUVR, standard uptake value ratio; HD_{NS}, hand-drawn in native space; ACG, anterior cingulate gyrus; AVS, anterior ventral striatum; FRC, frontal cortex; LTC, lateral temporal cortex; PAR, parietal cortex; PRC, precuneus cortex; GBL, global region; HD_{MNI}, hand-drawn in MNI; PA_{MNI}, PickAtlas and MNI; FS_{NS}, FreeSurfer in native space.

NOTE. Results from the repeated measures model are represented as mean estimated differences and 95% confidence interval using hand-drawn values as the reference method (*using Centiloid units measurements).

the level of agreement between two quantitative techniques by calculating limits of agreement represented by the difference between the two paired measurements versus the average between them. Its use follows the recommendation that 95% of these measurements lie within ±2 standard deviation of the mean differences.

3. Results

3.1. Descriptive statistics for each ROI

The average SUVR for all six regional ROIs (ACG, AVS, FRC, LTC, PAR, and PRC) and the global ROI was typically higher for HD_{NS} than that for all other methods (Table 1). Exceptions included PAR, which was the highest for FS_{NS}, and PRC, which was highest for HD_{MNI}.

The classification as PiB(+) by using the unconverted HD_{NS} cutoffs across methods was highly variable, showing that FS_{NS} and HD_{MNI} methods typically result in a lower number of subjects classified as PiB(+) when the unconverted HD_{NS} cutoff is used (Table 2). However, the PA_{MNI} method resulted in a similar number of subjects classified as PiB(+) using the ACG, FRC, PAR, and PRC regional ROIs or the GBL ROI. Also, when using the striatum, the HD_{NS} method shows the most subjects classified as PiB(+) compared with all other methods.

As intended, the Centiloid standardization results in very similar numbers of subjects classified as PiB(+), regardless of the normalization and ROI drawing methods used (Table 2).

3.2. Repeated measures model for each ROI

The results of the repeated measures model are presented as estimated mean differences along with the 95% CI between HD_{NS} SUVR values and each of the other methods for each individual ROI (Table 3). Consistent with Table 1, the SUVR values from the HD_{NS} method were typically higher for every region than those of the ROIs generated by automated methods. For example, for the ACG, the estimated SUVR mean difference between the HD_{MNI} and HD_{NS} was equal to −0.053 with a 95% CI of −0.074 to −0.031. This suggests that the range of differences in the mean SUVR values computed using the HD_{NS} method versus the SUVR computed by the HD_{MNI} method, for the ACG region, could range between −0.074 and −0.031. The largest estimated mean differences in SUVR values from the automated FS_{NS} ROIs compared with those from HD_{NS} ROIs were observed in the PRC (−0.238; 95% CI, −0.266 to −0.210) and the LTC (−0.176; 95% CI, −0.199 to −0.154) (Table 3). The estimated differences between the means of the SUVRs standardized to Centiloid units, as expected, are all zero with CIs of 1–3 CL units (Table 3).

3.3. ICC between methods

A high degree of reliability was found between the HD_{NS} method and the PA_{MNI} method for all ROIs (all ICCs are above 0.875; Table 4). A high degree of reliability was observed between the HD_{NS} method and the FS_{NS} method for ACG, FRC, PAR, and AVS (all ICC > 0.85; Table 4) but not for the LTC (ICC = 0.601) or PRC (ICC = 0.505), suggesting only moderate reliability for these two areas (Table 4). The ICC's reliability coefficients between HD_{NS} ROIs and HD_{MNI} were above 0.8 for ACG, PRC, and AVS, suggesting high reliability and moderate reliability for the FRC (ICC = 0.760), LTC (ICC = 0.731), and PAR (ICC = 0.604) (Table 4). The ICCs lower than 0.6 (LTC and PRC; Table 4) have wide CIs suggesting that these methods have higher variability and, in turn, lower reliability coefficients. The ICCs greater than 0.8 (ACG, FRC, PAR, and AVS; Table 4) have narrow CIs, indicating lower variability and higher reliability among the compared methods.

All ICCs were above 0.94 for Centiloid-converted ROIs, suggesting very high reliability when this type of standardization is used across ROI methods.

Table 4
ICC between methods

ROI	ICC (HD_{NS} − HD_{MNI}), (agreement) 95% CI	ICC (HD_{NS} − PA_{MNI}), (agreement) 95% CI	ICC (HD_{NS} − FS_{NS}), (agreement) 95% CI	ICC (HD_{NS} – Centiloid), (agreement) 95% CI
ACG	0.965 (0.944–0.978)	0.959 (0.934–0.974)	0.893 (0.831–0.933)	NA
FRC	0.760 (0.637–0.845)	0.930 (0.889–0.956)	0.856 (0.776–0.909)	NA
PAR	0.604 (0.427–0.736)	0.890 (0.827–0.931)	0.855 (0.775–0.908)	NA
PRC	0.960 (0.936–0.975)	0.954 (0.926–0.971)	0.505 (0.304–0.663)	NA
AVS	0.968 (0.949–0.980)	0.920 (0.871–0.949)	0.893(0.832–0.933)	0.873 (0.802–0.920)
LTC	0.731 (0.597–0.825)	0.875 (0.805–0.921)	0.601 (0.424–0.734)	NA
GBL	0.933 (0.893–0.958)	0.937 (0.899–0.961)	0.767 (0.648–0.850)	0.750 (0.623–0.838)
AVS* (Centiloid)	0.982 (0.971–0.989)	0.939 (0.903–0.962)	0.951 (0.921–0.969)	0.969 (0.950–0.981)
GBL* (Centiloid)	0.973 (0.956–0.983)	0.985 (0.977–0.991)	0.947 (0.915–0.967)	0.951 (0.922–0.970)

Abbreviations: ROI, region of interest; HD_{NS}, hand-drawn in native space; ACG, anterior cingulate gyrus; AVS, anterior ventral striatum; FRC, frontal cortex; LTC, lateral temporal cortex; PAR, parietal cortex; PRC, precuneus cortex; GBL, global region; HD_{MNI}, hand-drawn in MNI; PA_{MNI}, PickAtlas and MNI; FS_{NS}, FreeSurfer in native space; CI, confidence interval.

NOTE. ICCs and their associated 95% confidence interval (*using Centiloid units measurements).

4. Discussion

These data demonstrate how different methods for demarcation of brain ROIs can affect SUVR measurements and resultant classification of amyloid positivity. However, Centiloid standardization diminishes the variability among different methods of demarcation. In this study, we characterized the variability among different ROI demarcation techniques and demonstrated that a simple standardization (or linear scaling) to Centiloid units can greatly reduce the variability across methods and almost eliminate the discrepancies in amyloid-positivity classification. DS is a particularly relevant population to explore these effects, as adults with DS are almost uniformly affected by AD pathology. We compared three different ROI demarcation methods with our previously used HD_{NS} ROI method and their Centiloid translation, where applicable, focusing on (1) differences between the ROI demarcation methods; (2) reliability of the SUVR measures across the methods as compared with the HD_{NS} method; and (3) Centiloid standardization of classification cutoffs for application across ROI methods.

Our results indicate smaller differences in SUVR values between the HD_{NS} and PA_{MNI} methods as well as between the HD_{NS} and the HD_{MNI} methods for ACG, FRC, PRC, and LTC but larger differences for AVS and PAR. In contrast, the differences in SUVR values when the HD_{NS} method is compared with the FS_{NS} method are smaller for ACG, AVS, and PRC and much larger for FRC, PAR, and LTC. This variability in regional differences across methods could prove challenging when exploring longitudinal change in PiB retention or classification of amyloid positivity, either cross-sectionally or longitudinally. However, the Centiloid translation makes the ROI delineation methods comparable with little to no variability.

Even though the differences that we found in mean SUVR values between any two methods are relatively small, they can still influence the PiB classification of individual subjects, particularly those with SUVR values in the area around the cutoff. For example, a difference in SUVR of 0.05 between two ROI demarcation methods can result in a subject being classified differently. Therefore, method-specific cutoffs should be used for amyloid-positivity classification based on the ROI delineation method that was employed.

The PiB(+/−) classifications show very high variability across methods if one method-specific cutoff is used across all the different methods. Because the cutoffs were determined based on a k clustering method [30] and were specifically determined for the HD_{NS}, it is not advisable to use the same cutoff if a different ROI delineation method was used. However, we did find that conversion of method-specific data and cutoffs to Centiloid units reduced PiB(+/−) classification variability to a surprisingly great degree and may allow method-specific cutoffs to be applied in a valid manner to other methods.

It should be noted, though, that application of the Centiloid method resulted in failed normalizations, and like FreeSurfer, the other MRI-based automated method necessitated the exclusion of subjects from the analysis and PiB(+/−) classifications. Both SPM12's segmentation and FreeSurfer were developed based on structural MRI data from populations excluding DS subjects. SPM12's segmentation tissue probability maps were based on normal, healthy subjects (http://brain-development.org/ixi-dataset/), and FreeSurfer's cortical atlas was based on 40 non-DS subjects ranging in age from 19 to 86 years [26]. Relative differences in DS brain anatomy or motion artifacts likely led to the 4 failures with the Centiloid method and 9 failures with FreeSurfer.

Neither of the automated PET template methods, which do not require MRI data (HD_{MNI} and PA_{MNI}), were prone to failure. However, PET template methods can introduce additional biases and variability [36]. Also, the application of anatomically based partial volume correction (PVC) techniques, which are increasingly popular in longitudinal amyloid PET imaging [13,37,38], is not possible with PET template methods that do not contain significant anatomical information. This limitation must be weighed against the advantage of minimizing excluded data resulting from poor-quality MRI scans.

In this study, we did not explore PVC. Additional processing would be necessary to apply commonly used PVC techniques for each of the ROI methodologies with the exception of FreeSurfer, as it is the only method tested with tissue-specific ROIs (i.e., ROIs that do not contain both grey matter and white matter voxels). A major objective of PVC in amyloid PET imaging is the removal of influence of nonspecific binding in white matter [39]. The hand-drawn methods, either in native (HD_{NS}) or template space (HD_{MNI}), the Pick-Atlas method (PA_{MNI}), and the CTX (cortical) ROIs of the standard Centiloid method all contain substantial white matter voxels. However, the FreeSurfer method had the highest rate of failure, so alternative automated MRI-based DS-specific segmentation and ROI delineation techniques should be explored if PVC is to be applied in this population.

These data indicate that both SUVR values and PiB(+/−) classification are highly dependent on the choice of method for ROI demarcation. Small differences in SUVR values for individual ROIs can lead to differences in classification of subjects based on PiB status, especially when many in the population are near the cutoff value. However, conversion of these values to standardized Centiloid units results in far better agreement between methods in terms of both PiB(+/−) classification and SUVR values.

One limitation of this study is the use of different MRI scanners at both acquisition sites; however, such conditions will likely exist for most multisite investigations. Each of the MRI-based ROI methods may perform differently on MRI data acquired across different scanners. An additional limitation of the present method is that we are unable to control for random variations in individuals within the cohort and to variations based on change in imaging equipment over time. However, because the comparisons made between methods are in the same individuals, this is a small concern. Because the Centiloid method does not address variability across subjects or within subjects, only bias across methods, a future area of study should explore additional methods to address within/across subject variability, such as a next-level Centiloid linear regression that does incorporate a noise term.

RESEARCH IN CONTEXT

1. Systematic review: Amyloid PET data is typically quantified using regions of interest delineated on structural MRI images using manual, or more recently, automated methods. However, ROI segmentation on MRI images can be challenging in populations with brain structure abnormalities, such as Alzheimer's disease or Down syndrome, particularly when automated processing routines are employed. Differences in ROI delineation could substantially affect statistical outcomes when quantifying [^{11}C] PiB PET standardized uptake value ratio.
2. Interpretation: These data demonstrate how different methods for demarcation of brain ROIs can affect SUVR measurements and resultant classification of amyloid-positivity. However, Centiloid standardization diminishes the variability among different methods of demarcation. In this study, we characterized the variability among different ROI demarcation techniques and demonstrated that a simple standardization (or linear scaling) to Centiloid units can greatly reduce the variability across methods and almost eliminate the discrepancies in amyloid-positivity classification. DS is a particularly relevant population in which to explore these effects, as adults with DS are almost uniformly affected by AD pathology.
3. Future directions: Because the Centiloid method does not address variability across subjects or within subjects, only bias across methods, a future area of study should be to explore additional methods to address within/across subject variability, such as a next-level Centiloid linear regression that does incorporate a noise term.

Acknowledgments

The authors thank the psychologists, project coordinators, and imaging technologists at the University of Pittsburgh and University of Wisconsin, Madison. The authors also thank the participants and their families for their time and commitment to the research study. 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.; RF1 AG025516 to W.E.K; P01 AG025204 to W.E.K.; P50 AG005133 to Oscar Lopez).

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.dadm.2018.03.006.

References

[1] Cohen AD, Price JC, Weissfeld LA, James J, Rosario BL, Bi W, et al. Basal cerebral metabolism may modulate the cognitive effects of Abeta in mild cognitive impairment: an example of brain reserve. J Neurosci 2009;29:14770–8.
[2] Klunk WE, Engler H, Nordberg A, Wang YM, Blomqvist G, Holt DP, et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol 2004;55:306–19.
[3] Landau SM, Breault C, Joshi AD, Pontecorvo M, Mathis CA, Jagust WJ, et al., Alzheimer's Disease Neuroimaging, I. Amyloid-beta imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. J Nucl Med 2013;54:70–7.
[4] Lopresti BJ, Klunk WE, Mathis CA, Hoge JA, Ziolko SK, Lu X, et al. Simplified quantification of pittsburgh compound B amyloid imaging PET studies: a comparative analysis. J Nucl Med 2005;46:1959–72.

[5] Price JC, Ziolko SK, Weissfeld LA, Klunk WE, Lopresti BJ, Hoge JA, et al. FDG and PIBPET imaging in Alzheimer's disease and mild cognitive impairment. Neuropsychopharmacology 2005;30:S225.
[6] Schwarz CG, Jones DT, Gunter JL, Lowe VJ, Vemuri P, Senjem ML, et al., Alzheimer's Disease Neuroimaging, I. Contributions of imprecision in PET-MRI rigid registration to imprecision in amyloid PET SUVR measurements [published online ahead of print April 22, 2017]. Hum Brain Mapp 2017. https://doi.org/10.1002/hbm.23622.
[7] Su Y, D'Angelo GM, Vlassenko AG, Zhou G, Snyder AZ, Marcus DS, et al. Quantitative analysis of PiB-PET with FreeSurfer ROIs. PLoS One 2013;8:e73377.
[8] Lao PJ, Handen BL, Betthauser TJ, Mihaila I, Hartley SL, Cohen AD, et al. Longitudinal changes in amyloid positron emission tomography and volumetric magnetic resonance imaging in the nondemented Down syndrome population. Alzheimers Dement (Amst) 2017;9:1–9.
[9] Jack CR Jr, Wiste HJ, Weigand SD, Therneau TM, Lowe VJ, Knopman DS, et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. Alzheimers Dement 2017;13:205–16.
[10] Villemagne VL, Pike KE, Chetelat G, Ellis KA, Mulligan RS, Bourgeat P, et al. Longitudinal assessment of Abeta and cognition in aging and Alzheimer disease. Ann Neurol 2011;69:181–92.
[11] Yau WW, Tudorascu DL, McDade EM, Ikonomovic S, James JA, Minhas D, et al. Longitudinal assessment of neuroimaging and clinical markers in autosomal dominant Alzheimer's disease: a prospective cohort study. Lancet Neurol 2015;14:804–13.
[12] Rosario BL, Weissfeld LA, Laymon CM, Mathis CA, Klunk WE, Berginc MD, et al. Inter-rater reliability of manual and automated region-of-interest delineation for PiB PET. Neuroimage 2011; 55:933–41.
[13] Su Y, Blazey TM, Snyder AZ, Raichle ME, Marcus DS, Ances BM, et al., Dominantly Inherited Alzheimer, N. Partial volume correction in quantitative amyloid imaging. Neuroimage 2015;107:55–64.
[14] Klunk WE, Koeppe RA, Price JC, Benzinger TL, Devous MD Sr, Jagust WJ, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement 2015;11:1–15. e11–14.
[15] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341–55.
[16] Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT, et al. Sequence-independent segmentation of magnetic resonance images. Neuroimage 2004;1:S69–84.
[17] Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 2003;19:1233–9.
[18] Hyman BT, Tanzi RE. Amyloid, dementia and Alzheimer's disease. Curr Opin Neurol Neurosurg 1992;5:88–93.
[19] Lemere CA, Blusztajn JK, Yamaguchi H, Wisniewski T, Saido TC, Selkoe DJ. Sequence of deposition of heterogeneous amyloid beta-peptides and APO E in Down syndrome: implications for initial events in amyloid plaque formation. Neurobiol Dis 1996;3:16–32.
[20] Wisniewski KE, Dalton AJ, McLachlan C, Wen GY, Wisniewski HM. Alzheimer's disease in Down's syndrome: clinicopathologic studies. Neurology 1985;35:957–61.
[21] Klunk WE, Price JC, Mathis CA, Tsopelas ND, Lopresti BJ, Ziolko SK, et al. Amyloid deposition begins in the striatum of presenilin-1 mutation carriers from two unrelated pedigrees. J Neurosci 2007;27:6174–84.
[22] Visser FE, Aldenkamp AP, van Huffelen AC, Kuilman M, Overweg J, van Wijk J. Prospective study of the prevalence of Alzheimer-type dementia in institutionalized individuals with Down syndrome. Am J Ment Retard 1997;101:400–12.
[23] Zigman WB, Schupf N, Sersen E, Silverman W. Prevalence of dementia in adults with and without Down syndrome. Am J Ment Retard 1996;100:403–12.
[24] Lao PJ, Betthauser TJ, Hillmer AT, Price JC, Klunk WE, Mihaila I, et al. The effects of normal aging on amyloid-beta deposition in non-demented adults with Down syndrome as imaged by carbon 11-labeled Pittsburgh compound B. Alzheimers Dement 2016; 12:380–90.
[25] Jozsvai E, Hewitt S, Gedye A. Gedye Dementia Scale for Down Syndrome. In: Prasher V, ed. Neuropsychological Assessments of Dementia in Down Syndrome and Intellectual Disabilities. London: Springer; 2018.
[26] Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC. Automated image registration: I. General methods and intrasubject, intramodality validation. J Comput Assist Tomogr 22:139-52.
[27] Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31:968–80.
[28] Tziortzi AC, Searle GE, Tzimopoulou S, Salinas C, Beaver JD, Jenkinson M, et al. Imaging dopamine receptors in humans with [11C]-(+)-PHNO: dissection of D3 signal and anatomy. Neuroimage 2011;54:264–77.
[29] Malone IB, Leung KK, Clegg S, Barnes J, Whitwell JL, Ashburner J, et al. Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance. Neuroimage 2015;104:366–72.
[30] Cohen AD, Mowrey W, Weissfeld LA, Aizenstein HJ, McDade E, Mountz JM, et al. Classification of amyloid-positivity in controls: comparison of visual read and quantitative approaches. Neuroimage 2013;71:207–15.
[31] Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 1997;53:983–97.
[32] McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30.
[33] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420.
[34] Tukey JW. Comparing individual means in the analysis of variance. Biometrics 1949;5:99–114.
[35] Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135–60.
[36] Bourgeat P, Villemagne VL, Dore V, Brown B, Macaulay SL, Martins R, et al. Comparison of MR-less PiB SUVR quantification methods. Neurobiol Aging 2015;36 Suppl 1:S159–66.
[37] Brendel M, Hogenauer M, Delker A, Sauerbeck J, Bartenstein P, Seibyl J, et al., Alzheimer's Disease Neuroimaging, I. Improved longitudinal [(18)F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction. Neuroimage 2015; 108:450–9.
[38] Schwarz CG, Senjem ML, Gunter JL, Tosakulwong N, Weigand SD, Kemp BJ, et al. Optimizing PiB-PET SUVR change-over-time measurement by a large-scale analysis of longitudinal reliability, plausibility, separability, and correlation with MMSE. Neuroimage 2017b; 144:113–27.
[39] Thomas BA, Erlandsson K, Modat M, Thurfjell L, Vandenberghe R, Ourselin S, et al. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. Eur J Nucl Med Mol Imaging 2011;38:1104–19.