Title

β-amyloid PET harmonisation across longitudinal studies: Application to AIBL, ADNI and OASIS3

Authors

Bourgeat, Pierrick
Doré, Vincent
Burnham, Samantha C
et al.

Publication Date

2022-11-01

DOI

10.1016/j.neuroimage.2022.119527

Peer reviewed

HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2022 November 15.

Published in final edited form as:

Neuroimage. 2022 November 15; 262: 119527. doi:10.1016/j.neuroimage.2022.119527.

β-amyloid PET harmonisation across longitudinal studies: Application to AIBL, ADNI and OASIS3

Pierrick Bourgeat[a,*], **Vincent Doré**[a,b], **Samantha C. Burnham**[a], **Tammie Benzinger**[c], **Duygu Tosun**[d,g], **Shenpeng Li**[a], **Manu Goyal**[e], **Pamela LaMontagne**[e], **Liang Jin**[f], **Christopher C Rowe**[b,f], **Michael W. Weiner**[d,g], **John C Morris**[h], **Colin L Masters**[f], **Jurgen Fripp**[a], **Victor L Villemagne**[b,i,1],

Alzheimer's Disease Neuroimaging Initiative, OASIS3, and the AIBL research group

[a]CSIRO Health and Biosecurity, Brisbane, Australia

[b]Department of Molecular Imaging & Therapy, Austin Health, Melbourne, Australia

[c]Knight Alzheimer Disease Research Center, St. Louis, MO, USA

[d]San Francisco Veterans Affairs Medical Center, San Francisco, CA, USA,

[e]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, USA

[f]The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Melbourne, Australia

[*]Corresponding author. Pierrick.bourgeat@csiro.au (P. Bourgeat).

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Disclosures

Chris Rowe has received research grants from Piramal Imaging, GE Healthcare, Cerveau, Astra Zeneca, Biogen. Victor Villemagne is and has been a consultant or paid speaker at sponsored conference sessions for Eli Lilly, Piramal Imaging, Life Molecular Imaging, GE Healthcare, Abbvie, Lundbeck, Shanghai Green Valley Pharmaceutical Co Ltd, IX-ICO, and Hoffmann La Roche. Manu Goyal has stock equity in Moderna, BioNTech and IBM and has been a paid speaker at sponsored conferences funded by the Capital Medical University, Tancheng Talent Office and Shandong Madic Technology Co Ltd. JC Morris is funded by NIH grants # P30 AG066444; P01AG003991; P01AG026276; U19 AG032438; and U19 AG024904. Neither Dr. Morris nor his family owns stock or has equity interest (outside of mutual funds or other externally directed accounts) in any pharmaceutical or biotechnology company.

The AIBL data can be downloaded through LONI after registration at http://adni.loni.usc.edu/category/aibl-study-data/

The ADNI data can be downloaded through LONI after registration at http://adni.loni.usc.edu/data-samples/access-data/

The OASIS3 data can be downloaded through XNAT after registration at https://central.xnat.org/data/projects/OASIS3

The python code used to build the NMF models as well as the NMF models can be downloaded from 10.25919/5f8400a0b6a1e

Credit authorship contribution statement

Pierrick Bourgeat: Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft. **Vincent Doré:** Writing – review & editing, Supervision, Methodology. **Samantha C. Burnham:** Funding acquisition, Writing – review & editing, Supervision, Methodology. **Tammie Benzinger:** Writing – review & editing. **Duygu Tosun:** Writing – review & editing. **Shenpeng Li:** Writing – review & editing. **Manu Goyal:** Writing – review & editing. **Pamela LaMontagne:** Data curation. **Liang Jin:** Project administration. **Christopher C Rowe:** Writing – review & editing. **Michael W. Weiner:** Funding acquisition, Writing – review & editing, Conceptualization. **John C Morris:** Funding acquisition, Writing – review & editing, Conceptualization. **Colin L Masters:** Funding acquisition, Writing – review & editing, Conceptualization. **Jurgen Fripp:** Writing – review & editing, Supervision. **Victor L Villemagne:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119527.

[1]http://www.gaain.org/centiloid-project.

[g]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA

[h]Washington University in St. Louis, St. Louis, MO, USA,

[i]Department of Psychiatry, The University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Introduction: The Centiloid scale was developed to harmonise the quantification of β-amyloid (Aβ) PET images across tracers, scanners, and processing pipelines. However, several groups have reported differences across tracers and scanners even after centiloid conversion. In this study, we aim to evaluate the impact of different pre and post-processing harmonisation steps on the robustness of longitudinal Centiloid data across three large international cohort studies.

Methods: All Aβ PET data in AIBL (N = 3315), ADNI (N = 3442) and OASIS3 (N = 1398) were quantified using the MRI-based Centiloid standard SPM pipeline and the PET-only pipeline CapAIBL. SUVR were converted into Centiloids using each tracer's respective transform. Global Aβ burden from pre-defined target cortical regions in Centiloid units were quantified for both raw PET scans and PET scans smoothed to a uniform 8 mm full width half maximum (FWHM) effective smoothness. For Florbetapir, we assessed the performance of using both the standard Whole Cerebellum (WCb) and a composite white matter (WM)+WCb reference region. Additionally, our recently proposed quantification based on Non-negative Matrix Factorisation (NMF) was applied to all spatially and SUVR normalised images. Correlation with clinical severity measured by the Mini-Mental State Examination (MMSE) and effect size, as well as tracer agreement in ^{11}C-PiB-^{18}F-Florbetapir pairs and longitudinal consistency were evaluated.

Results: The smoothing to a uniform resolution partially reduced longitudinal variability, but did not improve inter-tracer agreement, effect size or correlation with MMSE. Using a Composite reference region for ^{18}F-Florbetapir improved inter-tracer agreement, effect size, correlation with MMSE, and longitudinal consistency. The best results were however obtained when using the NMF method which outperformed all other quantification approaches in all metrics used.

Conclusions: FWHM smoothing has limited impact on longitudinal consistency or outliers. A Composite reference region including subcortical WM should be used for computing both cross-sectional and longitudinal Florbetapir Centiloid. NMF improves Centiloid quantification on all metrics examined.

Keywords

Amyloid PET; Centiloid; Harmonisation

1. Introduction

The Centiloid (CL) scale was developed to harmonise all β−amyloid (Aβ) PET tracer quantification into a single universal scale (Klunk et al., 2015). In this scale, CL = 0 is anchored to group average of young healthy controls, and CL = 100 to group average of mild Alzheimer's disease (AD) patients. While the Centiloid scale was originally only calibrated for ^{11}C-PiB (PiB), it describes a framework where different tracers and methods

could be calibrated. The prescribed quantification pipeline based on SPM has since been calibrated for all F-18 Aβ tracers, namely ^{18}F-Florebetaben (FBB) (Rowe et al., 2017), ^{18}F-NAV4694 (NAV) (Rowe et al., 2016), ^{18}F-Flutemetamol (FLUTE) (Battle et al., 2018) and ^{18}F-Florbetapir (FBP) (Navitsky et al., 2016). These data were then made publicly available [1] so that other quantification approaches could be calibrated. This was later performed using different approaches including a number of MR-based methods such as PMOD (Battle et al., 2018; Hanseeuw et al., 2021), FSL (Battle et al., 2018), FreeSurfer (Royse et al., 2021; Su et al., 2018), and SPM5 (Schwarz et al., 2018), as well as PET-only methods including CapAIBL (Bourgeat et al., 2018). We have also seen non-traditional quantification methods based on image decomposition being also calibrated into Centiloids (Bourgeat et al., 2021).

While the Centiloid scale provides a good framework for harmonising across tracers and processing pipelines, there could still be significant residual non-biological variability, which could be attributable to heterogeneity in data collection, preprocessing framework or preprocessing steps. Such heterogeneity could hide subtle longitudinal changes which are important to improve our understanding of the progression of AD and its risk factors. These could also hamper the detection of small changes in anti-Aβ therapy and clinical trials. It is therefore important to evaluate existing quantification and harmonisation strategies in a large multi-centre datasets to quantify their impact on longitudinal variability of Aβ over time.

One of the main source of variability is the use of different PET scanners and reconstruction methods, which is inevitable in multi-site studies such as AIBL or ADNI. Differences in scanner geometry, underlying technology and reconstruction algorithms can lead to large differences in quantification (Aide et al., 2017; Joshi et al., 2009). Early work on scanner harmonisation was led by the work of Joshi et al. (2009) based on the scan of a Hoffman phantom used to estimate the amount of smoothing required to bring all the data to a uniform resolution. This method has been employed in ADNI as part of their standard pre-processing pipeline for all PET images and is often included in clinical studies and trials. While the initial validation was performed on FDG, its impact on Aβ image quantification acquired on different scanners has not been fully assessed.

The choice of reference region can also impact the reliability of Aβ quantification. While the whole cerebellum (WCb) is the prescribed reference region as it was shown to lead to the highest effect size between young controls and mild AD, its stability over time for each tracer has not been fully assessed. Previous work using the standardised uptake ratio (SUVR) has shown that WCb is suboptimal for FBP in longitudinal studies (Landau et al., 2015) and a composite region of subcortical white matter plus WCb (WM+WCb) led to improved longitudinal consistency and a rate of increase more congruous with quantification obtained using PiB. While including WM in the reference region is believed to improve quantification by counteracting the effects of the WM spilling into the cortical target regions (López-González et al., 2019), there remains concerns with including WM in a reference due its non-specific binding being significantly different from the cortex GM (Fodero-Tavoletti et al., 2009) and its lower tracer uptake in regions of WM injuries (Pietroboni et al., 2022) and demyelination (Moscoso et al., 2022). This composite reference region has been widely used for SUVR quantification, but has only recently been cross-sectionally evaluated for Centiloids (Royse et al., 2021).

Lastly, novel quantification methods which do not rely on predefined regions of interest have been proposed. These methods use image decomposition to separate specific from non-specific binding, as part of the Aβ quantification. These methods all show good correlation with standard CL or SUVR, while improving the separation between Healthy Controls (HC) and AD patients (Pegueroles et al., 2021; Whittington and Gunn, 2019), increasing the correlation with cognitive measures (Liu et al., 2021) and reducing longitudinal variability (Bourgeat et al., 2021; Whittington and Gunn, 2019). These methods include Non-negative Matrix Factorisation (NMF) (Bourgeat et al., 2021), AmyQ (Pegueroles et al., 2021) and Aβ-index (Leuzy et al., 2020) which both rely on a PCA decomposition, Amyloid Load (Amyloid^{IQ}) (Whittington and Gunn, 2019) which uses an image-base regression, and a more recent deep-learning based method which learns to separate the specific from the non-specific binding based on Aβ- scans (Liu et al., 2021). To our knowledge, our previous work on NMF was the only approach to explicitly enforce consistency between the decomposition of each tracer, and attempt to implicitly reduce the variability due to the use of different scanners. Moreover, it was validated on all five Aβ tracers currently in use and assessed in terms of longitudinal consistency in the multi-tracer/multi-scanner AIBL study. The validation however did not assess the effect of the uniform resolution, the importance of the choice in the reference region or its effectiveness in other studies.

Other work on PET harmonisation includes a recent deep learning approach (Shah et al., 2022) which allows to transform an image from an Amyloid tracer (FBP) to another Amyloid tracer (PiB). While this approach showed promising results, a major limitation is the need for a large number of paired scans to train the model ($N = 80$ used in the paper). The ComBat harmonisation method which is widely used in MR scanner harmonisation has also been recently used for FDG PET SUV harmonisation (Orlhac et al., 2022). However, to our knowledge, it has not been evaluated for Amyloid PET harmonisation.

In this work, we aim to assess the impact of smoothing to a uniform resolution, choice of the reference region and choice of the quantification method on the harmonization of the Aβ PET data in three large longitudinal cohorts, namely AIBL, ADNI and OASIS3 as part of the Alzheimer's Dementia Onset and Progression in International Cohorts (ADOPIC) study. We first evaluate the impact of smoothing the PET data to a uniform 8mm resolution. We then look at the stability of the reference region for each tracer and evaluate the impact of the choice of reference region for FBP. Lastly, we compare the quantification using the standard SPM8 pipeline and the more advanced NMF quantification approach. Since not all subjects can undergo an MRI, we also evaluated the impact of all these harmonisation strategies on our PET-only quantification method through CapAIBL, and its NMF extension on the same subset of subjects. We first compared the corresponding Centiloid values cross-sectionally to evaluate their impact on the quantification, before evaluating their consistency in longitudinal data.

2. Methods

2.1. Data

Data used in this study combined three of the largest and publicly available imaging studies in AD, namely AIBL (Ellis et al., 2009), ADNI (Petersen et al., 2010) and OASIS3

(LaMontagne et al., 2019). We extracted all Aβ PET data and corresponding T1W MRI acquired before the 31st of December 2020 in AIBL (N_{images} = 3315, $N_{subjects}$ = 1345), ADNI (N_{images} = 3516, $N_{subjects}$ = 1648) and OASIS3 (N_{images} = 1398, $N_{subjects}$ = 748) for a total of 8229 PET scans from 3741 participants. AIBL Aβ PET scans were acquired using one of five tracers (PiB, FBP, FBB, NAV, FLUTE), ADNI used three (PiB, FBP, FBB) and OASIS3 used two (PiB, FBP). The breakdown of the tracer's distribution is given in Table 1, showing that PIB is the most prevalent tracer in AIBL and OA-SIS, whereas FBP is the most used tracer in ADNI. AIBL has the highest proportion of subjects who were scanned with 2 of more tracers (41%), followed by OASIS (34%) and ADNI (3%). OASIS has the highest proportion of subjects who were scanned on 2 or more scanners (42%), followed by AIBL (37%) and ADNI (18%). When only considering subjects with 3 or more timepoints, OASIS has the highest proportion of subjects who were scanned with 2 or more tracers (79%), followed by AIBL (69%) and ADNI (9%). Similarly, OASIS has the highest proportion of subjects who were scanned on 2 or more scanners (96%), followed by AIBL (57%) and ADNI (42%). PET scans in AIBL were performed using 4 different scanners models, ADNI used 27 and OASIS used 3.

Both AIBL and OASIS had a higher proportion of healthy controls at baseline, whereas ADNI had similar proportion of HC and MCI patients. OASIS has no MCI patients. There was no significant difference in Age at baseline in any diagnostic group between AIBL and ADNI. The HC in OASIS were significantly younger, and AD patients significantly older. There were significant differences in MMSE between subjects in each of the diagnostics groups for each of the 3 studies. The number of imaging timepoints was generally higher in the HC and MCI than in the AD group.

Data used in the preparation of this article were partly obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

In OASIS, 120 subjects were scanned using both PiB and FBP within 7 months (median=8 days, max=6.5months). Since we do not expect significant increase of Aβ retention during this timeframe, this dataset was used to assess the pair-wise correlation between PiB CL and FBP CL. It should also be noted that most pairs were acquired on different scanners, as one of the PET imaging sessions was combined with the MRI visit by using the PET-MRI scanner in order to reduce participant burden and minimize missing data due to missed visits. FBP scans were acquired on 2 scanners (110 on BioGraph mMR, 10 on BioGraph 40) and PiB on 3 (1 on BioGraph mMR, 117 on BioGraph 40 and 2 on ECAT HRplus). Therefore, while we only refer to these 2 datasets by the tracer used in the rest of the manuscript, any difference measured will contain both a tracer and scanner effect, which cannot be easily isolated from each other. For the longitudinal analysis, having scans in such close proximity will artificially increase the error metrics, and is not representative

of the actual timespan between different scans of the same subject in longitudinal studies. Therefore, for the longitudinal analysis, only one of each scan pairs was used. However, as we sought to evaluate our methods in heterogeneous datasets, for each OASIS subject with 2 tracers at the same timepoint, the tracer that was the least represented in all timepoints for a given participant was kept, therefore enforcing a larger variability in tracers used for each subject.

2.2. Image analysis

We evaluated two quantification methods, the SPM-based quantification pipeline, as prescribed by the Centiloid consortium (Klunk et al., 2015), and CapAIBL, a PET-only quantification method which has been previously calibrated to provide Centiloids (Bourgeat et al., 2018). In the SPM-based quantification method, each T1W MR image is affinely registered to a T1 template. It is then segmented into GM, WM and CSF through an iterative expectation maximisation algorithm, which also includes bias field correction, and non-rigid alignment to the template. The corresponding PET image is then rigidly aligned to the T1W image and non-rigidly deformed using the T1 deformation field. Quantification of the PET is performed using the Centiloid masks in the normalised space (Klunk et al., 2015). With CapAIBL, the PET image is first affinely registered to a mean PET template. An adaptive PiB-PET template is optimised to match the pattern of Aβ retention in the image (Bourgeat et al., 2015). The optimal template is then used as a target for the non-rigid registration. Similar to SPM framework, the quantification is performed using the Centiloid masks in the normalised space. CL^{SPM} and CL^{Cap} will be used to refer to the Centiloids computed using the SPM pipeline, or the CapAIBL one.

To further test the stability of each method when using different PET scanners, we evaluate their performances when using raw PET images, compared to PET images which have been smoothed to a uniform point spread function. This is achieved using the methodology of Joshi et al. (2009), which is used in ADNI as part of their standard preprocessing pipeline. It requires the acquisition of a Hoffman phantom on each PET scanner. The scans are co-registered to a digital version of Hoffman phantom, which is smoothed using a 8 mm FWHM Gaussian filter. Each co-registered scan is smoothed with Gaussian filters of increasing FWHM. For each scanner, the FWHM which minimises the difference between the smoothed physical Hoffman and the smoothed digital one is then used to smooth all PET scans acquired on this scanner. This procedure was performed for both AIBL and OASIS using Hoffman phantoms scanned on each of the scanners used in each study. For ADNI, the pre-processed PET scans which follow the exact same preprocessing and are available on the LONI website were used. ^{Raw}CL and ^{Uni}CL will be used to refer to the CL computed using Raw images and images smoothed to a uniform resolution, respectively.

To assess the stability of the reference region, the subset of PET images from AIBL and ADNI which had valid SUV information in their DICOM files were scaled into SUV, so that their reference region mean SUV could be computed (We did not have access to the raw DICOMs for OASIS, and could not use them in this part of the analysis). To assess the impact of the choice of reference region for the FBP scans, two reference regions were evaluated, the whole cerebellum, and a composite reference region, as proposed by Landau

et al. (2015). The composite reference region includes subcortical white matter as well as the whole cerebellum. To minimize the contribution from voxels with the partial volume effects at the grey-white matter boundary, the white matter segmentation is first smoothed using an 8 mm Gaussian kernel and then thresholded at 70% of its maximum to erode the white matter mask away from grey matter (Landau et al., 2015), before being combined with the Centiloid whole cerebellum mask. For SPM, the WM segmentation from each corresponding T1W MR image was used to compute the composite mask. In CapAIBL, the WM segmentation of the T1 template, matching the PET template was used to build the composite mask. This means that for SPM, each scan used a subject-specific composite mask, whereas in CapAIBL, all scans used the same mask. CL_{WCb} and CL_{Comp} will be used to refer to the CL computed using the whole cerebellum as the reference region, or the composite WM region.

Lastly, the recently proposed NMF-based Centiloid quantification (Bourgeat et al., 2021) was evaluated. It relies on a decomposition of each PET image into its specific and non-specific binding components based on a 2 components NMF decomposition. The model used for the decomposition were built on the Centiloid calibration dataset, and the decomposition was performed so that the specific binding components of each tracer would match in the calibration paired data, therefore enforcing consistency across tracers. The method requires the PET images to be spatially normalised to a standard space, and SUVR normalised. While we've previously calibrated using SPM normalised images using WCb, we have here recalibrated the method for FBP images normalised using the Composite reference region, as well as PET images spatially normalised using the PET-only method CapAIBL. We will refer to the SPM and CapAIBL based NMF quantification as $CL^{SPM+NMF}$ and $CL^{CapAIBL+NMF}$.

Each pipeline and reference region were calibrated to the Centiloid scale following the level-2 calibration method described in the original Centiloid paper (Klunk et al., 2015). Since the original Centiloid calibration data from GAAIN do not include Hoffman phantoms, the calibration scans could not be smoothed to a uniform resolution. Therefore, there was no difference in the equations used to convert SUVR into ^{Raw}CL and ^{Uni}CL. Unless specified otherwise, all analysis were performed using all available data from all 3 studies.

2.3. Statistical analysis

2.3.1. Cross-sectional analysis—The effect of the uniform resolution smoothing on the Centiloid quantification compared to the raw data was first assessed cross-sectionally within each pipeline by looking at any bias in the linear equation between the CL values before and after smoothing to a uniform resolution and their correlation assessed using the coefficient of determination. The stability of the reference region SUV for each tracer against time was evaluated using a t-test, while controlling for the effect of multiple scanners. The impact of the reference region on cross-sectional Centiloid value was similarly assessed by looking at any bias in the linear equation and the correlation assessed using the coefficient of determination and ICC.

Using the paired data in OASIS, we also assessed the correlation between PIB and FBP using the coefficient of determination, and the correlation equation to identify any bias. Cohen's Kappa score was used to measure the inter-tracer agreement (PiB vs FBP) when classifying high (Aβ+) and low (Aβ−) scans based on a 20CL threshold.

To verify that the derived CL values are biologically meaningful, the strength of its correlation with MMSE was assessed using the coefficient of determination. The effect size between all baseline HC and AD was assessed using Cohen's *d*.

For all inter-tracer and pre-processing comparison, ICC was also computed to assess agreement.

2.3.2. Longitudinal analysis—For each subject, the rate of change for each method was defined as the slope of the CL value compared to the participant's age at the time of the scan and was reported in CL/year. Following the analysis done in Bourgeat et al. (2021), the longitudinal consistency (which we here define as the expectation that all timepoints follow a similar slope/trend) was first assessed using a linear regression of all available timepoints and measuring the fitting error, assuming the working hypothesis that Aβ accumulation is linear for each subject over the time-course of the study. We also measured the number of outliers, defined as successive timepoints having changes in CL/year larger or smaller than what is observed in 95% of the cases when a single tracer/single scanner is being used. The thresholds were computed using all 3 cohorts, but separately for the Aβ− and Aβ+ groups. Aβ+ was defined based on a threshold of 20 CL on the SPM CL_{WCb} Raw at baseline. Lastly, given that there is no expectations of linearity between the rate of CL change compared to baseline CL, their correlation was measured using the Spearman ρ.

Linear fit and correlations were computed using python's scipy (1.5.4). Cohen's Kappa was computed using python's sklearn (0.22.2). ICC was computed using python's pingouin (0.3.12).

3. Results

3.1. Studies characteristics

Studies and population characteristics are presented in Table 1.

3.2. Cross-sectional comparison

3.2.1. FWHM smoothing—The FWHM smoothing kernel (in mm) for each study was as follow (XY: mean [min,max], Z: mean [min,max]): AIBL (XY: 4.9 [0.5,7.0], Z: 7.1 [4.0,8.0]), ADNI (XY: 4.5 [2.0,6.0], Z: 3.9 [2.0,6.0]), OASIS (XY: 6.2 [5.5,6.5], Z: 6.8 [6.5,7.0]).

The ICC and R^2 between ^{Raw}CL and ^{Uni}CL for the different analysis methods is presented in Fig. 1. The WCb was used as the reference region for all analysis. The ICC between ^{Raw}CL and ^{Uni}CL was high for all quantification methods, and comparable between SPM (ICC = 0.999) and CapAIBL (ICC = 0.995). Using the FWHM smoothing led to an average reduction of CL by 3% when using SPM and 5% when using SPM+NMF compared to using

$^{Raw}CL^{SPM}$. When using CapAIBL, the reduction in CL was more pronounced with 8% with CapAIBL alone, and 7% with CapAIBL+NMF compared to using $^{Raw}CL^{CapAIBL}$.

Since the amplitude of FWHM smoothing is scanner specific, we also examined the variance in correction across scanners for each quantification method, with a smaller variance indicating that the correction has a similar effect on the quantification across all scanners, and a larger variance indicating a large range of effects across scanners. The individual correlations segregated by scanners for each cohort are illustrated in Suppl. Fig. 1. The variance of slopes between ^{Raw}CL and ^{Uni}CL across scanners was significantly smaller ($p < 0.02$) when using SPM ($2.5*10^{-4}$) compared to CapAIBL ($3.2*10^{-4}$), meaning that SPM had less variability between ^{Raw}CL and ^{Uni}CL across scanners. The variance of slopes was significantly higher ($p < 0.007$) using SPM NMF ($8.4*10^{-4}$) compared to SPM. There was no significant difference in the variance of slopes between CapAIBL and CapAIBL NMF.

Lastly, we checked if the smoothing could improve the concordance between different methods, especially given that the PET-only method might be more sensitive to the image appearance than MR-based one. There was however no change in the CL^{SPM} and $CL^{CapAIBL}$ agreement using raw data, or uniformly smoothed ones, with both yielding an ICC = 0.987.

3.3. Reference region

To evaluate the temporal stability of the reference regions, we computed the correlation between the SUV in the reference region and age in the subset of AIBL and ADNI data with valid SUVs. In AIBL, there was no correlation between the WCb SUV and the subject's age when using PiB ($p = 0.56$), NAV ($p = 0.30$) or FLUTE ($p = 0.89$). There was however a significant negative correlation when using FBP ($p = 0.049$). This correlation disappeared when using the composite WM+WCb SUV ($p = 0.22$). In ADNI, there was no correlation between the WCb SUV and the subject's age in FBB ($p = 0.88$), but there was a significant negative correlation in FBP ($p = 4*10^{-10}$). The correlation was reduced but remained significant when using the composite WM+WCb SUV ($p = 3*10^{-9}$). The scatter plots of SUV vs age are presented in Suppl. Fig. 2 for WCb and Suppl. Fig. 3 for the composite WM+WCb.

The correlation between the CL_{WCb} and CL_{Comp} for the different analysis methods is presented in Fig. 2. The uniform images (8mm FWHM) were used for the analysis. While the ICC between CL_{WCb} and CL_{Comp} was high for all analysis methods, the agreement was much higher (ICC > 0.98) when using NMF, meaning that the NMF-based quantification appears to be more robust to the choice of reference regions.

3.4. Head-to-head PiB-FBP comparison

The scatter plots comparing the PiB CL and their matching FBP CL in the OASIS pairs are presented in Fig. 3 and the ICC between PIB and FBP for each method is presented in suppl Table 1. It shows a strong bias when using SPM or CapAIBL, with FPB CL being overestimated compared to PiB CL. Using the NMF reduces the bias and improves the agreement with a higher ICC. The agreement between PiB and FBP for the classification into a Aβ+ (≥ 20CL) and Aβ− scan (< 20CL) was assessed using the Cohen's Kappa coefficient for each method and presented in Table 2. Using the SPM and CapAIBL

quantification methods, there was a greater agreement between PiB and FBP when FBP was normalised using the composite WM+WCb reference region. Using the uniform resolution smoothing, however, did not improve the agreement compared to using the raw data (shown in Suppl Fig. 4). The highest agreements were obtained using the NMF approach, which were systematically higher than their baseline methods. When using NMF, the choice of reference region had negligible effects on the agreement between PiB and FBP.

3.5. Correlation with MMSE and effect size

Using all subjects at baseline, we measured the correlation of CL with MMSE using the coefficient of determination (Table 3). There was no clear trend showing that the uniform smoothing improved the correlation. The correlation was however much stronger when using the composite WM+WCb reference region, and the NMF systematically improved the correlation compared to its baseline method. Similarly, we also computed the effect size between HC and AD at baseline (Table 4), leading to the same findings. Similar trends were observed when the analysis was conducted in each cohort separately (supplementary Tables 2 and 3).

3.6. Longitudinal comparison

3.6.1. Fitting error and number of outliers—In the Aβ−, 95% of the changes between consecutive pairs of scans acquired on the same scanner and using the same tracer were between −6.33 and 8CL/Y. In the Aβ+, those were between −16.6 and 20.13CL/Y.

The percentage of outliers in the whole population, including participants with a change of scanner and/or tracer, showing changes outside that range in the Aβ− and Aβ+ are presented in Tables 5 and 6, respectively. For all quantification methods, using images smoothed to a uniform 8mm resolution led to a systematic reduction of outliers compared to using the raw data. With both SPM and CapAIBL, using the composite WM+WCb reference region for FBP also led to a systematic reduction of outliers compared to using the WCb. This was also the case in the Aβ+ group when using the NMF. However, in the Aβ− group, the NMF gave the lowest number of outliers when the WCb was used. Overall, using the NMF led to a systematic reduction in the number of outliers in both groups, compared to their baseline method. Similar results were obtained with the mean standard error of the estimated slopes, with tables shown in Suppl. Tables 4 and 5. This reduction of outliers when using the NMF is illustrated in Suppl. Figs. 5 and 6 showing the longitudinal plots of Centiloid value against age for both SPM and CapAIBL, respectively.

3.7. Rate of change

The rate of CL change vs baseline CL for each method, as well as the corresponding Spearman correlation coefficients are shown in Fig. 4. The effect of the uniform smoothing on the correlation was negligeable (Suppl. Fig.7). The correlation with SPM and CapAIBL were stronger using the composite WM+WCb reference region for FBP, compared to using WCb. The correlations were the strongest using NMF, regardless of the pre-processing method or quantification pipeline used. The correlation using CapAIBL and CapAIBL+NMF were generally stronger than those obtained using SPM and SPM+NMF.

4. Discussion

In this paper, we have presented a comparison of different pre- and post-processing techniques applied for improving CL harmonisation. We assessed the use of FWHM resolution which was originally proposed to reduce inter-scanner differences in multi-centre studies, and later implemented in the default ADNI pre-processing pipeline. We then compared the use of different reference regions for FBP, deviating from the standard Centiloid protocol, but more in line with studies showing that the prescribed WCb reference region for Centiloid might not be adequate to observe longitudinal changes. These different pre-processing and normalisation were assessed with both the recommended SPM pipeline, and a PET-only quantification method that we previously calibrated to Centiloids. Lastly, our recently proposed NMF method, which was previously shown to improve longitudinal consistency in AIBL was evaluated on both pipelines. We will discuss each of these assessments, before providing overall recommendations and limitations of this study.

4.1. Uniform FWHM resolution

Smoothing to a uniform FWHM resolution was originally proposed for FDG (Joshi et al., 2009). While the authors showed a 20–50% reduction of variability across scanners on phantom data, the results on real subjects were a lot more modest, with only 15–25% reduction of variability. Given that we lack same tracer, head-to-head comparison on different scanners, it can be hard to assess how much improvement the smoothing brings to the CL quantification. It is however useful to quantify the effect of the smoothing to uniform resolution on the CL quantification. In our cross-sectional analysis, the effect was modest, with only 3% difference with SPM and 8% with CapAIBL. The difference between the 2 methods can be explained by the method used for the spatial normalisation. With SPM, the extra smoothing will have little to no impact on the accuracy of the co-registration between the PET and MRI, and therefore, most of the differences compared to using the raw data can be attributed to the change in signal intensity on the PET due to the extra smoothing. Since CapAIBL uses the PET directly for the non-linear registration to the template, it is more susceptible to biases due to changes in the PET appearance. As a result, the larger difference between using the raw and smoothed data can be attributed to both different errors in the registration as well as the differences in PET intensity. This was further illustrated by looking at the variance of the slopes between different scanners when comparing the CL computed before and after smoothing to uniform resolution for a given method. This variance was significantly higher with CapAIBL than SPM, indicating that when using CapAIBL, the CL quantification using raw data had a lot more variability across scanners compared to using raw data with SPM. This would indicate that PET-only quantification methods, such as CapAIBL could benefit from the FWHM smoothing to reduce variability in the spatial normalisation, whereas MR-based techniques, such as SPM, might not get as much of a benefit from it. It should however be noted that we did not observe any improvement in the agreement between SPM and CapAIBL when using raw or smoothed data, so while the smoothing had a greater effect on CapAIBL, it did not necessarily translate into a more accurate quantification.

In the head-to-head PIB-FBP comparison where 2 different scanners are used, the smoothing to a uniform resolution did not improve the agreement between the tracers, with similar ICC and bias obtained when comparing the raw PiB to the raw FBP, and the uniform PiB and uniform FBP. It did however modify the agreement between the two tracers for classifying Aβ+ from Aβ− scans based on a 20CL threshold, although there was no systematic trend, with some quantification methods leading to better agreement using the raw data. It should be noted that this head-to-head dataset is not optimal to evaluate the effect of smoothing to a uniform resolution, given that the 2 scanners use different technology, and MR-based attenuation has been previously shown to lead to an underestimation of SUVR compared to using a CT-based attenuation map (Su et al., 2016), which is independent of the scanner resolution.

The correlation with MMSE and effect size between HC and AD did not improve with the uniform smoothing, and while the differences were small, the results were often worse compared to using the raw data. It is therefore possible that the extra smoothing might reduce small changes, resulting in weaker correlations.

In the longitudinal analysis, where 56% of the subjects were scanned with 2 or more scanners, while the uniform resolution led to a reduction in both the number of outliers and in the standard error of the estimated slopes, it did not increase the correlation between the rate of change and the baseline CL. This is likely because we only used subjects with three or more timepoints in this analysis, with the linear regression, used to compute the rate of change, smoothing out the effects of outliers. The smoothing might have had a bigger impact if we had included subjects with only two timepoints.

4.2. Reference region

The correlation of the WCb SUV with age revealed that the WCb was stable over time for PiB, NAV, FBB and FLUTE, and therefore suitable to be used as a reference region. It also confirmed that it was not stable for FBP. The composite WM+WCb, however, was stable for FBP in AIBL. In ADNI, while it reduced the strength of the correlation, it remained significantly correlated. The disparity of results between AIBL and ADNI could be explained by the number of scanners being used. While AIBL used only 3 scanners to image FBP, 27 different scanner models have been used in ADNI, which could confound some of these effects since SUV can be dependent on the scanner used. It could also indicate that some age effects are still present in the composite reference region. Nevertheless, those results indicate that the composite reference is more stable over time, and therefore more suitable than WCb for FBP normalisation.

The choice of reference region had a strong impact on the CL quantification of FBP images, with the ICC between CL_{WCb} and CL_{Comp} being only 0.92 for both CapAIBL and SPM. The ICC was much higher (~0.98) when using NMF, indicating that NMF is quite robust irrespective of the choice of reference region. This is expected as the NMF model is fitted to the entire image and will therefore suffer less bias due to the intensity normalisation method.

In the head-to-head comparison, the use of CL_{Comp} for FBP did not reduce the bias, but improved the agreement between PiB and FBP, with higher ICC when using the standard

SPM or CapAIBL quantification pipeline. It also improved the agreement between both tracers in classifying Aβ+ from Aβ−. There was also a systematic improvement in the correlation of CL with MMSE when using CL_{Comp} compared to CL_{Raw}, as well as an increase in the effect size between HC and AD. These results indicate that using the composite WM+WCb reference region might improve the accuracy of FBP quantification in cross-sectional analysis.

In the longitudinal analysis, the results were in line with previous reports (Landau et al., 2015), showing that the use of the composite WM+WCb reference region generally reduced the number of outliers and the fitting error, especially in the Aβ+, as well as increasing the correlation between the rate of change and baseline CL.

Given the existing concerns with regards to using a reference region containing WM, we conducted further analysis testing a GM reference region using the cerebellum cortex (Cb). These results showed that the FBP SUV in the Cb was significantly correlated with age in both AIBL and ADNI (Suppl. Fig. 8). Using the Cb also led to a worse ICC in the head-to-head comparison compared to using WCb (Suppl. Fig. 9). In the longitudinal analysis, it also led to a larger number of outliers (Suppl. Tables 6 and 7) and worse Spearman correlation when comparing baseline CL against its rate of change (Suppl. Fig. 10).

4.3. Quantification methods

In all cross-sectional analysis, the results obtained using both SPM and CapAIBL were often comparable, with no quantification pipeline clearly outperforming the other. Neither quantification pipeline showed a strong benefit from the uniform resolution smoothing, while both showed a benefit from the use of the composite WM+WCb reference region for FBP. In the longitudinal analysis, while CapAIBL had fewer outliers in the Aβ−, SPM had fewer outliers in the Aβ+. This is likely due to the CapAIBL adaptive atlas only containing healthy controls, which might limit its ability to properly model AD cases with high CL values and lead to sub-optimal spatial normalisation. This was however not reflected in the correlation of baseline CL against its rate of change where CapAIBL generally had a higher correlation compared to SPM.

In all experiments, both cross-sectionally and longitudinally, the NMF systematically outperformed its baseline method. In the cross-sectional analysis, it led to the highest ICC between PiB and FBP in the head-to-head comparison, and the highest inter-tracer agreement when classifying Aβ+ from Aβ−. It also led to the strongest correlation with MMSE and highest effect-size between HC and AD. In the longitudinal analysis it also had the lowest number of outliers, and the strongest correlation between baseline CL against its rate of change. While there were small differences between SPM-NMF and CapAIBL-NMF, both versions performed similarly well.

4.4. Recommendations

These results indicate that while the smoothing to a uniform resolution can reduce the number of outliers in longitudinal studies, its impact on harmonisation appears to be quite limited, and in some cases detrimental to some metrics. Because of the overhead involved with acquiring a Hoffman phantom and smoothing the data, we do not consider smoothing

the images to a uniform resolution as a strong requirement for longitudinal studies. While this statement is valid for the studies considered, it should be noted that such advice might differ with the introduction of high-resolution scanners such as the Siemens Biograph Vision PET/CT, where significant differences in resolution and partial volume effect may have a stronger impact on the quantification. It should also be noted that the Centiloid neocortical mask is quite large and includes a large proportion of partial volume voxels. The results might therefore be different if a MR-based parcellation was used to define the neocortical mask, as it might be more susceptible to partial volume effects.

While previous studies have only recommended the use of the composite reference region for longitudinal studies using FBP, these results indicate that it also improves agreement with PIB in the head-to-head study, improves correlation with MMSE and increase the HC-AD effect size. The longitudinal analysis also confirmed that it reduces the number of outliers, decreases the fitting error and improves the correlation between baseline CL and its rate of change. These results indicate that the composite reference region should be used to normalise FBP images not only in longitudinal, but also in cross-sectional analysis when using SPM or CapAIBL. It should however be noted that the results presented in this study were obtained without partial volume correction (PVC) and recent work indicate that PVC could improve FBP quantification when using the Cb or WCb (López-González et al., 2019). Therefore, our recommendation does not apply to methods that use partial volume correction. When using the NMF, there was no systematic benefit from using the composite reference region.

In this study, SPM and CapAIBL had similar performances both cross-sectionally and longitudinally. Since CapAIBL does not need a matching MRI to perform the quantification, it can be run on a larger set of data in studies where the MR is missing, and therefore could become the preferred analysis method since it allows an increase in the number of images that can be quantified, especially in AIBL where 20% of the subjects were unable to undergo an MRI. The NMF proved to be more versatile tool as it could improve the quantification of both CapAIBL and SPM on all the metrics used both cross-sectionally and longitudinally. We would therefore recommend using this method for any future analysis relying on SPM or CapAIBL.

The NMF code and models used in this study are available at 10.25919/5f8400a0b6a1e.

4.5. Limitations

While we looked at reducing the effect of different PET scanner resolution by smoothing the images to a lower resolution, we did not investigate how PVC could be used to achieve a similar goal. While smoothing to a uniform resolution is a fairly standard procedure, there is a wide range of techniques for PVC which can lead to quite different quantification results (Schwarz et al., 2019). PVC would also preclude the use of NMF in our study's framework, as it would require perfect matching of the cortical GM across patients, which the current pipeline based on SPM does not provide. Therefore, the potential gains from using PVC would need to outperform the clear benefits that we've demonstrated by using NMF. While such comparison would be valuable, it is outside the scope of this paper.

Another limitation of our evaluation is that we used the same Centiloid transforms for both raw and uniformly FWHM smoothed PET data, which could introduce a bias in the analysis as the transforms derived from the raw calibration data are not optimal for the uniformly smoothed data. Deriving a new transform for uniformly smoothed data is not possible using the existing calibration dataset as they do not have phantom data. An alternative could be to use an external dataset to recalibrate the Centiloid, but this would require a large number of paired scans for all tracer which is currently not available in our study. The application of the FWHM smoothing was also performed uniformly throughout the brain when the resolution is known to vary across the field of view and depending on the type of reconstruction used. Future studies should seek to estimate and apply spatially varying image smoothing which could improve the accuracy of the uniform resolution harmonization step.

We also did not investigate the use of different reference regions for the other tracers, noting that for quantification using SUVR, the cerebellar cortex is typically the prescribed reference region for PiB, NAV and FBB, and the pons for FLUTE. There is however little literature indicating the inadequacy of using the whole cerebellum for these tracers, compared to the well documented issues with longitudinal FBP, and our analysis of the stability of the SUV in the reference region over time supports these conclusions. That said, one interesting finding from the current study was to show that the NMF was relatively robust irrespective of the choice of reference region, and while it was only tested on FBP, and only two reference regions were compared, we do expect these results to generalise to other tracers and reference regions. This would however need to be confirmed in further studies.

Similarly to our previous work (Bourgeat et al., 2021), our longitudinal validation relies on the assumption that Aβ accumulation is linear over a period < 10 years, when the accumulation is believed to follow trajectory close to a sigmoid (Villemagne et al., 2013). However, half of the participants had their last timepoints within 3.3 years for AIBL, 3.9 for ADNI and 5.0 for OASIS, a fairly short timeframe where changes can be approximated as linear. For participants scanned over a longer period of time, 54% of AIBL participants, 43% of ADNI and 68% of OASIS had a CL remaining bellow 10, meaning that they had very little changes over time.

Lastly, it should be noted that all our validation experiments rely on surrogate markers, and while NMF improves on all of them, it does not necessarily mean that the method is more accurate. Further evaluation of all quantification methods using actual ground truth data such as autopsy (although this is not viable in large studies), phantoms (although those are often unrealistic) and Monte Carlo simulations (López-González et al,. 2019; Paredes-Pacheco et al., 2021) is therefore warranted. We have also limited this analysis to two quantification pipelines, which was again done for the sake of clarity. More quantification pipeline could be included in further studies now that the impact of the pre-processing steps has been clarified.

5. Conclusions

With the availability of large imaging datasets, data harmonisation has become an important topic not only for combining multiple studies, but also to ensure that the findings can be replicated in the clinic where different PET tracers and scanners might be used. In this study, we quantified the impact that each pre-processing step can have on the final PET quantification, and its consistency over time. We also compared two state of the art PET quantification methods and demonstrated that NMF can further reduce inter-tracer differences, improve concordance with cognitive measures and separation between HC and AD as well as reduce variability over time. These improvements will help detect smaller variations in the dynamics of Aβ accumulation and better relate those to genetic, lifestyle and cognitive differences, leading to a better understanding of the progression of AD and its risk factors. Improving the detection of small changes of Aβ over time, will improve the sensitivity to detect the effects of anti-Aβ therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Health and Medical Research Council [GA16788] and the National Institutes of Health [R01-AG058676-01A1].

Data collection and sharing for this project was funded in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data were provided in part by OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

References

Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R, 2017. EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. Eur. J. Nucl. Med. Mol. Imaging 44, 17–31. doi:10.1007/s00259-017-3740-2.

Battle MR, Pillay LC, Lowe VJ, Knopman D, Kemp B, Rowe CC, Doré V, Villemagne VL, Buckley CJ, 2018. Centiloid scaling for quantification of brain amyloid with [18F]flutemetamol using multiple processing methods. EJNMMI Res. 8, 107. doi:10.1186/s13550-018-0456-7. [PubMed: 30519791]

Bourgeat P, Bourgeat P, Dore V, Doecke J, Ames D, Masters C, Rowe C, Fripp J, Villemagne V, 2021. Non-negative matrix factorization improves Centiloid robustness in longitudinal studies. Neuroimage 226, 117593. doi:10.1016/j.neuroimage.2020.117593. [PubMed: 33248259]

Bourgeat P, Doré V, Fripp J, Ames D, Masters CL, Salvado O, Villemagne VL, Rowe CC, 2018. Implementing the Centiloid transformation for 11C-PiB and β-amyloid 18F-PET tracers using CapAIBL. NeuroImage 183, 387–393. doi:10.1016/j.neuroimage.2018.08.044. [PubMed: 30130643]

Bourgeat P, Villemagne VL, Dore V, Brown B, Macaulay SL, Martins R, Masters CL, Ames D, Ellis K, Rowe CC, Salvado O, Fripp J, 2015. Comparison of MR-less PiB SUVR quantification methods. Neurobiol. Aging 36 (1), S159–S166. doi:10.1016/j.neurobiolaging.2014.04.033, Suppl. [PubMed: 25257985]

Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P, Masters C, Milner A, Pike K, Rowe C, Savage G, Szoeke C, Taddei K, Villemagne V, Woodward M, Ames D, 2009. The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int. Psychogeriatr. IPA 21, 672–687. doi:10.1017/S1041610209009405.

Fodero-Tavoletti MT, Rowe CC, McLean CA, Leone L, Li QX, Masters CL, Cappai R, Villemagne VL, 2009. Characterization of PiB binding to white matter in Alzheimer disease and other dementias. J. Nucl. Med. 50, 198–204. doi:10.2967/jnumed.108.057984, Off. Publ. Soc. Nucl. Med.. [PubMed: 19164220]

Hanseeuw BJ, Malotaux V, Dricot L, Quenon L, Sznajer Y, Cerman J, Woodard JL, Buckley C, Farrar G, Ivanoiu A, Lhofel R, 2021. Defining a Centiloid scale threshold predicting long-term progression to dementia in patients attending the memory clinic: an [18F] flutemetamol amyloid PET study. Eur. J. Nucl. Med. Mol. Imaging 48, 302–310. doi:10.1007/s00259-020-04942-4. [PubMed: 32601802]

Joshi A, Koeppe RA, Fessler JA, 2009. Reducing between scanner differences in multi-center PET studies. NeuroImage 46, 154–159. doi:10.1016/j.neuroimage.2009.01.057. [PubMed: 19457369]

Klunk WE, Koeppe RA, Price JC, Benzinger T, Devous MD, Jagust W, Johnson K, Mathis CA, Minhas D, Pontecorvo MJ, Rowe CC, Skovronsky D, Mintun M, 2015. The Centiloid project: standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement. J. 11, 1–15. doi:10.1016/j.jalz.2014.07.003, Alzheimers Assoc.e4.

LaMontagne PJ, Benzinger TL, Morris JC, Keefe S, Hornbeck R, Xiong C, Grant E, Hassenstab J, Moulder K, Vlassenko AG, Raichle ME, Cruchaga C, Marcus D, 2019. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medRxiv 2019.12.13.19014902. 10.1101/2019.12.13.19014902

Landau SM, Fero A, Baker SL, Koeppe R, Mintun M, Chen K, Reiman EM, Jagust WJ, 2015. Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. J. Nucl. Med. 56, 567–574. doi:10.2967/jnumed.114.148981, Off. Publ. Soc. Nucl. Med.. [PubMed: 25745095]

Leuzy A, Lilja J, Buckley CJ, Ossenkoppele R, Palmqvist S, Battle M, Farrar G, Thal DR, Janelidze S, Stomrud E, Strandberg O, Smith R, Hansson O, 2020. Derivation and utility of an Aβ-PET pathology accumulation index to estimate Aβ load. Neurology 95, e2834–e2844. doi:10.1212/WNL.0000000000011031. [PubMed: 33077542]

Liu H, Nai YH, Saridin F, Tanaka T, O' Doherty J, Hilal S, Gyanwali B, Chen CP, Robins EG, Reilhac A, 2021. Improved amyloid burden quantification with nonspecific estimates using deep learning. Eur. J. Nucl. Med. Mol. Imaging 48, 1842–1853. doi:10.1007/s00259-020-05131-z. [PubMed: 33415430]

López-González FJ, Moscoso A, Efthimiou N, Fernández-Ferreiro A, Piñeiro-Fiel M, Archibald SJ, Aguiar P, Silva-Rodríguez J, 2019. Spill-in counts in the quantification of 18F-florbetapir on Aβ-negative subjects: the effect of including white matter in the reference region. EJNMMI Phys. 6, 27. doi:10.1186/s40658-019-0258-7. [PubMed: 31858289]

Moscoso A, Silva-Rodríguez J, Aldrey JM, Cortés J, Pías-Peleteiro JM, Ruibal Á, Aguiar P, 2022. 18F-florbetapir PET as a marker of myelin integrity across the Alzheimer's disease spectrum. Eur. J. Nucl. Med. Mol. Imaging 49, 1242–1253. doi:10.1007/s00259-021-05493-y. [PubMed: 34581847]

Navitsky M, Joshi AD, Devous MD, Pontecorvo MJ, Lu M, Klunk WE, Rowe CC, Wong DF, Mintun MA, 2016. Conversion of amyloid quantitation with florbetapir SUVR to the Centiloid scale. Alzheimers Dement. J. 12, P25–P26. doi:10.1016/j.jalz.2016.06.032, Alzheimers Assoc..

Orlhac F, Eertink JJ, Cottereau AS, Zijlstra JM, Thieblemont C, Meignan M, Boellaard R, Buvat I, 2022. A guide to combat harmonization of imaging biomarkers in multicenter studies. J. Nucl. Med. 63, 172–179. doi:10.2967/jnumed.121.262464, Off. Publ. Soc. Nucl. Med.. [PubMed: 34531263]

Paredes-Pacheco J, López-González FJ, Silva-Rodríguez J, Efthimiou N, Niñerola-Baizán A, Ruibal Á, Roé-Vellvé N, Aguiar P, 2021. SimPET—an open online platform for the Monte Carlo simulation of realistic brain PET data. Validation for 18F-FDG scans. Med. Phys 48, 2482–2493. doi:10.1002/mp.14838. [PubMed: 33713354]

Pegueroles J, Montal V, Bejanin A, Vilaplana E, Aranha M, Santos-Santos MA, Alcolea D, Carrió I, Camacho V, Blesa R, Lleó A, Fortea J, 2021. AMYQ: an index to standardize quantitative amyloid load across PET tracers. Alzheimers Dement. doi:10.1002/alz.12317, n/a.

Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR, Jagust WJ, Shaw LM, Toga AW, Trojanowski JQ, Weiner MW, 2010. Alzheimer's disease neuroimaging initiative (ADNI). Neurology 74, 201–209. doi:10.1212/WNL.0b013e3181cb3e25. [PubMed: 20042704]

Pietroboni AM, Colombi A, Carandini T, Sacchi L, Fenoglio C, Marotta G, Arighi A, De Riz MA, Fumagalli GG, Castellani M, Bozzali M, Scarpini E, Galimberti D, 2022. Amyloid PET imaging and dementias: potential applications in detecting and quantifying early white matter damage. Alzheimers Res. Ther. 14, 33. doi:10.1186/s13195-021-00933-1. [PubMed: 35151361]

Rowe CC, Doré V, Jones G, Baxendale D, Mulligan RS, Bullich S, Stephens AW, Santi SD, Masters CL, Dinkelborg L, Villemagne VL, 2017. 18F-florbetaben PET beta-amyloid binding expressed in Centiloids. Eur. J. Nucl. Med. Mol. Imaging 44, 2053–2059. doi:10.1007/s00259-017-3749-6. [PubMed: 28643043]

Rowe CC, Jones G, Doré V, Pejoska S, Margison L, Mulligan RS, Chan JG, Young K, Villemagne VL, 2016. Standardized expression of ^{18}F-NAV4694 and ^{11}C-PiB β-amyloid pet results with the Centiloid scale. J. Nucl. Med. 57, 1233–1237. doi:10.2967/jnumed.115.171595. [PubMed: 26912446]

Royse SK, Minhas DS, Lopresti BJ, Murphy A, Ward T, Koeppe RA, Bullich S, DeSanti S, Jagust WJ, Landau SM, 2021. Validation of amyloid PET positivity thresholds in Centiloids: a multisite PET study approach. Alzheimers Res. Ther. 13, 99. doi:10.1186/s13195-021-00836-1. [PubMed: 33971965]

Schwarz CG, Gunter JL, Lowe VJ, Weigand S, Vemuri P, Senjem ML, Petersen RC, Knopman DS, Jack CR, 2019. A comparison of partial volume correction techniques for measuring change in serial amyloid PET SUVR. J. Alzheimers Dis. 67, 181–195. doi:10.3233/JAD-180749. [PubMed: 30475770]

Schwarz CG, Tosakulwong N, Senjem ML, Gunter JL, Therneau TM, Vemuri P, Lowe VJ, Jack CR, 2018. Considerations for performing level-2 Centiloid transformations for amyloid PET SUVR values. Sci. Rep. 8, 7421. doi:10.1038/s41598-018-25459-9. [PubMed: 29743601]

Shah J, Gao F, Li B, Ghisays V, Luo J, Chen Y, Lee W, Zhou Y, Benzinger TLS, Reiman EM, Chen K, Su Y, Wu T, 2022. Deep residual inception encoder-decoder network for amyloid PET harmonization. Alzheimers Dement. doi:10.1002/alz.12564, n/a.

Su Y, Flores S, Hornbeck RC, Speidel B, Vlassenko AG, Gordon BA, Koeppe RA, Klunk WE, Xiong C, Morris JC, Benzinger TLS, 2018. Utilizing the Centiloid scale in cross-sectional and longitudinal PiB PET studies. NeuroImage Clin. 19, 406–416. doi:10.1016/j.nicl.2018.04.022. [PubMed: 30035025]

Su Y, Rubin BB, McConathy J, Laforest R, Qi J, Sharma A, Priatna A, Benzinger TLS, 2016. Impact of MR-based attenuation correction on neurologic PET studies. J. Nucl. Med. 57, 913–917. doi:10.2967/jnumed.115.164822. [PubMed: 26823562]

Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, Szoeke C, Macaulay SL, Martins R, Maruff P, Ames D, Rowe CC, Masters CL, 2013. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol. 12, 357–367. doi:10.1016/S1474-4422(13)70044-9. [PubMed: 23477989]

Whittington A, Gunn RN, 2019. Amyloid load: a more sensitive biomarker for amyloid imaging. J. Nucl. Med. 60, 536–540. doi:10.2967/jnumed.118.210518. [PubMed: 30190305]

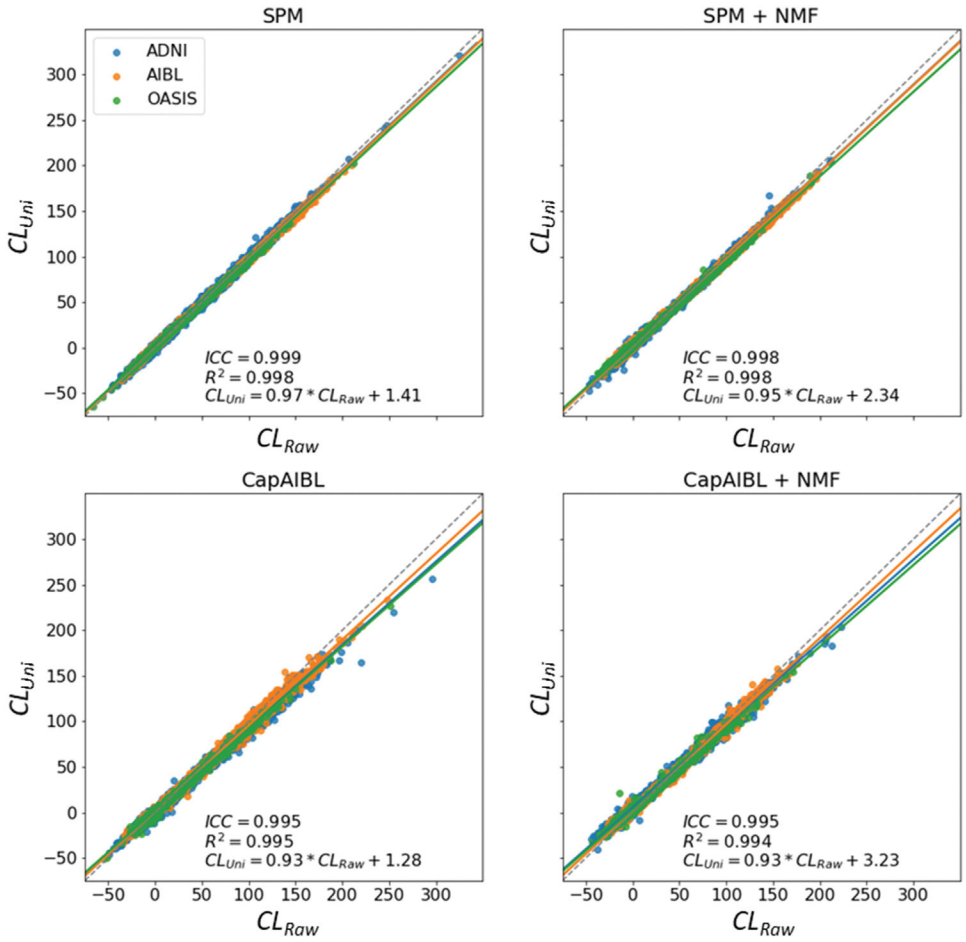

Fig. 1.

Scatter plot of the Centiloid computed using the raw data (^{Raw}CL) compared to the Centiloid computed using images smoothed to a uniform 8mm resolution (^{Uni}CL) quantified using SPM, CapAIBL and their NMF extension. This shows the limited impact of uniform smoothing on CL quantification.

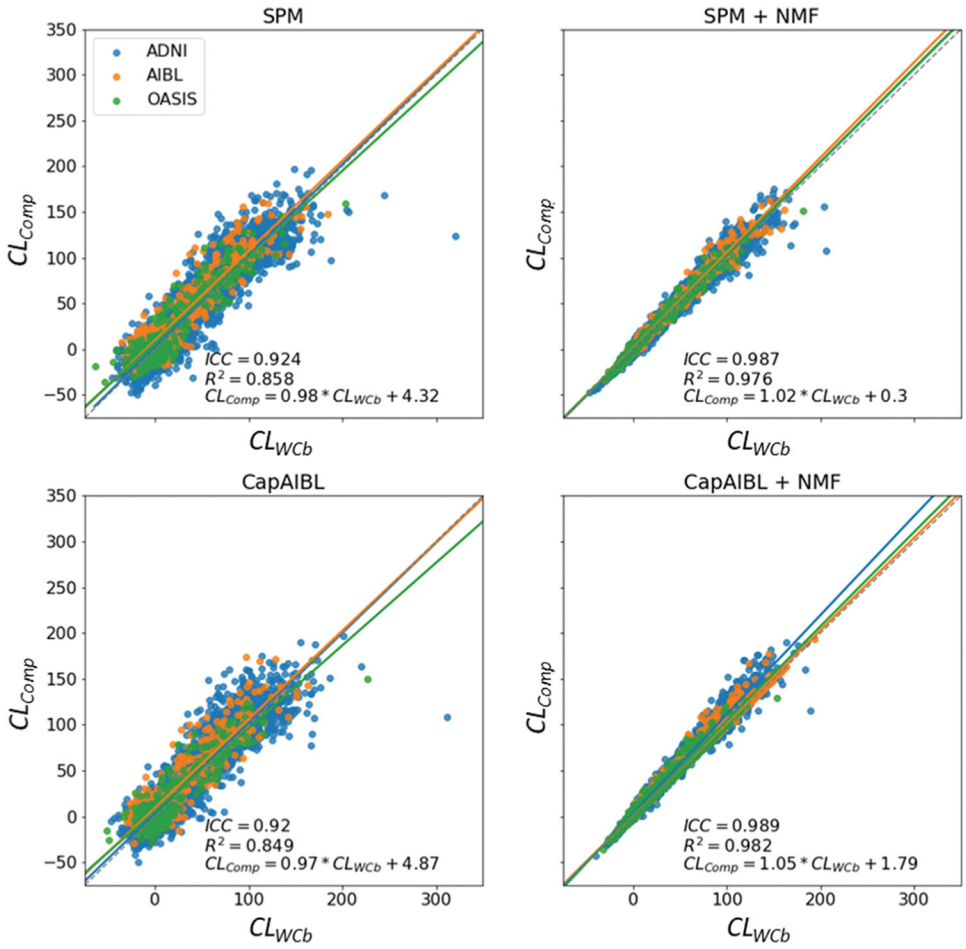

Fig. 2.

Scatter plot of FBP CL_{WCb} and CL_{Comp} quantified using SPM, CapAIBL and their NMF extension.

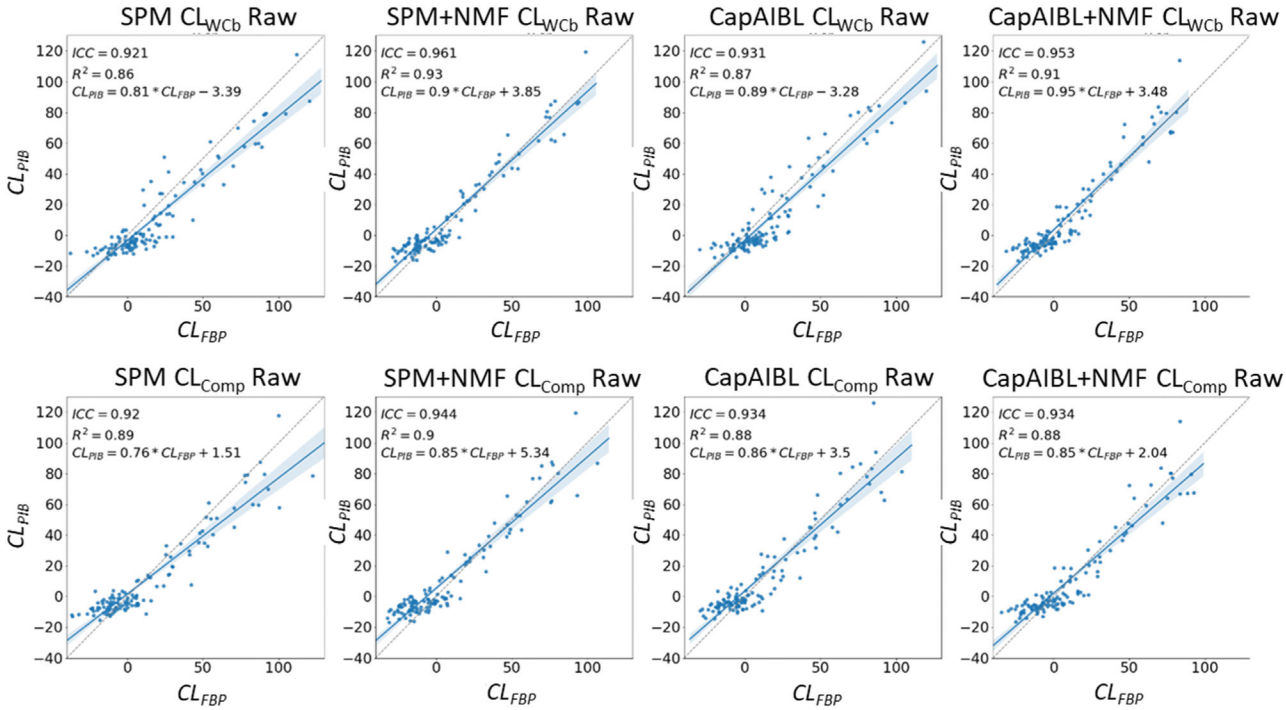

Fig. 3.

Scatter plots of the PiB-FBP CL pairs, using different preprocessing and quantification methods.

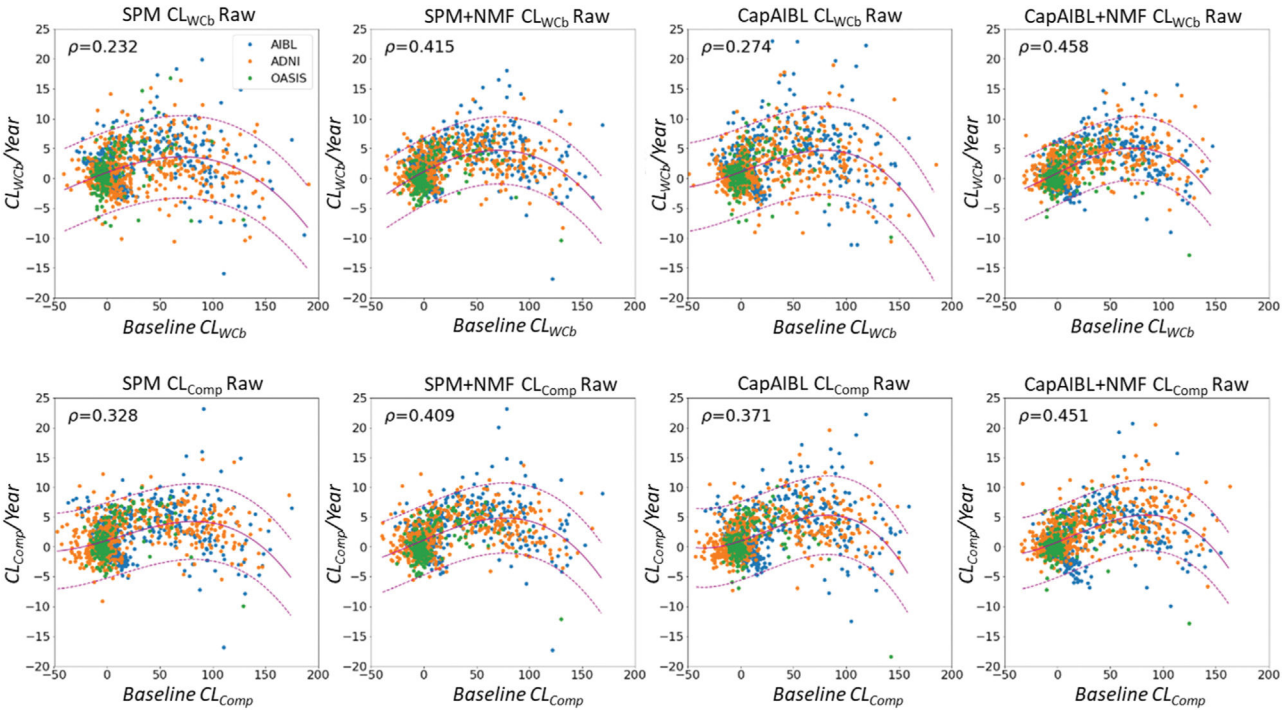

Fig. 4.

Rate of change in CL/year against baseline CL value for CL measured using different preprocessing and quantification methods.

Table 1

Basic demographics and distribution of the number of scans per tracers used in each study.

	AIBL	ADNI	OASIS
Number of scans per tracer PIB/FBP/FBB/NAV/FLUTE	1307/627/14/849/518	226/2901/389/-/-	958/440/-/-/-
Subjects with change of tracer	41.1%	2.9%	34.0%
Subjects with change of scanner	37.2%	18.2%	41.8%
Number of scanner models	4	27	3
Diagnosis at baseline (%) HC/MCI/AD/Others	66/19/13/2	40/44/16/1	83/0/11/6
Age at baseline (Mean [Std]) HC/MCI/AD	72/73/74 [6/8/8]	73/73/75 [7/8/8]	69/-/77 [9/-/8]
MMSE at baseline (Mean [Std]) HC/MCI/AD	28/26/22 [1/2/5]	29/28/23 [1/2/3]	29/-/25 [1/-/4]
Number of timepoints (Mean [Std]) HC/MCI/AD	2.8/2.1/1.6 [1.7/1.4/0.9]	2.3/2.3/1.3 [1.3/1.4/0.5]	1.9/-/1.1 [0.8/-/0.3]
Length of follow-up (Mean [Std]) in years	4.2/3.7/3.4 [2.9/2.5/2.1]	4.2/3.7/3.4 [2.4/2.3/2.2]	4.6/-/4.7 [2.3/-/2.5]

Neuroimage. Author manuscript; available in PMC 2022 November 15.

Table 2

Cohen's Kappa score for the inter-tracer (PiB vs FBP) agreement for classifying Aβ+ and Aβ− scans based on a 20CL threshold. Higher Cohen's Kappa means greater agreement. For each quantification method, the pre-processing leading to the highest agreement is shown as a bold value. The overall best agreement is underlined.

	Preprocessing		*Quantification method*			
	Uniform FWHM	**Reference Region**	**SPM**	**SPM+NMF**	**CapAIBL**	**CapAIBL+ NMF**
$^{Raw}CL_{WCb}$	No	WCb	0.73	0.89	0.69	0.88
$^{Uni}CL_{WCb}$	Yes	WCb	0.68	**<u>0.94</u>**	0.74	0.89
$^{Raw}CL_{Comp}$	No	Composite	**0.87**	0.91	**0.85**	**0.91**
$^{Uni}CL_{Comp}$	Yes	Composite	0.78	0.92	0.81	0.85

Neuroimage. Author manuscript; available in PMC 2022 November 15.

Table 3

Correlation of CL with MMSE. For each quantification method, the pre-processing method leading to the highest R^2 is shown in bold. The overall highest R^2 is underlined.

	Preprocessing		***Quantification method***			
	Uniform FWHM	**Reference Region**	**SPM**	**SPM+NMF**	**CapAIBL**	**CapAIBL+ NMF**
$^{Raw}CL_{WCb}$	No	WCb	0.191	**0.226**	0.195	0.234
$^{Uni}CL_{WCb}$	Yes	WCb	0.188	0.225	0.198	0.235
$^{Raw}CL_{Comp}$	No	Composite	**0.218**	**0.236**	0.229	**<u>0.238</u>**
$^{Uni}CL_{Comp}$	Yes	Composite	0.216	0.235	**0.232**	0.237

Table 4

Effect-size (ES) between HC and AD at baseline based on the CL value. For each quantification method, the pre-processing method leading to the highest effect size is shown in bold. The overall highest effect size is underlined.

	Preprocessing		***Quantification method***			
	Uniform FWHM	**Reference Region**	**SPM**	**SPM+NMF**	**CapAIBL**	**CapAIBL+ NMF**
$^{Raw}CL_{WCb}$	No	WCb	1.654	1.824	1.653	1.837
$^{Uni}CL_{WCb}$	Yes	WCb	1.635	1.807	1.65	1.836
$^{Raw}CL_{Comp}$	No	Composite	**1.826**	**1.876**	**1.839**	**1.849**
$^{Uni}CL_{Comp}$	Yes	Composite	1.801	1.858	1.829	1.841

Table 5

Percentage of outliers in the A$\beta-$ group with changes smaller than −6.33 CL/Y or larger than 8 CL/Y. For each quantification method, the pre-processing leading to the smallest number of outliers is shown as a bold value. The overall lowest percentage of outliers is underlined.

	Preprocessing		***Quantification method***			
	Uniform FWHM	**Reference Region (FBP)**	**SPM**	**SPM+NMF**	**CapAIBL**	**CapAIBL+ NMF**
$^{Raw}CL_{WCb}$	No	WCb	6.43	4.39	6.90	2.75
$^{Uni}CL_{WCb}$	Yes	WCb	6.04	**3.45**	5.41	**2.04**
$^{Raw}CL_{Comp}$	No	Composite	7.57	6.31	6.59	4.63
$^{Uni}CL_{Comp}$	Yes	Composite	**5.22**	5.18	**3.92**	4.75

Neuroimage. Author manuscript; available in PMC 2022 November 15.

Table 6

Percentage of outliers in the Aβ+ group with changes smaller than −16.6 CL/Y or larger than 20.13CL/Y. For each quantification method, the pre-processing leading to the smallest number of outliers is shown as a bold value. The overall lowest percentage of outliers is underlined.

	Preprocessing		***Quantification method***			
	Uniform FWHM	**Reference Region (FBP)**	**SPM**	**SPM+NMF**	**CapAIBL**	**CapAIBL+ NMF**
$^{Raw}CL_{WCb}$	No	WCb	6.41	2.07	7.69	2.78
$^{Uni}CL_{WCb}$	Yes	WCb	5.70	1.64	5.41	1.99
$^{Raw}CL_{Comp}$	No	Composite	2.64	1.64	3.56	1.92
$^{Uni}CL_{Comp}$	Yes	Composite	**2.21**	**1.21**	**2.56**	**1.35**