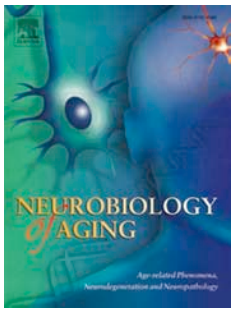
Author/s:
Bourgeat, P;Villemagne, VL;Dore, V;Brown, B;Macaulay, SL;Martins, R;Masters, CL;Ames, D;Ellis, K;Rowe, CC;Salvado, O;Fripp, J

Title:
Comparison of MR-less PiB SUVR quantification methods

Date:
2015-01-01

Citation:
Bourgeat, P., Villemagne, V. L., Dore, V., Brown, B., Macaulay, S. L., Martins, R., Masters, C. L., Ames, D., Ellis, K., Rowe, C. C., Salvado, O. & Fripp, J. (2015). Comparison of MR-less PiB SUVR quantification methods. Neurobiology of Aging, 36 (S1), pp.S159-S166. https://doi.org/10.1016/j.neurobiolaging.2014.04.033.

Persistent Link:
https://hdl.handle.net/11343/43820



Comparison of MR-less PiB SUVR quantification methods

P. Bourgeat, V.L. Villemagne, V. Dore, B. Brown, S.L. Macaulay, R. Martins, C.L. Masters, D. Ames, K. Ellis, C.C. Rowe, O. Salvado, J. Fripp

PII: S0197-4580(14)00545-4

DOI: 10.1016/j.neurobiolaging.2014.04.033

Reference: NBA 9018

To appear in: *Neurobiology of Aging*

Received Date: 9 May 2013

Revised Date: 19 February 2014

Accepted Date: 15 April 2014

Please cite this article as: Bourgeat, P., Villemagne, V.L., Dore, V., Brown, B., Macaulay, S.L., Martins, R., Masters, C.L., Ames, D., Ellis, K., Rowe, C.C., Salvado, O., Fripp, J., for the AIBL Research Group, Comparison of MR-less PiB SUVR quantification methods, *Neurobiology of Aging* (2014), doi: 10.1016/j.neurobiolaging.2014.04.033.

Comparison of MR-less PiB SUVR quantification methods

P. Bourgeat[1], V.L. Villemagne[2,7], V. Dore[1,7], B. Brown[3], S.L. Macaulay[4], R. Martins[3], C.L. Masters[2], D. Ames[5], K. Ellis[6], C.C. Rowe[7], O. Salvado[1], J. Fripp[1], for the AIBL Research Group[8].

1. Preventative Health Flagship, CSIRO Computational Informatics, Herston, QLD 4029, Australia

2. The Mental Health Research Institute, University of Melbourne, Parkville VIC 3052, Australia

3. Centre of Excellence for Alzheimer's Disease Research & Care, Edith Cowan University, Joondalup WA 6027, Australia

4. Preventative Health Flagship, CSIRO CMSE, Parkville VIC 3052, Australia

5. National Ageing Research Institute, University of Melbourne, Parkville, VIC 3052, Australia

6. Department of Psychiatry, University of Melbourne, Parkville, VIC 3052, Australia

7. Department of Nuclear Medicine and Centre for PET, Austin Health, Heidelberg, VIC 3084, Australia

8. http://www.aibl.csiro.au/partners.html.

Corresponding author:

Pierrick Bourgeat

The Australian eHealth Research Centre

Level 5, UQ Health Sciences Building 901/16

Royal Brisbane and Women's Hospital

Herston, Queensland 4029

Australia

Tel: +61 7 3253 3659

Fax: +61 7 3253 3690

Email: pierrick.bourgeat@csiro.au

Abstract

^{11}C-Pittsburgh Compound B (PiB) is a PET tracer designed to bind to β-amyloid (Aβ) plaques, one of the hallmarks of Alzheimer's disease. The potential of PiB as an early marker of Alzheimer's disease led to the increasing use of PiB in clinical research studies and development of several F-18–labeled Aβ radiotracers. Automatic quantification of PiB images requires an accurate parcellation of the brain's gray matter. Typically this relies on a co-registered MRI to extract the cerebellar gray matter, compute the standardized uptake value ratio (SUVR) and provide parcellation and segmentation for quantification of regional and global SUVR. However, not all subjects can undergo MRI, in which case, an MR-less method is desirable. In this study, we assess three PET-only quantification methods: a mean atlas, an adaptive atlas and a multi-atlas approach on a database of 237 subjects having been imaged with both PiB PET and MRI. The PET-only methods were compared against MR-based SUVR quantification, and evaluated in terms of correlation, average error, and performance in classifying subjects with low and high Aβ deposition. The mean atlas method suffered from a significant bias between the estimated neocortical SUVR and the PiB status, resulting in an overall error of 5.6% (R^2=0.98), compared to the adaptive and multi-atlas approaches which had errors of 3.06% and 2.74% respectively (R^2=0.98) and no significant bias. In classifying PiB negative from PiB positive subjects, the mean atlas had 10 misclassified subjects compared to 0 for the adaptive and 1 for the multi-atlas approach. Overall, the adaptive and the multi-atlas approaches performed similarly well against the MR-based quantification, and would be a suitable replacements for PiB quantification when no MRI is available.

Keywords: PET quantification, Alzheimer's disease

1 Introduction

^{11}C-Pittsburgh Compound B (PiB) is a PET tracer designed to bind to β-amyloid (Aβ) plaques, one of the hallmarks of Alzheimer's disease (AD) (Masters and Beyreuther, 2006; Klunk et al., 2004). The Australian Imaging Biomarkers and Lifestyle (AIBL) was the first large scale longitudinal study examining the potential of PiB as an early marker of Alzheimer's disease (Ellis et al., 2009; Rowe et al., 2010). Promising results from this and other studies led to increased use of PiB in research clinical studies and the development of several F-18–labeled Aβ radiotracers. (Rowe et al., 2008; Wong et al., 2010, Vandenberghe et al., 2010) The analysis of PiB images requires accurate quantification of its retention in the brain. Image quantification typically relies on the parcellation of a corresponding MRI to extract the cerebellar cortex and scale the PiB intensity using the standardised uptake value ratio (SUVR) (Lopresti et al., 2005). Cortical and subcortical masks are also derived from the MRI to provide quantification of PiB retention. Along with the

increasing demand for Aβ imaging studies, there is an increasing number of subjects where MRI is contraindicated (pacemakers, metal implants, claustrophobia, etc) or studies where MRI is not part of the patient evaluation. For instance, 20% of the subjects enrolled in AIBL did not undergo an MRI scan at baseline for various reasons, which would have meant that using current MR image analysis techniques, their PiB retention could not have been quantitatively assessed.

The automatic parcellation and quantification of PiB PET images without any structural information is a complex problem due to the lack of anatomical information, partial volume effects and the variability in intensity distribution across the brain. There has been extensive work on quantifying Fluorodeoxyglucose (FDG) PET images without MRI (Ishii et al., 2001, Matsunari et al., 2007). These approaches generally rely on the registration of the PET image to a template, or the search for a peak in intensity in the cortical gray matter (GM) (Neurostat: Minoshima et al., 1995). These strategies work well on FDG images which have less variability in their pattern of retention and are less susceptible to partial voluming, therefore better representing the underlying anatomy. In contrast, PiB PET images exhibit a large variability in the pattern of retention and are more susceptible to partial volume effects and noise. They also suffer from an inversion in contrast between the gray and white matter (WM) in subjects with low to no Aβ burden (where PiB retention in WM is greater than that of the GM), and subjects with a high Aβ burden (where PiB retention in GM is greater than that of WM). The variability in PiB retention, coupled with the limited anatomical resolution of PET scans renders FDG-specific methods unsuitable for PiB quantification.

Previous approaches of PiB quantification have looked at directly segmenting the PET image into GM and WM (Raniga et al., 2007) using an EM approach with a Spline model to account for intensity variations. Such approach can give reasonable GM/WM segmentation, but does not provide a parcellation into anatomical regions of interest (such as the cerebellum) which is required for SUVR quantification. A model-based approach was later proposed (Fripp et al., 2008a) where a subject specific atlas was iteratively generated using a statistical appearance model of PiB retention, and the registration to this atlas was constrained by a statistical deformation model. One major drawback of this approach is that it is model driven, and atypical patterns not initially modelled by the statistical model can result in inaccurate quantification. This approach was also computationally onerous requiring over 6 hours for each image. More recently (Edison et al., 2012), a non-rigid normalisation to a mean PiB atlas was evaluated, showing that the PET-only quantification lead to significant differences compared to MR-based quantification, with an overestimation of the neocortical SUVR in the healthy control (HC) group of 14.4% (p<0.001) and an underestimation in the AD group of 1.1% (not significant). The authors however concluded that these differences did not significantly impact the classification between subjects having a high PiB retention (referred to as PiB positive), and those having a low PiB retention (referred to as PiB negative). Furthermore, the authors suggested that such approach would be suitable for

clinical practice, but an MR-based quantification would be more appropriate for clinical research. To address some of the issues arising from using a mean atlas, we recently proposed a multi-atlas MR-less quantification approach (Bourgeat et al., 2012a), which relies on a strategy commonly used in MRI parcellation (Aljabar et al., 2009), where the *k* most similar PET images are used as atlases, and registered to the target PET image. The computed transforms are used to propagate the corresponding parcellations to the target image and a consensus parcellation is obtained by majority voting. This approach showed a good agreement compared to MR-based quantification, and outperformed the mean atlas approach. This approach is however computationally demanding, requiring *k* non-rigid registrations to be performed for each subjects. It also requires access to a large database of images to be used as atlases. Given the good performance of the previously proposed model-based approach (Fripp et al., 2008a), we are seeking to evaluate a simpler adaptive atlas model which can reduce the computational burden, and is simpler to implement.

In this study, we compared three different approaches: the multi-atlas approach, the mean atlas quantification and a simplified model based approach which adaptively builds an atlas as a weighted sum of a PiB-positive and PiB-negative atlas. Each approach was compared to the MR-based approach in terms of SUVR correlation and error, and in the classification error between PiB positive and PiB-negative cases.

2 Materials and Methods

Data used in the preparation of this paper were obtained from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study (http://www.aibl.csiro.au/) (Ellis et al., 2009). All subjects who underwent both MRI and PiB PET scans at baseline were used. The 239 participants included 38 patients who met National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association criteria for probable AD; (McKhan et al., 1984) 44 participants who met criteria for mild cognitive impairment (MCI); (Petersen et al., 1999), and 157 HC. As previously described (Rowe et al., 2010), a 30-min acquisition scan was performed with Phillips Allegro™ PET cameras at Austin Health for the Melbourne participants and the Sir Charles Gardiner Hospital for the Perth participants, starting 40 minutes after injection of ~370 MBq ^{11}C-PiB. MR imaging was performed using a Siemens 3T Trio (~60%) and a Siemens 1.5T Magnetom Avanto (~40%). The imaging protocol was defined following ADNI's guideline on the 3T scanner (http://adni.loni.ucla.edu/research/protocols/mri-protocols) and a custom MPRage sequence was used on the 1.5T scanner.

2.1 MRI Analysis

All T1W images were first corrected for bias field using the N4 algorithm (Tustison et al., 2010), and smoothed using anisotropic filtering. T2W images were motion corrected using inverse interpolation (Rohlfing et al., 2008). All images were first segmented into GM, WM and CSF in their native space using an in-house implementation of the Expectation Maximisation Segmentation algorithm (Van Leemput et al., 2010). The resulting segmentation was used as a brain mask for skull-stripping, and all the skull stripped T1W images were rigidly registered to the MNI average brain. Each image was re-segmented in the MNI space to generate the final GM, WM and CSF segmentations.

The MRI parcellation was performed using Learning Embeddings for Atlas Propagation (LEAP) following the work of Wolz et al. (2010). This technique has been shown to reduce bias when parcellating a population of elderly subjects by segmentation propagation. LEAP seeks to minimise the amount of deformation required to propagate the parcellation of a set of *M* healthy atlases to a database of atrophic brains, by defining a number of *N* steps following a path going from healthy to atrophic brains, so that the amount of deformation between each step remains small.

LEAP relies on the construction of an affinity matrix which captures the similarity between each pair of images in the database. The affinity matrix is then decomposed using the Eigen Maps decomposition and the first two modes are kept. In Wolz et al. (2010), the metric used to compute the similarity between each pair of images was the Normalised Mutual information (NMI). To normalize the MRI acquisitions across field strengths so that variations in NMI only represent morphological changes, a composite segmentation image of GM, WM and CSF was created for each case, by assigning each tissue type a different value (CSF=1, GM=2 and WM=3). The NMI was then computed between pairs of composite segmentation images using 3 bins. All images were mapped on a 2D manifold, where images with similar level of atrophy can be clustered together. By defining *N* clusters on the manifold, the parcellations from the cluster of *M* atlases can be propagated to the closest cluster by segmentation propagation. The new set of images is subsequently used as atlases to segment the images in the next closest cluster.

We used *M*=20 initial atlases in our approach which were parcellated following the Automated Anatomic Labeling (AAL) parcellation (Tzourio-Mazoyer et al., 2002), and the cerebellum crus I and crus II (Schmahmann et al., 1999) were combined and subsequently used as the reference region for SUVR normalisation. Affine registrations were performed using a block matching approach (Ourselin et al., 2000). Non-rigid registrations were performed using a free form deformation approach using the nifty-reg software (Modat et al., 2010), with a final control point spacing of 2mm. The number of clusters was empirically set to *N*=5. At the end of the procedure, the propagated and voted AAL parcellations were masked by the GM segmentation to generate the final cerebellum and neocortical mask for

subsequent computation of neocortical SUVR uptake. A full description of the MRI parcellation pipeline can be found in (Bourgeat et al., 2012b).

2.2 PiB Quantification

2.2.1 Preprocessing

PiB images were cropped to remove the top and bottom 4 slices which mostly contain ring reconstruction artefacts.

2.2.2 MR-based SUVR scaling

Each PiB image was co-registered to its corresponding MRI in MNI space, and SUVR scaled using the GM masked cerebellum mask. The neocortical SUVR value was computed using the mean SUVR in the GM masked neocortical region composed of the frontal, superior parietal, lateral temporal, occipital and anterior and posterior cingulate regions of the AAL atlas. The computed SUVR values were used as ground truth for all statistical analysis.

2.2.3 Population stratification

The proposed PET-only methods all require a PiB PET atlas to spatially normalise the PET images to a standardised space. To keep the atlas independent from the population of interest, the dataset was split into 2 groups with similar distributions of neocortical SUVR retention, with the atlas from group A (respectively B) used to spatially normalise the images from group B (respectively A). From here onward, each group is analysed independently using the atlas from the other group.

2.2.4 PET Atlas generation

Within each group, all T1W images were first averaged to create an average MRI atlas. This atlas was used as a target for the affine registration (Ourselin et al., 2000) of all T1W images, and the resulting images were averaged to generate a new atlas. The procedure was repeated 3 times. The affine registration was then replaced by a non-rigid registration with a control point spacing of 10mm performed using nifty-reg (Modat et al., 2010). The procedure was repeated 3 times. The segmentations from all T1W images were propagated to the last iteration of the atlas and averaged to generate new population specific priors of GM, WM and CSF. The priors were discretised into binary segmentations

according to the class having the highest probability. Similarly, the AAL parcellations were propagated, and voted to generate a consensus AAL labelling of the mean atlas. Lastly, the co-registered, SUVR normalised PiB images were propagated and averaged to create a mean PiB PET atlas. The atlas and corresponding parcellation were subsequently used to generate SUVR scaled PET images in the mean and adaptive atlas approaches.

2.2.5 PiB only SUVR scaling

Mean atlas SUVR scaling

In an approach similar to that presented in (Edison et al., 2013)**,** all raw PiB images were first affinely and then non-rigidly normalised to the PiB atlas (20mm control point spacing). The GM masked AAL parcellation on the PiB atlas was then used to generate the neocortical SUVR. The mean atlas is illustrated in Figure 1.

Adaptive atlas SUVR scaling

Edison et al. (2013) reported that using a mean atlas introduces a bias, with an overall over-estimation of SUVR values compared to an MR-based quantification. The authors argued against using a disease specific atlas as choosing between the HC or AD atlas based on the clinical classification could lead to a bias in the measurements. To improve the matching between the atlas and the image, while avoiding introducing any bias, we implemented an adaptive subject-specific atlas which is generated for each subject by linearly combining a PiB negative and PiB positive atlas. This is a simplified version of the adaptive atlas presented by Fripp et al. (2008a), where we assume that changes in Aβ retention can be modelled using a linear model. This is a sensible assumption given that close to 80% of the variability in PiB images can be explained by a linear model (Fripp et al., 2008b). The PiB negative atlas A_{neg} was built using cases with neocortical SUVR <= 1.5 (N=59 in group A and N=59 in group B), and the positive atlas A_{pos} using cases with SUVR > 1.5 (N=60 in group A and N=59 in group B). The cut-off value of 1.5 was not selected to split the positive and negative cases in groups of equal size, but was based on a previously published and commonly used cut-off value (Jack et al., 2008, Rowe et al., 2010). The resulting atlases had a mean neocortical SUVR of 1.24 and 2.18 respectively in group A and 1.25 and 2.10 respectively in group B. A weight w governs the contribution of each atlas to the adaptive atlas $A(w)$ so that :

$$\boldsymbol{A(w) = w * A_{neg} + (1 - w) * A_{pos}}$$

The weight *w* was optimised by maximising the normalised mutual information (NMI) between the adaptive atlas and the target *T* so that:

$$\arg\max_{w\epsilon[0,1]} NMI\left(T, A(w)\right)$$

The optimisation was run once using a Powel optimiser (Powell, 1964) after the affine registration and before the non-rigid registration. The optimisation was initialised using *w*=0.5. The weighting between the positive and the negative atlases is solely driven by the image data, and is not introducing any bias regarding the clinical status of the subject. The GM masked AAL parcellation on the PiB atlas was then used to compute the neocortical SUVR. The negative and positive atlases are illustrated in Figure 1.

Multi-atlas SUVR scaling

All PiB images were affinely registered to the PiB atlas, NMI similarity was computed between each pair of images and a 2D manifold was generated. Each PiB was mapped on the manifold, and its *k* closest PiB neighbours on the manifold were identified and registered to the PiB image first affinely, and then non-rigidly with nifty-reg (10mm control point spacing). Their corresponding AAL parcellation and GM mask were then propagated. Consensus parcellations and masks were generated by voting. The consensus GM masked AAL parcellations were then used to compute the neocortical SUVR. The number of neighbours *k* was set to 19. The procedure was performed in a leave-one-out fashion, so that the target PiB image was not part of the selected *k* neighbours. A full description of the multi-atlas parcellation pipeline has been previously published (Bourgeat et al., 2012a).

An overview of all the SUVR quantification methods is available in Figure 2.

2.3 Statistics

Group differences between group A and B were tested using a t-test for differences in terms of age, CDR, MMSE and neocortical SUVR, and a chi-square test for sex, handedness and diagnosis.

In the first set of experiments, we compared the accuracy of the estimation of neocortical SUVR, using both the mean, adaptive and multi-atlas approaches. Neocortical SUVR values were compared to the MR-based normalisation in terms of average retention within each subgroup, with a paired t-test to check the significance of any difference between

the PET-only methods and the MR-based method. The error in the measurements was assessed using the coefficient of determination (R^2) and mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{R_i - T_i}{R_i}\right|$$

with n the number of subjects, R_i the reference measure and T_i the test measure. The agreement between the PET-based methods and the MR-based quantification was also evaluated using Bland Altman plots (Altman and Bland, 1983).

In the second set of experiment, the classification between PiB-positive and PiB-negative in all 3 methods was compared to the classification based on MR-based quantification. A chi-square test was used to test for significant differences in the number of correctly classified subjects.

3 Results

No group differences were observed between group A and B in terms of age, CDR, MMSE, sex, handedness, diagnosis and neocortical SUVR value for the MRI approach or any of the PET-only approaches. Since no differences were found, the results of Group A and B were pooled together for all subsequent statistical analysis.

Mean neocortical SUVR values between all four methods are presented in Table 1. The neocortical SUVR computed using the mean atlas approach was significantly different to the MR-based neocortical SUVR in all but the MCI group, with PiB-based neocortical SUVR being higher than MR-based neocortical SUVR in PiB negative subjects (1.34 vs 1.26) and lower than MR-based neocortical SUVR in PiB positive subjects (2.10 vs 2.18). There was a significant correlation between the signed error and the neocortical SUVR retention (R^2=0.67, $p<0.0001$). There was no significant difference between the neocortical SUVR computed using the adaptive or multi-atlas approaches and the MR-based quantification. There was also no correlation in either approach between the signed error and the neocortical SUVR retention (R^2<0.07).

The coefficient of determination and mean absolute percentage error between the three PiB-only neocortical SUVR quantifications and the MR-based quantification are presented in Table 2, and Bland-Altman plots of the neocortical, frontal and cingulate regions are presented in Figure 3. The mean atlas approach had overall good correlation with the MR-based quantification (R^2=0.98), but also the highest error overall (5.6%) and in all subgroups. The adaptive and multi-atlas approaches showed good agreement with the MRI quantification with overall R^2=0.98 and absolute error of 3.06% for the adaptive approach and 2.74% for the multi-atlas, with the correlation being the highest in the MCI/HC groups and the lowest in the AD group. The error was also worse in the AD group for both approaches. The multi-atlas

approach had lower error in the HC, AD and PiB negative and positive groups, and the adaptive atlas approach had lower error in the MCI group.

In the individual regions, the multi-atlas approach had the lowest overall error in the all regions. The coefficient of determination and mean absolute percentage error for all regions are available in Supplementary Table 1 and Bland-Altman plots for the temporal, occipital and parietal regions are available as Supplementary Figure 1.

Table 3 shows the number of misclassified PiB positive/negative cases for the 3 PET only methods compared to the MR-based method used as the ground truth. The adaptive and multi-atlas approaches showed the lowest number of misclassified cases with 0 and 1 subject misclassified respectively, whereas 10 subjects were misclassified using the mean atlas approach. These differences were however not significant.

Table 1: Neocortical SUVR value presented as mean +/- standard deviation for all four methods. For the three PET-only methods, the p value corresponding to a paired t-test between each approach and the MR-based approach are presented in brackets. The bold values highlight results which are significantly different.

Neocortical SUVR	*MRI*	*Mean Atlas*	*Adaptive Atlas*	*Multi-Atlas*
HC	1.52+/-0.43	**1.56+/-0.36 (<0.001)**	1.52+/-0.42 (0.529)	1.52+/-0.42 (0.709)
MCI	1.93+/-0.58	1.90+/-0.48 (0.0899)	1.93+/-0.56 (0.682)	1.92+/-0.55 (0.499)
AD	2.32+/-0.31	**2.20+/-0.24 (<0.001)**	2.30+/-0.26 (0.347)	2.32+/-0.26 (0.962)
PiB negative	1.26+/-0.085	**1.34+/-0.11 (<0.001)**	1.26+/-0.093 (0.268)	1.26+/-0.082 (0.826)
PiB positive	2.18+/-0.36	**2.10+/-0.29 (<0.001)**	2.17+/-0.34 (0.0577)	2.18+/-0.33 (0.466)

Table 2: Neocortical SUVR estimation compared to the MR-based quantification. Results are presented in terms of mean absolute percentage error, and coefficient of determination: MAPE (R^2). The lowest error for each group is highlighted in a bold font.

MAPE (R^2)	*Mean Atlas*	*Adaptive Atlas*	*Multi-Atlas*
HC	5.97 % (0.97)	2.98 % (0.98)	**2.57 % (0.99)**
MCI	5.23 % (0.98)	**2.49 % (0.99)**	2.57 % (0.98)
AD	5.30 % (0.88)	3.91 % (0.87)	**3.67 % (0.88)**
PiB negative	6.91 % (0.70)	2.93 % (0.76)	**2.41 % (0.80)**
PiB positive	4.55 % (0.94)	3.14 % (0.94)	**3.07 % (0.94)**
All	5.60 % (0.98)	3.06 % (0.98)	**2.74 % (0.98)**

Table 3: Number of misclassified cases between PiB-positive and PiB-negative for the 3 PET-only methods, compared to the MR-based method (lower is better).

	Mean Atlas	*Adaptive Atlas*	*Multi-Atlas*
HC (157)	*9*	-	*1*
MCI (43)	1	-	-
AD (37)	-	-	-
All (237)	*10*	***0***	***1***

4 Discussion

This study extends the work of Edison et al. (2013) to provide a more comprehensive comparison of PET-only methods for SUVR normalisation when no MRI is available. We compared the MR-based normalisation technique to a mean atlas approach (Edison et al., 2013), an adaptive atlas approach (adapted from Fripp et al., 2008a) and a multi-atlas approach (Bourgeat et al., 2012a). These methods were tested on a much larger database of 237 participants (157 HC, 43 MCI and 37 AD). The results demonstrate the limitation of the mean atlas approach which introduces a bias depending on PiB status, that more advanced analysis techniques can mitigate.

The mean atlas approach had overall good correlation, but also the highest error. There was a strong bias between the PiB status and the error. This bias was also reported by Edison et al. (2013), where neocortical SUVR was over-estimated for all cases with a SUVR <1.5 and under estimated for all cases with a SUVR > 2.0. Similar results were obtained in the present study, where SUVR was overestimated in 92% of the cases with a SUVR <1.5 and

underestimated in 91% of the cases with a SUVR > 2.0. This effect was observed for all the cortical regions, except in the cingulate region where there was no significant difference in bias between the PiB negative and positive subjects. The cingulate, being deeper in the cortex, is less susceptible to registration errors and partial volume effects. To better understand the source of this bias, the difference in SUV in the reference region between the mean atlas and the MRI approaches was measured. The difference was less than 1% in both the negative and positive groups, which means that the difference is unlikely caused by an error in the extraction of the reference region, but is most likely due to subtle mis-registrations of the cortical regions. The registration process being intensity based, the high WM retention and low GM retention in the PiB negative subjects tends to result in a small shift of the target's WM towards the mean atlas GM, leading to an increased sampling of the WM and an over-estimation of the neocortical retention. Similarly, the low WM retention and high GM retention in the PiB-positive subjects tends to result in a small shift of the target's GM towards the mean atlas WM, leading to an increased sampling of the CSF and an under-estimation of the neocortical retention. Moreover, we found that the cut-off value between under and over-estimation relates closely to the neocortical SUVR value of the mean atlas (1.70 and 1.67 for group A and B respectively). The strong bias in the mean atlas approach results in higher errors in the estimated SUVR values and a higher number of misclassified cases when separating PiB positive from PiB negative subjects.

The adaptive atlas and multi-atlas approaches do not suffer from this bias, as they both match the atlas(es) to the image before the non-rigid registration. They both, however, tend to underestimates the neocortical SUVR in subjects with high neocortical retention. As observed in Figure 3, the amplitude of the error tends to increase with the neocortical SUVR value. This shows the limitation of our model based approach which is too simple to properly match cases with atypical shape or pattern of deposition. PiB-negative cases tend to form a tight cluster in terms of pattern of PiB retention and gray matter atrophy, which means that they can be easily modelled by the adaptive atlas approach. PiB positive subjects, however, are more likely to have pronounced atrophy and various degrees of asymmetric or focal Aβ deposition. They are also more susceptible to have mixed pathology with a reported 45% of probable AD and 54 % of MCI having mixed pathology (Schneider et al., 2009). The adaptive atlas, does not explicitly take atrophy into account, which can result in a non optimal choice of the weighting factor in the case of subjects with pronounced atrophy. Focal or asymmetric deposition cannot be easily modelled by the adaptive atlas which also results in larger registration errors and lower accuracy in SUVR quantification.

Although the multi-atlas approach should be less susceptible to the effect of atrophy, as the closest neighbours are also likely to be present with similar levels of atrophy, it is also affected by the same shortcomings as the adaptive

approach for similar reasons. Cases in the PiB negative groups are quite similar to each other in terms of pattern of deposition and atrophy, which results in lower registration errors when registering these cases together. The PiB positive group is however much more heterogeneous, with a larger spread of atrophy and pattern of deposition. As a result, the closest neighbours are often very dissimilar to each other and to the target image, which results in an increase in registration errors. This is further illustrated by the much smaller error in the PiB negative group (2.4%) compared to the PiB positive group (3.1%). Previous studies on MRI (Aljabar et al., 2009) have shown that the accuracy of the registration is directly related to the similarity between the atlases and the target. Increasing the number of PiB positive subjects in the database would increase the number of potential candidates, reducing the heterogeneity within the group of closest neighbours and improving the overall accuracy of this approach, especially in the PiB positive group. This would however require a much larger population than we currently have access to.

The choice of control point spacing, or coarseness, of the registration can play a major role in the overall results when little anatomical information is present in the images to be registered. We evaluated various degrees of coarseness for the registration in the various approaches, repeating each experiment with a control point spacing of 20, 10 and 5mm (results not shown). For the mean and adaptive atlas approaches, we found that a coarse registration at 20mm provided the optimal results in terms correlation and error, as a finer grid tends to introduce folding in the deformation field, leading to improbable registrations. The multi-atlas approach did however benefit from a finer registration with optimal results found at 10mm. This also leads to some folding, but the subsequent voting process can smooth out registration errors. Similarly, the choice of the number of neighbours used in the voting step can impact the overall performance of the multi-atlas approach. The voting step was repeated with $k\epsilon[10,20]$, and we found that k=19 provided optimal results in terms correlation, SUVR estimation error and classification error.

While the multi and adaptive atlas approach had similar levels of accuracy, these methods have various degrees of computational complexity, with the adaptive atlas method being the least demanding, requiring only 1 non-rigid registration, whereas the multi-atlas approach requires k non-rigid registration. When assessing both accuracy and computation time, the adaptive atlas approach provides the best trade off, having both low computational burden and high accuracy devoid of significant bias. The error in neocortical SUVR estimation of 3.06% compares well the variability of PiB of 3.5% measured in test retest in the same population (Villemagne et al., 2011). This would make the adaptive atlas approach more suitable for clinical diagnosis application where short computational time is required. For clinical research where processing time is not as critical, the multi-atlas approach can offer a higher accuracy.

In conclusion, we have presented a comprehensive comparison of PET-only SUVR normalisation techniques. The results showed that using adaptive or multi-atlas approaches can lead to a more accurate quantification of SUVR values, as compared to using a mean atlas approach. As Aβ tracers become more prevalent and used in the clinic, MR-less approaches will become critical. Future work will test the robustness of the PET only quantification methods for longitudinal studies, and validate their use with some of the new F18 Aβ tracers.

Acknowledgements

AIBL is a large collaborative study and a complete list of contributors can be found at our website www.aibl.csiro.au. We thank all who took part in the study. This research is supported by the Science and Industry Endowment Fund (www.sief.org.au).

Disclosure statement

The authors report no conflicts of interest.

References

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage. 46, 726-738.

Bland, J., Altman, D., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 327, 307-310.

Bourgeat, P., Raniga, P., Dore, V., Zhou, L., Macaulay, S.L., Martins, R., Masters, C., Ames, D., Ellis, K.A., Villemagne, V., Rowe, C., Salvado, O., Fripp, J. On behalf of the AIBL research group. 2012a. Manifold driven MR-less PiB SUVR normalisation. In MICCAI 2012 Workshop on Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD'12), Nice, France, October 5.

Bourgeat, P., Dore, V., Shen, K.-K., Raniga, P., Salvado, O., Fripp, J. And the Alzheimer's Disease Neuroimaging Initiative. 2012b. Atrophy challenge: Enforcing Longitudinal Consistency in Longitudinal analysis using multi-atlas segmentation. In MICCAI 2012 Workshop on Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD'12), Nice, France, October 5.

Edison, P., Carter, S., Rinne, J., Gelosa, G., Herholz, K., Nordberg, A., Brooks, D., Hinz, R., 2013. Comparison of MRI based and PET template based approaches in the quantitative analysis of amyloid imaging with PIB-PET. Neuroimage. 70, 423-433.

Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., 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. 21, 672-687.

Fripp, J., Bourgeat, P., Raniga, P., Acosta, O., Villemagne, V., Jones, G., O'keefe, G., Rowe, C., Ourselin, S., Salvado, O., 2008. MR-less high dimensional spatial normalization of 11C PiB PET images on a population of elderly, mild cognitive impaired and Alzheimer disease patients, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008a. Springer, pp. 442-449.

Fripp, J., Bourgeat, P., Acosta, O., Raniga, P., Modat, M., Pike, K.E., Jones, G., O'Keefe, G., Masters, C.L., Ames, D., Ellis, K.A., Maruff, P., Currie, J., Villemagne, V.L., Rowe, C.C., Salvado, O., Ourselin, S., 2008b. Appearance modeling of 11C PiB PET images: characterizing amyloid deposition in Alzheimer's disease, mild cognitive impairment and healthy aging. Neuroimage. 43, 430-439.

Ishii, K., Willoch, F., Minoshima, S., Drzezga, A., Ficaro, E.P., Cross, D.J., Kuhl, D.E., Schwaiger, M., 2001. Statistical brain mapping of 18F-FDG PET in Alzheimer's disease: validation of anatomic standardization for atrophied brains. J Nucl Med. 42, 548-557.

Jack, C.R., Lowe, V.J., Senjem, M.L., Weigand, S.D., Kemp, B.J., Shiung, M.M., Knopman, D.S., Boeve, B.F., Klunk, W.E., Mathis, C.A., Petersen, R.C., 2008. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain 131, 665-680.

Klunk, W.E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D.P., Bergström, M., Savitcheva, I., Huang, G., Estrada, S., 2004. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol. 55, 306-319.

Lopresti, B.J., Klunk, W.E., Mathis, C.A., Hoge, J.A., Ziolko, S.K., Lu, X., Meltzer, C.C., Schimmel, K., Tsopelas, N.D., DeKosky, S.T., 2005. Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis. J Nucl Med. 46, 1959-1972.

Masters, C.L., Beyreuther, K., 2006. Alzheimer's centennial legacy: prospects for rational therapeutic intervention targeting the Aβ amyloid pathway. Brain. 129, 2823-2839.

Matsunari, I., Samuraki, M., Chen, W., Yanase, D., Takeda, N., Ono, K., Yoshita, M., Matsuda, H., Yamada, M., Kinuya, S., 2007. Comparison of 18F-FDG PET and optimized voxel-based morphometry for detection of Alzheimer's disease: aging effect on diagnostic performance. J Nucl Med. 48, 1961-1970.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 34, 939-939.

Minoshima, S., Frey, K.A., Koeppe, R.A., Foster, N.L., Kuhl, D.E., 1995. A Diagnostic Approach in Alzheimer's Disease Using Three-Dimensional Stereotactic Surface Projections of Fluorine-18-FDG PET. J Nucl Med. 36,1238-1248.

Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. Comput. Methods Programs Biomed. 98, 278-284.

Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N., 2001. Reconstructing a 3D structure from serial histological sections. Image Vision Comput. 19, 25-31.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 303-308.

Powell, M.J.D. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput J. 7, 155-162.

Raniga, P., Bourgeat, P., Villemagne, V., O'Keefe, G., Rowe, C., Ourselin, S., 2007. Spline based inhomogeneity correction for 11C-PIB PET segmentation using expectation maximization, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007. Springer, pp. 228-235.

Rohlfing, T., Rademacher, M.H., Pfefferbaum, A., 2008. Volume Reconstruction by Inverse Interpolation: Application to Interleaved MR Motion Correction, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008. Springer, pp. 798-806.

Rowe, C.C., Ackerman, U., Browne, W., Mulligan, R., Pike, K.L., O'Keefe, G., Tochon-Danguy, H., Chan, G., Berlangieri, S.U., Jones, G., 2008. Imaging of amyloid β in Alzheimer's disease with 18F-BAY94-9172, a novel PET tracer: proof of mechanism. Lancet Neurol, 7, 129-135.

Rowe, C.C., Ellis, K.A., Rimajova, M., Bourgeat, P., Pike, K.E., Jones, G., Fripp, J., Tochon-Danguy, H., Morandeau, L., O'Keefe, G., 2010. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol Aging. 31, 1275-1283.

Schneider, J.A., Arvanitakis, Z., Leurgans, S.E., Bennett, D.A., 2009. The neuropathology of probable Alzheimer disease and mild cognitive impairment. Ann Neurol. 66, 200-208.

Schmahmann, J.D., Doyon, J., McDonald, D., Holmes, C., Lavoie, K., Hurwitz, A.S., Kabani, N., Toga, A., Evans, A., Petrides, M. 1999. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. Neuroimage. 10, 233-60.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 29, 1310-1320.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15, 273-289.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imaging. 18, 897-908.

Vandenberghe, R., Van Laere, K., Ivanoiu, A., Salmon, E., Bastin, C., Triau, E., Hasselbalch, S., Law, I., Andersen, A., Korner, A., 2010. 18F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: A phase 2 trial. Ann Neurol. 68, 319-329.

Villemagne, V.L., Pike, K.E., Chételat, G., Ellis, K.A., Mulligan, R.S., Bourgeat, P., Ackermann, U., Jones, G., Szoeke, C., Salvado, O., 2011. Longitudinal assessment of Aβ and cognition in aging and Alzheimer disease. Ann Neurol. 69, 181-192.

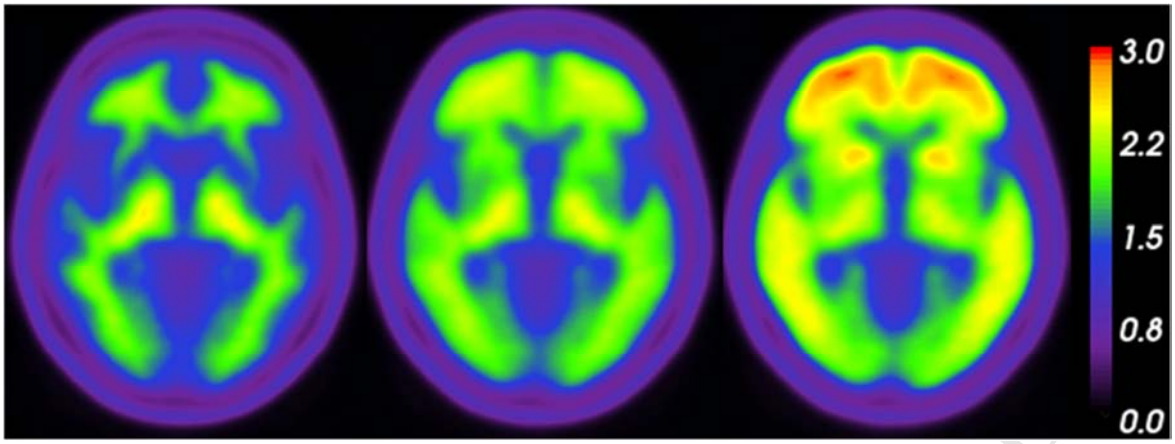
Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D., 2010. LEAP: Learning embeddings for atlas propagation. Neuroimage. 49, 1316-1325.

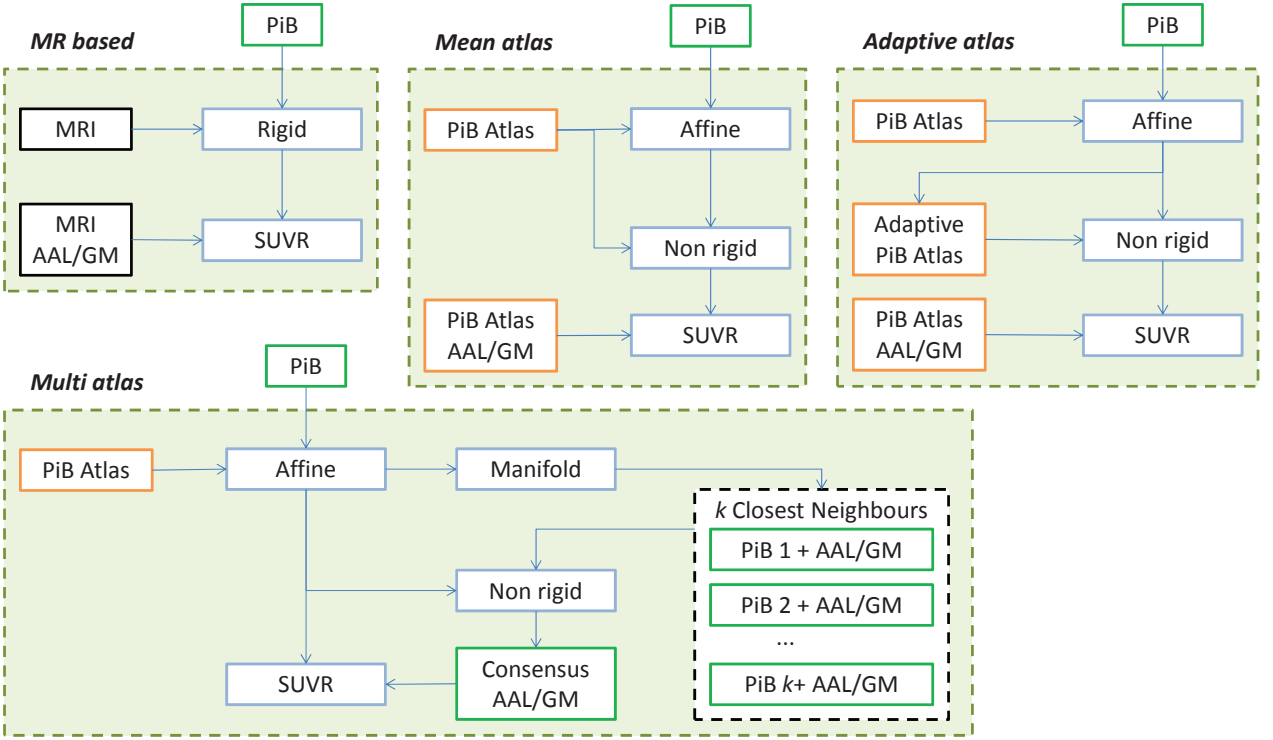
Wong, D.F., Rosenberg, P.B., Zhou, Y., Kumar, A., Raymont, V., Ravert, H.T., Dannals, R.F., Nandi, A., Brašić, J.R., Ye, W., 2010. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18). J Nucl Med. 51, 913-920.

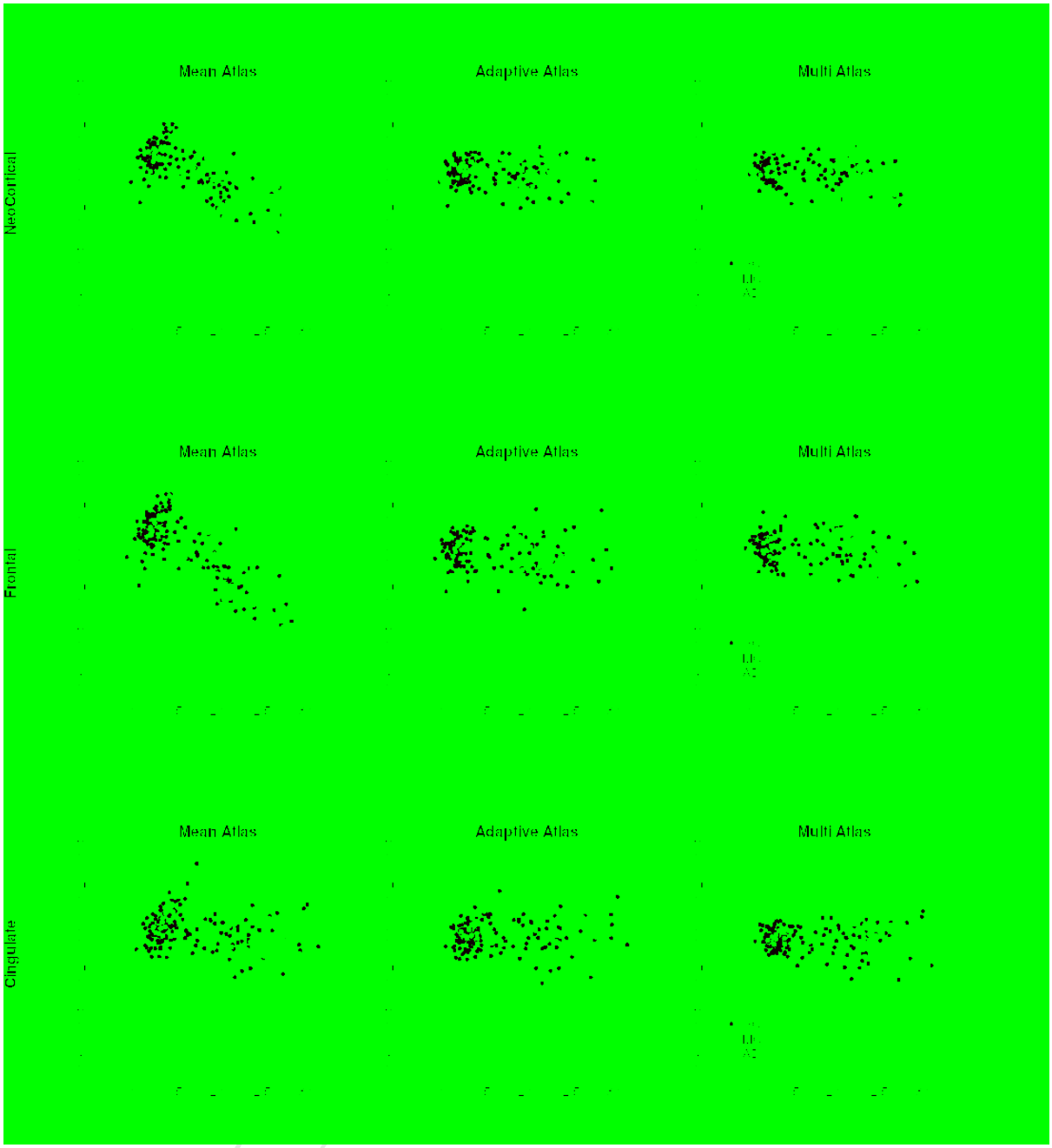
Figure 1. The negative (left) and positive (right) atlas used in the adaptive atlas approach, and the mean atlas (middle).

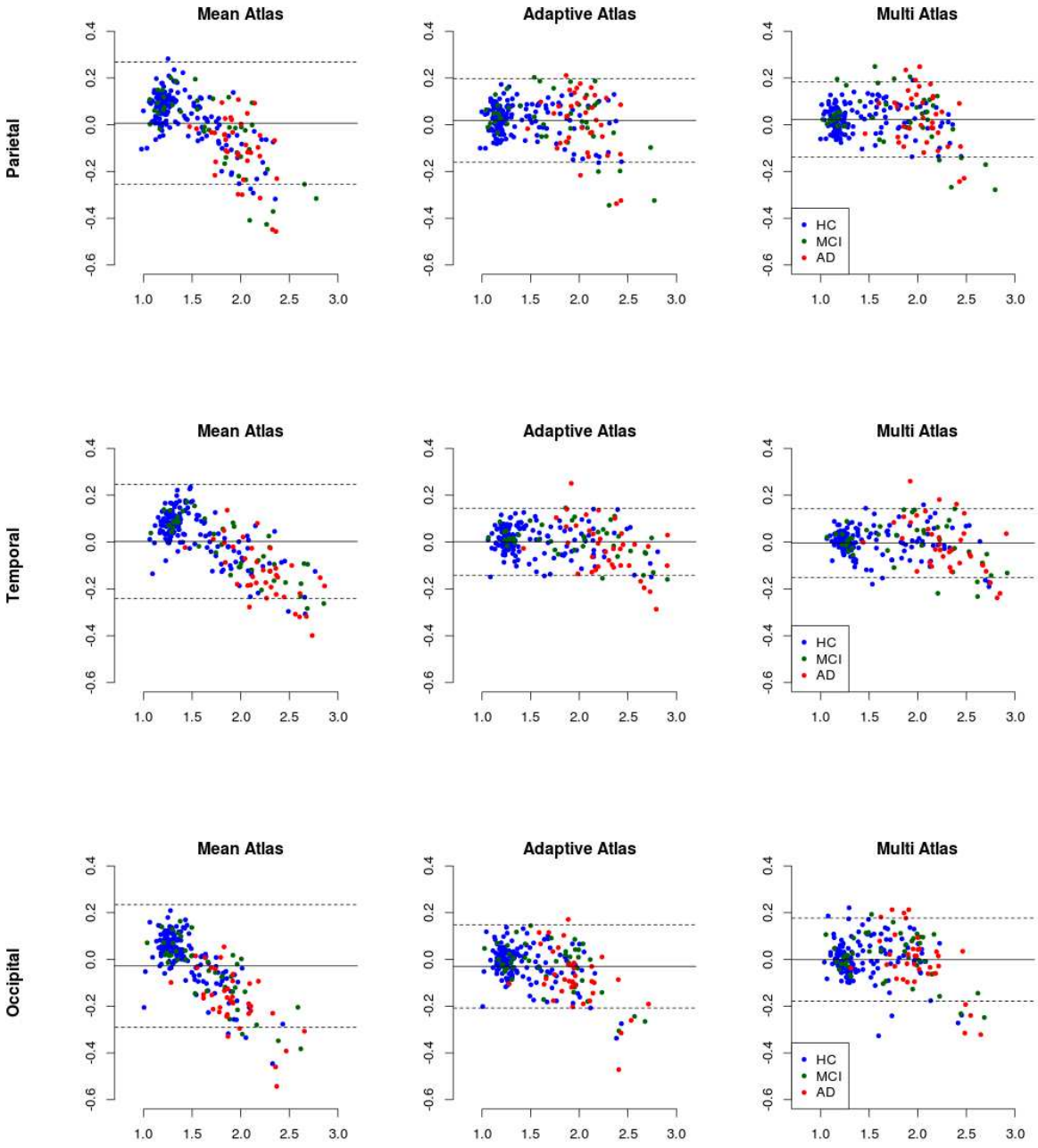
Figure 2. Flowchart of the four quantification methods evaluated, with the MR-based (top left), mean atlas (top center), adaptive atlas (top right) and multi-atlas (bottom) approaches.

Figure 3: Bland-Altman plots of the neocortical (top row), frontal (middle row) and cingulate (bottom row) SUVR estimation using the mean atlas (left), adaptive atlas (middle) and the multi-atlas (right) approaches compared to the MR-based quantification. The y axis represents the difference between the PET-only and MR-based approaches, and the x axis is the average of both methods. The solid horizontal line is the mean difference between the 2 approaches, and the dash lines are the 95% confidence interval.




MR based
PiB
MRI
Rigid
MRI
AAL/GM
SUVR
Mean atlas
PiB
PiB Atlas
Affine
Non rigid
PiB Atlas
AAL/GM
SUVR
Adaptive atlas
PiB
PiB Atlas
Affine
Adaptive
PiB Atlas
Non rigid
PiB Atlas
AAL/GM
SUVR
Multi atlas
PiB
PiB Atlas
Affine
Manifold
k Closest Neighbours
PiB 1 + AAL/GM
PiB 2 + AAL/GM
...
PiB k+ AAL/GM
Non rigid
SUVR
Consensus
AAL/GM





Supplementary figure 1: Bland-Altman plots of the parietal (top row), temporal (middle row) and occipital (bottom row) SUVR estimation using the mean atlas (left), adaptive atlas (middle) and the multi atlas (right) approaches compared to the MR-based quantification. The y axis represents the difference between the PET-only and MR-based approaches, and the x axis is the average of both methods.

Supplementary table 1: SUVR estimation for the frontal, parietal, cingulate, temporal and occipital regions compared to the MR-based quantification. Results are presented in terms of mean absolute percentage error, and coefficient of determination: MAPE (R^2). The lowest error for each group is highlighted in a bold font.

Frontal	*Mean Atlas*	*Adaptive Atlas*	*Multi-Atlas*
HC	8.94 % (0.96)	3.87 % (0.98)	**3.53 % (0.98)**
MCI	7.44 % (0.97)	**3.11 % (0.98)**	3.39 % (0.98)
AD	7.65 % (0.87)	4.38 % (0.87)	**4.03 % (0.89)**
PiB Negative	10.32 % (0.64)	3.73 % (0.70)	**3.50 % (0.71)**
PiB Positive	6.63 % (0.92)	3.90 % (0.92)	**3.66 % (0.93)**
All	8.47 % (0.97)	3.81 % (0.98)	**3.58 % (0.98)**

Parietal	*Mean Atlas*	*Adaptive Atlas*	*Multi-Atlas*
HC	6.98 % (0.94)	4.32 % (0.96)	**3.81 % (0.97)**
MCI	7.22 % (0.94)	5.03 % (0.95)	**4.80 % (0.96)**
AD	5.84 % (0.72)	5.11 % (0.73)	**4.84 % (0.79)**
PiB Negative	7.95 % (0.57)	4.39 % (0.63)	**3.59 % (0.66)**
PiB Positive	5.75 % (0.87)	4.75 % (0.89)	**4.71 % (0.91)**
All	6.85 % (0.95)	4.57 % (0.96)	**4.15 % (0.97)**

Cingulate	*Mean Atlas*	*Adaptive Atlas*	*Multi-Atlas*
HC	3.71 % (0.97)	3.27 % (0.98)	**3.11 % (0.99)**
MCI	3.51 % (0.98)	**2.79 % (0.98)**	2.90 % (0.99)
AD	3.63 % (0.87)	3.85 % (0.85)	**3.43 % (0.87)**
PiB Negative	4.26 % (0.77)	3.32 % (0.81)	**3.01 % (0.86)**
PiB Positive	3.07 % (0.94)	**3.23 % (0.93)**	**3.23 % (0.94)**
All	3.66 % (0.98)	3.27 % (0.98)	**3.12 % (0.98)**

Temporal	*Mean Atlas*	*Adaptive Atlas*	*Multi-Atlas*
HC	6.34 % (0.96)	3.28 % (0.98)	**2.70 % (0.98)**
MCI	5.74 % (0.98)	**2.64 % (0.99)**	3.35 % (0.98)
AD	6.14 % (0.92)	**3.94 % (0.92)**	3.97 % (0.92)
PiB Negative	7.35 % (0.66)	3.14 % (0.68)	**2.32 % (0.77)**
PiB Positive	5.05 % (0.95)	**3.39 % (0.95)**	3.70 % (0.94)
All	6.20 % (0.97)	3.27 % (0.98)	**3.02 % (0.98)**

Occipital	*Mean Atlas*	*Adaptive Atlas*	*Multi-Atlas*
HC	5.78 % (0.92)	**3.72 % (0.94)**	3.80 % (0.93)
MCI	6.00 % (0.96)	**3.96 % (0.96)**	4.40 % (0.95)
AD	8.42 % (0.88)	5.72 % (0.87)	**4.54 % (0.88)**
PiB Negative	5.87 % (0.62)	3.33 % (0.69)	**3.07 % (0.76)**
PiB Positive	6.58 % (0.91)	**4.81 % (0.92)**	4.98 % (0.90)
All	6.23 % (0.95)	4.08 % (0.96)	**4.03 % (0.95)**