NeuroImage 108 (2015) 450–459

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Improved longitudinal [^{18}F]-AV45 amyloid PET by white matter reference and VOI-based partial volume effect correction

Matthias Brendel [a], Marcus Högenauer [a], Andreas Delker [a], Julia Sauerbeck [a], Peter Bartenstein [a], John Seibyl [b], Axel Rominger [a,*], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Dept. of Nuclear Medicine, University of Munich, Germany
[b] MNI, New Haven, USA

ARTICLE INFO

Article history:
Accepted 26 November 2014
Available online 4 December 2014

Keywords:
^{18}F-AV45-PET
Alzheimer's disease
Reference region
Partial volume effect correction

ABSTRACT

Amyloid positron-emission-tomography (PET) offers an important research and diagnostic tool for investigating Alzheimer's disease (AD). The majority of amyloid PET studies have used the cerebellum as a reference region, and clinical studies have not accounted for atrophy-based partial volume effects (PVE). Longitudinal studies using cerebellum as reference tissue have revealed only small mean increases and high inter-subject variability in amyloid binding. We aimed to test the effects of different reference regions and PVE-correction (PVEC) on the discriminatory power and longitudinal performance of amyloid PET.

We analyzed [^{18}F]-AV45 PET and T1-weighted MRI data of 962 subjects at baseline and two-year follow-up data of 258 subjects. Cortical composite volume-of-interest (VOI) values (COMP) for tracer uptake were generated using either full brain atlas VOIs, gray matter segmented VOIs or gray matter segmented VOIs after VOI-based PVEC. Standard-uptake-value ratios (SUVR) were calculated by scaling the COMP values to uptake in cerebellum ($SUVR_{CBL}$), brainstem ($SUVR_{BST}$) or white matter ($SUVR_{WM}$). Mean SUV, SUVR, and changes after PVEC were compared at baseline between diagnostic groups of healthy controls (HC; N = 316), mild cognitive impairment (MCI; N = 483) and AD (N = 163). Receiver operating characteristics (ROC) were calculated for the discriminations between HC, MCI and AD, and expressed as area under the curve (AUC). Finally, the longitudinal [^{18}F]-AV45-PET data were used to analyze the impact of quantitation procedures on apparent changes in amyloid load over time.

Reference region SUV was most constant between diagnosis groups for the white matter. PVEC led to decreases of COMP-SUV in HC (−18%) and MCI (−10%), but increases in AD (+7%). Highest AUCs were found when using PVEC with white matter scaling for the contrast between HC/AD (0.907) or with brainstem scaling for the contrast between HC/MCI (0.658). Longitudinal increases were greatest in all diagnosis groups with application of PVEC, and inter-subject variability was lowest for the white matter reference.

Thus, discriminatory power of [^{18}F]-AV45-PET was improved by use of a VOI-based PVEC and white matter or brainstem rather than cerebellum reference region. Detection of longitudinal amyloid increases was optimized with PVEC and white matter reference tissue.

Abbreviations: Aß, ß-amyloid; AD, Alzheimer's disease; ADAS-cog, Alzheimer's disease assessment scale — cognitive subscale; ADNI, Alzheimer's Disease Neuroimaging Initiative; AUC, area under the curve; BST, brainstem; CBL, cerebellum; COMP, combined region of frontal, parietal, temporal and posterior cingulate cortices; CSF, cerebrospinal fluid; FAD, familiar Alzheimer's disease; FDA, Food and Drug Administration; FDG, fluorodeoxyglucose; FULL, full atlas VOIs; GM, gray matter; HC, cognitively healthy; MANOVA, multivariate analysis of covariance; MCI, mild cognitive impairment; MMSE, mini mental state examination; MNI, Montreal Neurological Institute; MRI, magnetic resonance imaging; NIA, National Institute on Aging; NIBIB, National Institute of Biomedical Imaging and Bioengineering; PET, positron emission tomography; PVE, partial volume effect; PVEC, partial volume effect correction; R, Pearson's coefficient of correlation; REF, reference configuration from tracer validation; ROC, receiver operating characteristics; SAD, sporadic Alzheimer's disease; SUV, standard uptake value; SUVR, standard uptake value ratio; T1w, T1 weighted; VOI, volume of interest; WM, white matter; FWHM, Full-width-at-half-maximum.

* Corresponding author at: Department of Nuclear Medicine, University of Munich, Germany. Fax: +49 89 4400 77646.
E-mail address: axel.rominger@med.uni-muenchen.de (A. Rominger).

[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.

http://dx.doi.org/10.1016/j.neuroimage.2014.11.055
1053-8119/

Introduction

Alzheimer's disease (AD) is the most common form of dementia; its incidence increases exponentially as a function of age, which is imposing an onerous burden on health care systems in societies with aging populations (Ziegler-Graham et al., 2008). Neurofibrillary tangles and amyloid plaques together comprise the hallmark neuropathology of AD (Braak and Braak, 1991). Elevated brain amyloid burden is now clearly associated with cognitive decline in the healthy elderly (HC) (Lim et al., 2012) and in cases of mild cognitive impairment (MCI) (Lim et al., 2014). Amyloid PET offers a feasible tool for the early detection of brain amyloidosis, and the recent development of fluorine-18 labeled amyloid radioligands such as [^{18}F]-AV45 has made this technique available to PET centers lacking an on-site cyclotron/radiochemistry facility.

In clinical PET practice, Aβ-positivity and -negativity are visually assessed with good inter- and intra-reader agreement (Clark et al., 2012). However, a semiquantitative approach is better suited especially to the requirements of longitudinal clinical trials of amyloidosis progression and treatment. The issue of defining an optimal reference region has been extensively discussed for normalization of [^{18}F]-fluorodeoxyglucose-(FDG) PET relative to cerebellum, pons/brainstem, global mean, or a reference cluster (Bohnen et al., 2012; Borghammer et al., 2009; Dukart et al., 2013; Yakushev et al., 2008).

In PET imaging with [^{18}F]-AV45- and [^{11}C]-PiB, the entire cerebellum and the cerebellar gray matter (GM) have emerged as the most widely used reference regions for quantitation of amyloid burden (Weiner et al., 2013). However, a recent longitudinal [^{11}C]-PiB PET study of mild cognitive impairment (MCI) and AD showed high inter-subject variability based on a cerebellar GM reference (van Berckel et al., 2013). Furthermore, amyloid PET results are potentially biased by partial volume effects (PVE), which have a considerable impact in patients with pronounced atrophy (Thomas et al., 2011), which is particularly problematic in longitudinal studies.

Given these considerations, we aimed to compare systematically the quantitation of [^{18}F]-AV45-PET results for different reference regions, using as our material the Alzheimer's Disease Neuroimaging Initiative (ADNI)-dataset, which includes more than 1000 amyloid PET cases. Furthermore, we set about to investigate the impact of a volume-of-interest (VOI)-based partial volume effect correction (PVEC) on the semiquantitative analyses. Receiver operating characteristics (ROC) were obtained for the baseline discrimination of HC from MCI and AD cases in order to identify the most sensitive amyloid-PET analysis. Finally, two-year longitudinal [^{18}F]-AV45-PET data from 258 patients used to test the impact of the above factors on apparent changes in amyloid load with time.

Materials and methods

Alzheimer's disease neuroimaging initiative

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. 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). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California — San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Data from ADNI-GO/-2 were included in this work. Pre-processed brain [^{18}F]-AV45-PET images and temporally corresponding T1-weighted MPRAGE images were downloaded from the ADNI database as available on Jan 16th, 2014.

Patient selection and study design

1018 subjects from ADNI-GO and ADNI-2 who had undergone brain [^{18}F]-AV45-PET and T1 weighted MPRAGE (T1w) MRI at their study baseline were included in this investigation. Clinical diagnoses as provided by the ADNI database at the time of PET-imaging were: 336 HC, 508 MCI and 174 AD. A subgroup of 278 subjects (94 HC, 158 MCI and 26 AD) who underwent a second [^{18}F]-AV45-PET and T1w MRI at two-year follow-up was used for longitudinal analyses.

Image data

ADNI [^{18}F]-AV45-PET acquisition and pre-processing

The [^{18}F]-AV45-PET images had been acquired using Siemens, GE and Philips PET scanners according to a standard dynamic 50–70 min protocol following the intravenous injection of 370 ± 37 MBq of [^{18}F]-AV45. Data were corrected for both scatter and measured attenuation, which was determined using the CT scan for PET/CT scanners, or a transmission scan with [^{68}Ge] or [^{137}Cs] rotating rod sources for PET-only scanners. Images were reconstructed using scanner-specific algorithms, and sent to the University of Michigan, where they were reviewed for artifacts and transmitted to the Laboratory of NeuroImaging (LONI) for storage.

Downloaded [^{18}F]-AV45-PET images in DICOM format had been pre-processed in four steps: 1) motion correction by co-registration of single five minute frames; 2) time frame averaging (50–70 min p.i.); 3) co-registration of longitudinal data to the baseline scan and reorientation in a standardized 160 × 160 × 96 matrix with 1.5 mm cubic voxels; 4) smoothing with a scanner-specific filter function to an isotropic resolution of 8 mm. Further details are provided in Supplement 1.

ADNI MRI acquisition and pre-processing

T1-weighted MRI scans had been acquired using Siemens, GE or Philips MRI scanners according to a standard protocol (Jack et al., 2008) involving acquisitions of two 3-D MPRAGE imaging sequences per subject. Of the two images acquired per subject and time-point, the ADNI quality assurance team selected the better image for pre-processing, based on the presence and severity of commonly occurring image artifacts.

MRI preprocessing involved: 1) application of a scanner-specific correction for gradient nonlinearity distortion (Gradwarp) (Jovicich et al., 2006); 2) correction for image intensity non-uniformity (B1) (Jack et al., 2008); 3) histogram peak sharpening algorithm for bias field correction (N3) (Sled et al., 1998.); 4) application of spatial scaling factors obtained by phantom measurements. For images acquired on Philips scanners, B1 correction was already implemented, and the gradient systems with this instrument tended to be linear (Jack et al., 2008).

PET analyses

Co-registration and quality control

All procedures were performed according to an automatic protocol using the PMOD PNEURO tool (V. 3.5 PMOD Technologies, Zürich): PET images were rigidly co-registered to the corresponding MRI to calculate a linear transformation (PET-2-MRI). Individual MRI images were nonlinearly co-registered to the standard Montreal Neurological Institute (MNI) space MRI template (MRI-2-MNI), while PET-2-MRI and MRI-2-MNI transformations were used to transform PET images into the MNI space.

T1-weighted MR images were segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) (Ashburner and Friston, 2005) to generate a total of 83 individual brain VOIs for each subject in the MNI space, according to the atlas of Hammers (Hammers et al., 2003). Hammers Atlas VOIs were also defined and assessed without segmentation (FULL). In addition to the Hammers VOIs, three white matter VOIs were defined and tested: 1) WM, for which the area of individual cortical GM VOIs was subtracted from entire brain volume, thus segmenting all white matter voxels included in the whole atlas VOI (this, separately for cortex and cerebellum). 2) WM-seg, for which the segmentation process was used to generate WM VOIs containing all voxels declared as white matter in the segmentation (separately for cortex and cerebellum). 3) CSO, a manually drawn bilateral VOI for the centrum semiovale, of volume 2.4 cm^3. More detailed illustrations of the different white matter volumes are given in Supplement 2 (Fig. S1). Results for WM are reported in the manuscript, while results for WM-seg and CSO are provided in Supplement 2 (Tables S1–S4).

Inverse PET-2-MNI transformations were used to resample VOIs from the Hammers atlas to the native PET space, in which VOI calculations were subsequently performed. GM volumes were assessed in MNI space to account for different brain sizes. Further details of the analysis procedure are provided in Supplement 3 (Fig. S2).

All images were visually checked for correct co-registration and appropriate segmentation. From a total of 1018 (baseline) PET images, 36 (3.5%) were identified as misregistered, and were consequently processed semi-automatically by inclusion of hot spot alignment or masking of extracerebral structures prior to final co-registration. A total of 56 (5.5%) MRI segmentations were found to be inadequate (due failed GM/WM/CSF identification or visible artifacts) when calculated using the standard settings. These 56 subjects were necessarily excluded from further analyses, as changing standard parameters could have biased the resulting VOI configuration. For the 278 follow-up PET images, eight (3%) had to be processed semi-automatically and 20 (7.2%) were excluded due to failed GM/WM/CSF identification or visible artifacts (Fig. 1). Details of the resulting study groups including demographics are provided in Table 1.

VOI-based PVEC

A VOI-based PVEC (Rousset et al., 1998, 2008) was performed in PET space for individual GM VOIs from the Hammers Atlas (again obtained through multiplication by the inverse PET-2-MNI transformation). Background and white matter regions were included in the correction algorithm to ensure that all voxels included in the atlas were assigned to a VOI. These VOIs were subsequently used for calculation of PVE-corrected values by the geometric transfer matrix method,

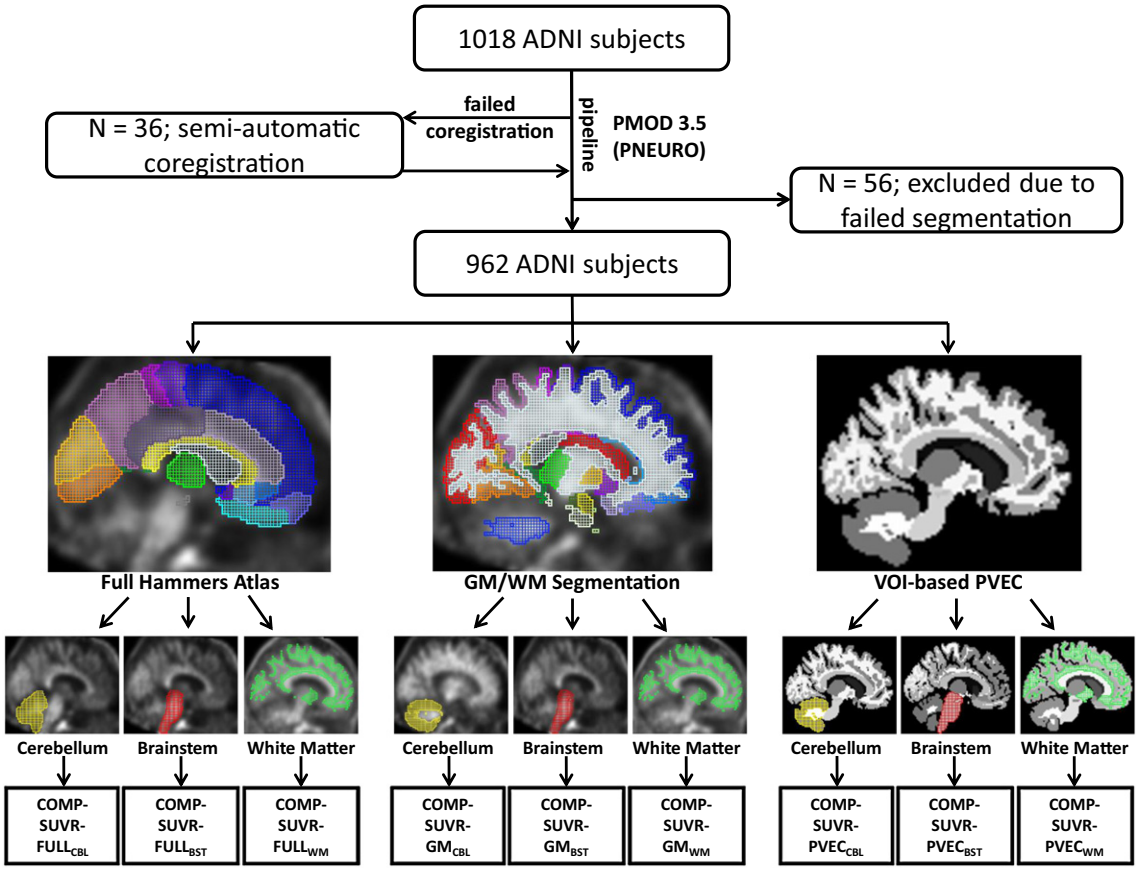

Fig. 1. Workflow from ADNI database to VOI-based analyses. All subjects receiving an [^{18}F]-AV45 PET scan and a T1 weighted MRI at baseline were processed using the full-automatic PNEURO (PMOD 3.5) pipeline. Subsequent visual inspection was followed either by manual coregistration when mis-registrations occurred, or by exclusion when segmentation errors or extensive artifacts were detected. Remaining subjects were analyzed using full atlas VOIs (Hammers et al., 2003), segmented GM VOIs and segmented GM VOIs after PVEC. Resulting VOI values were consecutively scaled by either the cerebellum (yellow), the brainstem (red) or the white matter (green). A tenth methodological configuration consisted of full atlas VOIs and scaling by the whole cerebellum and served as a reference (Johnson et al., 2013; Joshi et al., 2012).

Table 1
Demographics of healthy controls (HC), mild cognitive impaired patients (MCI) and patients suffering from Alzheimer's disease (AD) included in this study. Column 9 indicates subjects with available 2-year follow-up [^{18}F]-AV45-PET and MRI. * p< 0.05; ** p< 0.01; *** p< 0.001.

DX-group	Baseline N (Aß+, %)	Age (y ± SD)	Gender (f/m)	Education (y ± SD)	APOE ε4 (N of alleles ± SD)	ADAS-cog (0–70 ± SD)	MMSE (0–30 ± SD)	2-year follow up N (Aß+, %)
HC	316 (108, 32%)	74.6 ± 6.5	148 (47%) f/168 (53%) m	16.4 ± 2.6	0.31 ± 0.51	9.1 ± 4.6	29.0 ± 1.2	88 (26, 30%)
MCI	483 (267, 55%)	72.5 ± 7.9**	272 (57%) f/210 (43%) m	16.1 ± 2.7	0.56 ± 0.65*	14.7 ± 6.7***	28.0 ± 1.7**	148 (66, 45%)
AD	163 (143, 88%)	75.1 ± 7.7	99 (61%) f/64 (39%) m	16.0 ± 2.6	0.88 ± 0.71**	31.2 ± 9.5***	22.7 ± 3.2***	22 (16, 73%)

cited above. The gray matter threshold was set at 0.3 while a full-width-at-half-maximum (FWHM) of 8 mm was used, determined as the uniform spatial resolution after ADNI preprocessing. Both uncorrected and PVE-corrected results were generated (implemented in PNEURO) for individual GM VOIs. More details and visualization of the processing steps are provided in Supplement 3 (Fig. S2).

Assessment of VOI values

Following analyses of PET data were performed (Fig. 1):

a) Uncorrected and PVE-corrected standard uptake values (SUV) were calculated for a combined composite VOI (COMP) consisting of frontal, parietal, temporal cortex and precuneal/posterior cingulate gyrus. Cerebellum (CBL), brainstem (BST), and white matter (WM) were assessed as reference regions. COMP and CBL were assessed both as full atlas regions (FULL-VOI) and as segmented GM regions (GM-VOI). Additionally PVE-corrected values were obtained for the GM-VOI (PVEC-GM-VOI) of COMP and CBL.
b) Ten different composite VOI standard uptake value ratios (SUVR) were calculated: Composite FULL-VOI scaled either by (1) the cerebellar FULL-VOI (COMP-SUVR-FULL$_{CBL}$), (2) the brainstem VOI (COMP-SUVR-FULL$_{BST}$), or (3) the white matter VOI (COMP-SUVR-FULL$_{WM}$); composite GM-VOI scaled either by (4) the cerebellar GM-VOI (COMP-SUVR-GM$_{CBL}$), (5) the brainstem VOI (COMP-SUVR-GM$_{BST}$), or (6) the white matter VOI (COMP-SUVR-GM$_{WM}$); composite PVEC-GM-VOI scaled either by (7) the cerebellar PVEC-GM-VOI (COMP-SUVR-PVEC$_{CBL}$), (8) the brainstem VOI (COMP-SUVR-PVEC$_{BST}$), or (9) the white matter VOI (COMP-SUVR-PVEC$_{WM}$). The final configuration (10) consisted of the composite GM-VOI scaled by the cerebellar FULL-VOI (Johnson et al., 2013; Joshi et al., 2012), and served as validation of our analyses against this reference standard (COMP-SUVR-REF). Aβ-positivity/-negativity was obtained according to the COMP-SUVR-REF threshold of >/≤1.1 in ADNI database.
c) To ascertain the effect of VOI-based PVEC, changes in raw PVE-corrected VOI values (SUVs) were calculated relative to their uncorrected values (% change). GM volumes were extracted after segmentation to estimate the global GM atrophy.
d) Longitudinal changes after two years in [^{18}F]-AV45 uptake (Δ%-COMP-SUVR) were assessed for the ten composite SUVR according to Eq. (1):

$$\Delta\%-COMP-SUVR = \frac{[COMP-SUVR\ (2Y-Follow-Up)]-[COMP-SUVR\ (Baseline)]}{COMP-SUVR\ (Baseline)} \times 100\% \quad (1)$$

Statistics

a) Mean (± SD) SUV for the reference regions were calculated for HC, MCI and AD groups and compared using multivariate analysis of variance with subsequent Bonferroni correction (MANOVA).
b) Mean (± SD) COMP-SUVR were calculated for HC, MCI and AD groups according to all ten different evaluations listed above, and compared using MANOVA.
c) Mean (± SD) % changes after PVEC in GM-SUV of composite and cerebellum as well as GM volumes were compared between HC, MCI and AD groups using MANOVA.
d) ROCs were calculated for the discriminatory power between AD/HC, AD/MCI and MCI/HC for the ten composite SUVRs, using SPSS (version 22.0; SPSS, Chicago, IL). ROCs for the nine test configurations were contrasted against COMP-SUVR-REF using ROC_2_curves_tool version 6 (ACOMED statistic, Leipzig Germany) including the non-parametric approach of Delong (DeLong et al., 1988). [^{18}F]-AV45 cut-off values, sensitivities and specificities were calculated for revealing the best tradeoff between sensitivity and specificity for the discrimination between AD/HC, AD/MCI and MCI/HC.
e) Results from all 74 (33 HC, 39 MCI, 2 AD) non-amyloid accumulating subjects (those who were Aβ-negative in both [^{18}F]-AV45 PET assessments) were used to perform correlation analysis (Pearson's R) between baseline and follow-up COMP-SUVR for all ten different configurations with regard to longitudinal stability.
f) Mean (± SD) Δ%-COMP-SUVR were calculated by diagnosis groups for all ten COMP-SUVR evaluations. Paired t-tests were used for the comparison of different configurations against the COMP-SUVR-REF configuration. Mean (± SD) Δ%-volume of COMP and CBL were calculated by diagnosis groups and compared using MANOVA.

Different PET scanners (N = 17) of the multi-center study were included as a covariate for all group-wise calculations. For all statistical tests p-values < 0.05 were assigned to be significant.

Results

Demographics

See Table 1 for details of the study population. At baseline, 108/316 (32%) HC subjects, 267/483 (55%) MCI subjects and 143/163 (88%) AD subjects had a positive [^{18}F]-AV45 scan (SUVR > 1.10). From those subjects with an additional 2-year follow-up scan, 26/88 (30%) HC subjects, 66/148 (45%) MCI subjects and 16/22 (73%) AD subjects had a positive [^{18}F]-AV45 baseline scan. Nine converters to MCI and one converter to AD were observed in HC subjects. Five re-converters to HC and 12 converters to AD were found in the MCI group, while all AD subjects retained AD-positive status.

Comparison of reference region SUVs

All mean values (± SD) for the SUVs of reference regions by clinical group are provided in Table 2. Strong decreases in SUVs of AD subjects when compared to HC were observed for CBL-FULL-VOI (−4.3%;

Table 2
Mean (± SD) SUV for all reference regions (column 1) by diagnosis groups (columns 2–4). F-statistics and p-values for diagnosis related MANOVA are presented in columns 5–6. Significant differences against HC are indicated by: *p < 0.05; **p < 0.01; ***p < 0.001.

SUV region	HC	MCI	AD	F-statistic	p-value
CBL-FULL	1.09 ± 0.07	1.08 ± 0.07	1.04 ± 0.07***	23.8	<0.001
CBL-GM	1.03 ± 0.06	1.03 ± 0.05	1.02 ± 0.05*	3.5	<0.05
BST	1.63 ± 0.20	1.59 ± 0.19***	1.47 ± 0.18***	34.3	<0.001
WM	1.52 ± 0.23	1.54 ± 0.23	1.54 ± 0.25	1.1	n.s.

$p < 0.001$) and BST-VOI (−10.4%; $p < 0.001$), while a small but significant decrease was found for CBL-GM-VOI (−1.4%; $p < 0.05$) in this contrast. WM-VOI SUV did not show any significant differences between diagnosis groups.

Comparison of differently assessed composite SUVR

All mean values (± SD) for the ten differently assessed COMP-SUVRs by clinical group are provided in Table 3. Differences between MCI/HC and AD/HC were highly significant regardless of the assessment configuration used (all $p < 0.001$). The COMP-SUVRs were consistently lower when using the brainstem or the white matter reference when compared with cerebellum reference counterparts. Segmentation and PVEC application led to elevated differences in diagnosis-related COMP-SUVR. The most prominent distinctions between mean SUVR in both MCI/HC (+22%) and AD/HC (+70%) were observed when using COMP-SUVR-PVEC$_{BST}$ ($F = 183$, $p < 0.001$) or COMP-SUVR-PVEC$_{WM}$ (MCI/HC: +15%; AD/HC: +52%; $F = 182$, $p < 0.001$).

SUV changes after VOI-based PVEC

Magnitudes of % change after PVEC differed significantly between diagnosis groups for the GM segmented COMP-VOI ($F = 166$; $p < 0.001$), as SUV decreases were observed for HC (−18 ± 12%) and MCI subjects (−10 ± 15%) while SUV of AD subjects increased after PVEC (+7 ± 13%) (Fig. 2). Differences between the diagnosis groups regarding changes after PVEC were also found in the GM segmented CBL-VOI ($F = 5.1$, $p < 0.01$) with decreases in HC (−8 ± 5%), MCI (−7 ± 5%) and AD (−6 ± 5%) subjects. Mean GM volumes for COMP results were 485 ± 37 cm^3 (HC), 480 ± 41 cm^3 (MCI; $p = 0.08$ vs. HC) and 452 ± 37 cm^3 (AD; $p < 0.001$ vs. HC and MCI). Corresponding GM volumes for CBL resulted in 127 ± 9 cm^3 (HC), 128 ± 8 cm^3 (MCI) and 124 ± 8 cm^3 (AD; $p < 0.001$ vs. HC and MCI).

ROC analyses

COMP-SUVR-REF values showed a high correlation ($R = 0.98$; $p < 0.001$) with the corresponding available values obtained from the ADNI database, which were likewise derived from computation of a GM composite target VOI scaled by the whole cerebellum (Johnson et al., 2013; Joshi et al., 2012). Slightly higher values were obtained in our analyses, which have a linear relation defined by $y = 1.07x + 0.03$ with respect to different segmentation methods.

All values for ROC analyses and comparisons are detailed in Table 4. When using COMP-SUVR-REF values, ROC analyses gave an AUC of 0.823 (95%-CI: 0.786–0.856) for the contrast between AD and HC. Respective AUCs for the MCI/HC and the AD/MCI contrasts were 0.622 (95%-CI: 0.587–0.656) and 0.717 (95%-CI: 0.680–0.751), respectively.

When contrasting AD against HC by method, the highest AUC of 0.907 (95%-CI: 0.876–0.938) was found using the white matter VOI as a reference region and VOI-based PVEC performed after GM-segmentation (COMP-SUVR-PVEC$_{WM}$). The ROC comparison of COMP-SUVR-PVEC$_{WM}$ differed very significantly from that of COMP-SUVR-REF ($p < 0.00001$) (Fig. 3A). The AUC for the brainstem reference region was 0.897 (95%-CI: 0.864–0.929) when using VOI-based PVEC after GM segmentation ($p < 0.00001$ vs. COMP-SUVR-REF) (Fig. 3B), but only 0.778 (95%-CI: 0.738–0.815) when using full atlas VOIs without PVEC and the whole cerebellum as a reference region ($p < 0.00001$ vs. COMP-SUVR-REF) (Fig. 3C).

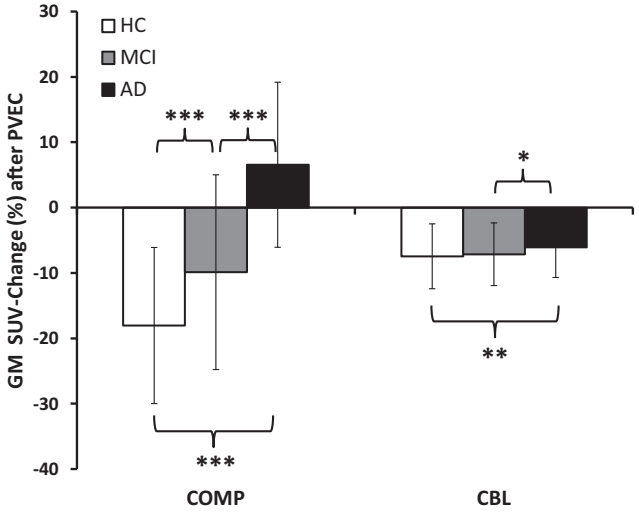

Fig. 2. SUV changes after PVEC. % (± SD) changes after VOI-based partial volume effect correction (PVEC) in the gray matter segmented composite VOI (COMP) and the cerebellar gray matter (CBL). Mean values are given for HC (white; $N = 316$), MCI (gray; $N = 483$) and AD (black; $N = 163$) subjects. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

For the contrast of MCI against HC, the AUC for the white matter reference region was 0.651 (95%-CI: 0.613–0.688) when using VOI-based PVEC after GM segmentation (p = n.s. vs. COMP-SUVR-REF) (Fig. 3D). The highest AUC of 0.658 (95%-CI: 0.620–0.696) was obtained through use of brainstem as reference region and performance of VOI-based PVEC after GM-segmentation (COMP-SUVR-PVEC$_{BST}$). The ROC comparison of COMP-SUVR-PVEC$_{BST}$ against COMP-SUVR-REF revealed a significant difference ($p < 0.01$) (Fig. 3E), while the AUC was as low as 0.610 (95%-CI: 0.575–0.644) when using full atlas VOIs without PVEC and the whole cerebellum as a reference region (p = n.s. vs. COMP-SUVR-REF) (Fig. 3F).

Longitudinal analyses

Findings in stable Aβ-negative subjects

Correlation analyses in 74 stable Aβ-negative subjects revealed lowest agreement between baseline and two year follow-up COMP-SUVR for the cerebellar reference region (FULL: $R = 0.68$; GM: $R = 0.55$; PVEC: $R = 0.60$; Fig. 4A). Scaling to the brainstem uptake resulted at

Table 3
Mean (± SD) COMP-SUVR when assessed by the ten different methodological configurations (column 1) by diagnosis groups (columns 2–4) as well as % differences between diagnosis groups (columns 5–6). F-statistics and p-values for diagnosis related MANOVA are presented in columns 7–8.

Configuration	HC	MCI	AD	MCI vs. HC	AD vs. HC	F-statistic	p-value
COMP-SUVR-FULL$_{CBL}$	1.29 ± 0.17	1.36 ± 0.20	1.48 ± 0.21	5.6%	14.9%	54.4	<0.001
COMP-SUVR-FULL$_{BST}$	0.86 ± 0.11	0.93 ± 0.14	1.05 ± 0.16	7.6%	21.9%	106.6	<0.001
COMP-SUVR-FULL$_{WM}$	0.92 ± 0.04	0.95 ± 0.06	1.01 ± 0.04	3.3%	9.8%	143.6	<0.001
COMP-SUVR-GM$_{CBL}$	1.29 ± 0.20	1.39 ± 0.25	1.55 ± 0.25	7.8%	21.0%	71.3	<0.001
COMP-SUVR-GM$_{BST}$	0.82 ± 0.13	0.90 ± 0.17	1.08 ± 0.18	10.3%	31.9%	146.8	<0.001
COMP-SUVR-GM$_{WM}$	0.88 ± 0.08	0.92 ± 0.10	1.03 ± 0.08	4.5%	17.0%	159.1	<0.001
COMP-SUVR-PVEC$_{CBL}$	1.16 ± 0.35	1.38 ± 0.47	1.79 ± 0.47	19.0%	53.9%	111.0	<0.001
COMP-SUVR-PVEC$_{BST}$	0.69 ± 0.20	0.84 ± 0.28	1.17 ± 0.29	22.2%	70.2%	183.4	<0.001
COMP-SUVR-PVEC$_{WM}$	0.73 ± 0.17	0.85 ± 0.23	1.11 ± 0.21	15.1%	52.1%	181.9	<0.001
COMP-SUVR-REF	1.22 ± 0.19	1.32 ± 0.24	1.52 ± 0.24	8.3%	24.5%	96.0	<0.001

Table 4
ROC calculations for AD/HC (A), MCI/HC (B) and AD/MCI (C) discrimination by the ten different methodological configurations (column 1). Respective areas under the curve (AUC) with standard errors (SE) and 95% confidence interval (CI) are depicted in columns 2–5. All ROCs from our analyses were compared (DeLong et al., 1988) against the widely used reference method COMP-SUVR-REF as (column 6). [^{18}F]-AV45 cut-off values, sensitivities and specificities (best tradeoff) for the discrimination between diagnosis groups are provided in columns 7–9.

Configuration	AUC	SE	Lower 95% CI	Higher 95% CI	p-value	CUT-OFF	Sensitivity	Specificity
A) AD vs HC								
COMP-SUVR-REF	0.823	0.022	0.786	0.856		1.36	80%	80%
COMP-SUVR-FULL$_{CBL}$	0.778	0.024	0.738	0.815	0.00000	1.36	75%	75%
COMP-SUVR-GM$_{CBL}$	0.799	0.023	0.760	0.834	0.00000	1.39	78%	78%
COMP-SUVR-PVEC$_{CBL}$	0.848	0.020	0.813	0.879	0.00065	1.42	82%	82%
COMP-SUVR-FULL$_{BST}$	0.826	0.022	0.789	0.859	0.76820	0.95	81%	81%
COMP-SUVR-GM$_{BST}$	0.860	0.020	0.825	0.890	0.00105	0.94	82%	82%
COMP-SUVR-PVEC$_{BST}$	0.897	0.017	0.864	0.929	0.00000	0.89	85%	83%
COMP-SUVR-FULL$_{WM}$	0.902	0.016	0.870	0.933	0.00000	0.97	85%	85%
COMP-SUVR-GM$_{WM}$	0.907	0.016	0.877	0.938	0.00000	0.96	84%	84%
COMP-SUVR-PVEC$_{WM}$	0.907	0.016	0.876	0.938	0.00000	0.93	85%	85%
B) MCI vs HC								
COMP-SUVR-REF	0.622	0.020	0.587	0.656		1.19	61%	61%
COMP-SUVR-FULL$_{CBL}$	0.610	0.020	0.575	0.644	0.05198	1.27	59%	58%
COMP-SUVR-GM$_{CBL}$	0.621	0.020	0.586	0.654	0.75479	1.25	60%	61%
COMP-SUVR-PVEC$_{CBL}$	0.639	0.020	0.604	0.672	0.04346	1.07	61%	60%
COMP-SUVR-FULL$_{BST}$	0.641	0.019	0.606	0.674	0.15284	0.87	60%	63%
COMP-SUVR-GM$_{BST}$	0.647	0.019	0.613	0.681	0.03741	0.80	64%	60%
COMP-SUVR-PVEC$_{BST}$	0.658	0.019	0.620	0.696	0.00922	0.65	62%	62%
COMP-SUVR-FULL$_{WM}$	0.632	0.019	0.594	0.670	0.53321	0.93	60%	58%
COMP-SUVR-GM$_{WM}$	0.632	0.019	0.594	0.670	0.54871	0.88	60%	58%
COMP-SUVR-PVEC$_{WM}$	0.651	0.019	0.613	0.688	0.07752	0.72	62%	60%
C) AD vs MCI								
COMP-SUVR-REF	0.717	0.025	0.680	0.751		1.45	69%	70%
COMP-SUVR-FULL$_{CBL}$	0.674	0.026	0.636	0.710	0.00000	1.41	65%	64%
COMP-SUVR-GM$_{CBL}$	0.691	0.025	0.653	0.726	0.00000	1.47	66%	66%
COMP-SUVR-PVEC$_{CBL}$	0.731	0.024	0.694	0.764	0.02492	1.61	70%	69%
COMP-SUVR-FULL$_{BST}$	0.724	0.025	0.688	0.758	0.60534	1.01	68%	68%
COMP-SUVR-GM$_{BST}$	0.760	0.024	0.725	0.793	0.00025	1.02	72%	72%
COMP-SUVR-PVEC$_{BST}$	0.792	0.021	0.751	0.833	0.00000	0.92	74%	74%
COMP-SUVR-FULL$_{WM}$	0.792	0.020	0.753	0.830	0.00000	1.00	74%	74%
COMP-SUVR-GM$_{WM}$	0.805	0.019	0.767	0.843	0.00000	1.00	74%	73%
COMP-SUVR-PVEC$_{WM}$	0.803	0.020	0.764	0.842	0.00000	1.03	74%	74%

an intermediate correlation (FULL: R = 0.82; GM: R = 0.80; PVEC: R = 0.81; Fig. 4B), while the WM reference showed the highest agreement (FULL: R = 0.92; GM: R = 0.93; PVEC: R = 0.87; Fig. 4C). The COMP-SUVR-REF configuration showed an R of 0.58.

Changes in GM volume and longitudinal [^{18}F]-AV45 uptake

To two-year follow-up, GM Δ%-volumes for COMP decreased −1.0 ± 4.3% in HC, −1.9 ± 3.0% in MCI and −3.3 ± 4.4% in AD ($p < 0.01$ vs. HC). Corresponding GM Δ%-volumes for CBL decreased −1.3 ± 2.8% in HC, −1.4 ± 2.2% in MCI and −1.8 ± 3.2% in AD (all n.s.). Low longitudinal changes over two years were found in reference region SUVs of CBL-FULL-VOI (+0.1 −0.2%), CBL-GM-VOI (+0.3–0.6%), BST-VOI (+0.2–0.7%), and WM-VOI (+0.1–0.2%), with no differences between diagnosis groups. Longitudinal SUV decreases were observed in a small subset of nine converting HC (CBL-FULL: −2.0 ± 2.6%; CBL-GM: −1.2 ± 1.8%; BST: −4.3 ± 6.2%; WM: −4.7 ± 4.4%) and twelve converting MCI (CBL-FULL: −1.3 ± 2.8%; CBL-GM: −0.1 ± 2.2%; BST: −2.4 ± 6.3%; WM: −2.5 ± 8.1%) subjects. All mean [^{18}F]-AV45 changes (% ± SD) to follow-up for the ten differently-assessed Δ-COMP-SUVRs are presented in Table 5. Greatest increases over time were observed when using PVEC regardless of the reference region (Fig. 5A). When only Aβ-positive subjects were considered, highest increases were again found when using PVEC (Fig. 5B). Inter-subject variability was consequently lowest for white matter, intermediate for brainstem and highest for cerebellum as reference region (Fig. 5A + B).

Discussion

We present a systematic investigation of the discriminatory performance of different reference regions for [^{18}F]-AV45-PET quantitation, making use of the hitherto largest dataset of PET examinations in AD patients, along with MCI and HC groups. In addition, we tested effects of additionally performed MRI-based segmentation and VOI-based PVEC in our nearly 1000 cases. ROC analyses revealed that the discriminatory power between MCI or AD and the healthy control population was increased by using the white matter or the brainstem rather than cerebellum as reference region, and was also increased by application of a VOI-based PVEC. However, for the particular case of longitudinal analyses, white matter emerged as the best-suited reference region, due to the relatively low group variability it provided. Furthermore, the expected increases with disease progression in cortical amyloidosis were best captured using additional GM segmentation and PVEC.

Reference region

White matter has recently been evaluated as a reference region for [^{18}F]-FDDNP PET in a small sample of AD patients (Wong et al., 2010). The authors of that study concluded that more accurate and less biased binding estimates could be established when using a white matter reference instead of the cerebellum. In our study, cerebellum and brainstem reference quantitation showed significantly lower SUV in AD subjects when compared to HC. In this contrast, which is critical for monitoring early disease progression, the white matter reference gave the more stable SUV results between diagnosis groups at baseline, and therefore performed best with regard to invariance with the investigated pathology, i.e. GM amyloidosis.

The lower [^{18}F]-AV45 SUVs observed in cerebellum and brainstem were not expected in AD subjects, based on knowledge of the distribution of amyloid, and thus need to be further investigated. Interestingly, we found decreases in all reference region SUVs in the group of subjects

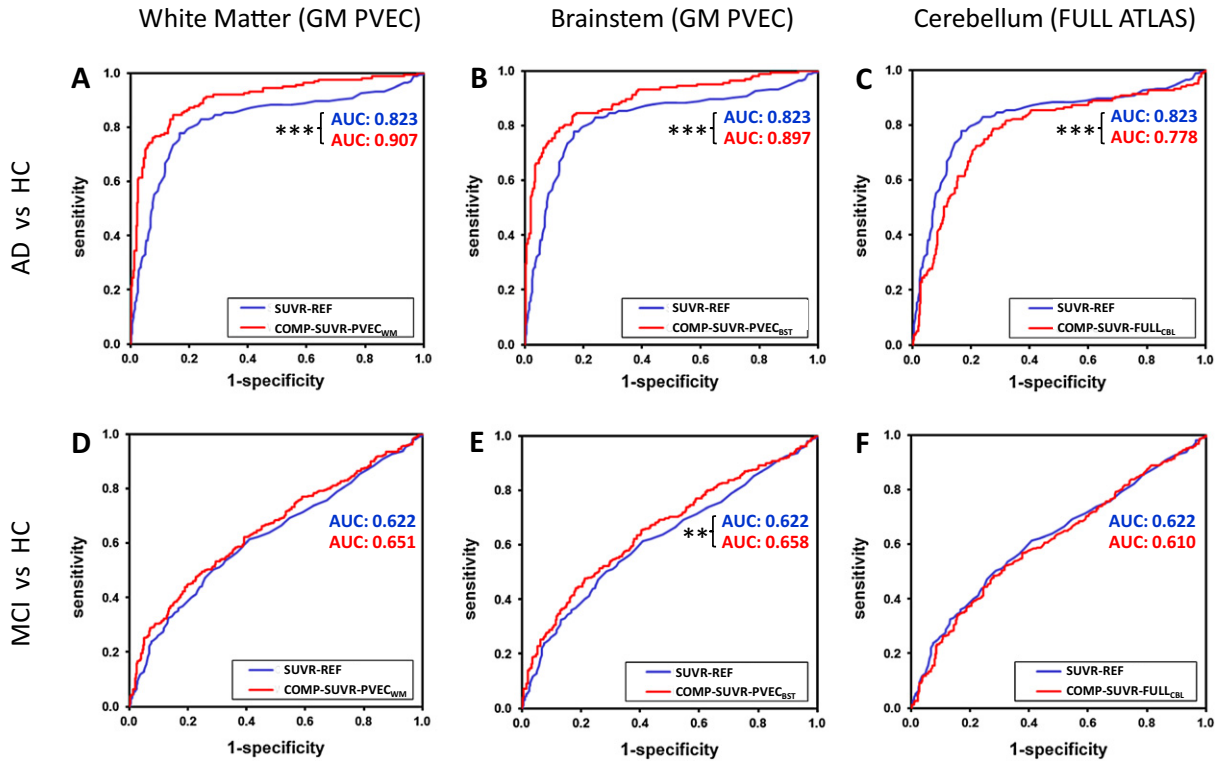

Fig. 3. ROC analyses for AD/HC and MCI/HC discrimination. ROCs were calculated and compared for discriminations by [^{18}F]-AV45 uptake between AD and HC (upper row, A–C) as well as between MCI and HC (lower row, D–F) subjects. Exemplary ROCs for COMP-SUVR-PVEC$_{WM}$ (column 1), COMP-SUVR-PVEC$_{BST}$ (column 2) and COMP-SUVR-FULL$_{CBL}$ (column 3) configurations (all red) were consequently contrasted against the COMP-SUVR-REF configuration (blue). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

converting from HC to MCI (N = 9) or MCI to AD (N = 12). However these data are not in accord with stable results between diagnosis groups in the WM SUV, as this region was affected as well, a phenomenon which is probably related to the small sample size of converters. A higher inter-subject variation (SD: 15.1%) in WM SUV was observed when compared to brainstem SUV (SD: 12.1%) or cerebellum SUV (SD: 5.8%) regardless of the diagnosis group. From this side most stable cortical quantification would have been expected from the cerebellar reference. However, the longitudinal correlation of COMP-SUVR between baseline and 2-year follow-up in 74 Aβ-negative subjects was by far the highest when using the WM reference region (Fig. 4). These findings in presumably pathology free patients support the suitability of WM reference for a stable cortical quantification within individual subjects.

Interestingly, the WM reference region including not only true WM voxels performed better than did the WM-seg reference region. For example, the CSF space in the depths of sulci was present in the WM VOI configuration, arising from the volumes in the full atlas. Therefore, some extra-cerebral volumes contributed to those "non-true" WM voxels. The longitudinal investigation of amyloid negative subjects resulted in a R of 0.85 when correlating baseline versus 2-year follow-up of COMP-GM-SUV/"non-true" WM. We note that using these "non-true" WM voxels as a stand-alone reference region did not result in a performance level quite as high as for WM (which had an R = 0.93), but that the correlation was still higher than we had expected. We suppose that WM reference region may be stabilized by inclusion of the "non-true" WM voxels lying adjacent to the GM target region. In particular, spill-over from variable amounts of nonspecific GM uptake may be counterbalanced by including voxels adjacent to the GM (e.g. outer CSF space) in the reference region. Alternately, we suppose that inter-subject variability in nonspecific [^{18}F]-AV45 binding in "non-true" WM voxels would influence the signal attributed to GM.

A recent methodologically-oriented amyloid PET study comparing [^{18}F]-AV45, [^{18}F]-flutemetamol and [^{11}C]-PiB has also calculated composite SUVR with a brainstem reference (Landau et al., 2014). The [^{18}F]-AV45 findings in their 32 cases showed consequently lower composite SUVR$_{BST}$ compared to SUVR$_{CBL}$, which is in line with our results, as were also their estimates for [^{18}F]-AV45 in mostly MCI subjects for the magnitudes of SUVR$_{BST}$ (0.88 vs. 0.90) and SUVR$_{CBL}$ (1.38 vs. 1.39). In another methods study, Edison showed that a target-to-pons ratio for [^{11}C]-PiB offered reliable estimations of the cortical radiotracer uptake, especially in cases where use of the cerebellum is not appropriate, such as in familial AD (Edison et al., 2012). Although the rather low target-to-WM or target-to-BST ratios, which were frequently < 1, may perplex the observer, we can only agree with these earlier findings, as we likewise obtained very stable and robust results for all assessed COMP-SUVR$_{WM}$ or COMP-SUVR$_{BST}$, which gave an equal relative standard deviation when compared to COMP-SUVR$_{CBL}$.

Our ROC analyses proved better discrimination between the present large groups of AD, MCI and HC subjects using the white matter or the brainstem as a reference region (Fig. 3). Indeed, the differentiation of these clinical groups is an important precondition for preventive therapy trials. Another important observation of our study was the notably lower AUC ($p < 0.001$ compared to COMP-SUVR-REF) when plain atlas VOIs without GM segmentation or PVEC were used, a finding which was also found for whole cerebellum as a reference tissue. Use of precisely this approach therefore probably lowers the discriminatory power between MCI/HC and AD/HC, when compared to the analysis strategies investigated for the initial validation of this tracer (Clark et al., 2012; Johnson et al., 2013). The use of a brainstem reference with full atlas VOIs showed comparable results with COMP-SUVR-REF, and can therefore serve adequately even when segmentation/PVEC are not available (e.g. without MRI information).

VOI-based PVEC

In an extensive review, Erlandsson emphasized the effects of marked cortical atrophy in neurodegenerative diseases, which lead to exaggerated PVE when performing brain PET studies (Erlandsson

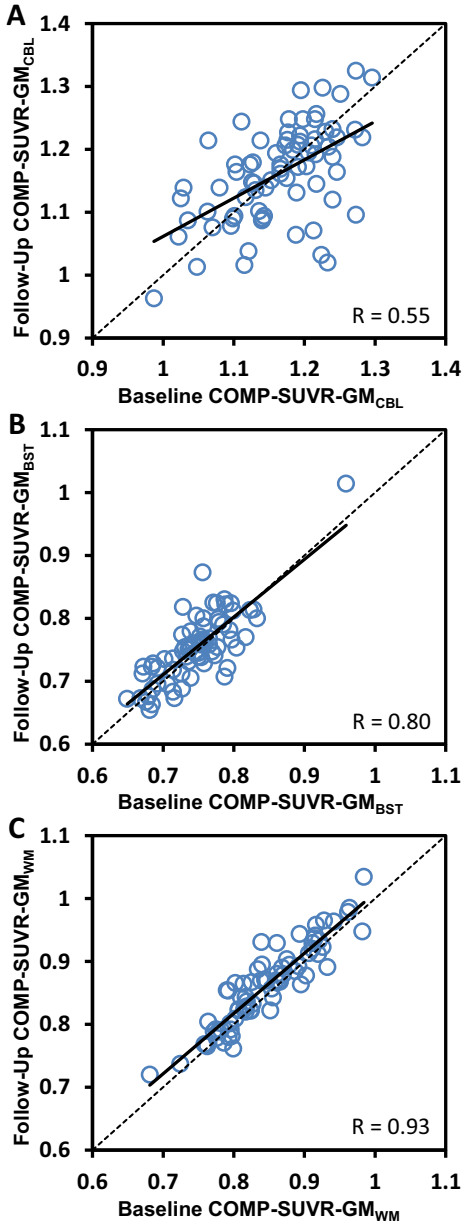

Fig. 4. Longitudinal findings in stable Aβ-negative subjects. Agreement between baseline and follow-up COMP-SUVR-GM for 74 (N = 33 HC, N = 39 MCI, N = 2 AD) stable non-amyloid accumulating subjects. Scaling was performed either by the cerebellar gray matter (A), the brainstem (B) or the white matter (C). Trend lines of correlation are illustrated continuously while dotted lines represent assumed perfect identity. R give estimated Pearson's coefficients of correlation.

et al., 2012). Even apparent hypometabolism in normal aging as detected by [^{18}F]-FDG PET was partially recoverable by MRI-based PVEC (Curiati et al., 2011), and is therefore likely in large part a bias due to atrophy and loss of signal. On the other hand, age-related increases in oxygen extraction fraction are likely robust to atrophy and PVE (Aanerud et al., 2012).

For [^{18}F]-flutemetamol amyloid PET data, a region-based voxel-wise PVEC proved suitable (Thomas et al., 2011) after validation by phantom experiments. The authors noted that not only atrophy and the spill-over from a region, but also the spill-in must be considered when performing PVEC. They concluded that techniques accounting for multiple regions and tissue types are more appropriate for amyloid PET, especially given that significant WM uptake is a well-known attribute of [^{18}F]-flutemetamol, [^{18}F]-AV45 and [^{18}F]-florbetaben. We concur with their position, and recently undertook a microPET study testing effects of VOI-based PVEC of [^{18}F]-florbetaben uptake in brain of AD transgenic mouse brain (Brendel et al., 2014). We were able to verify by gamma counting and autoradiography ex vivo (where PVE is avoided) that distinct overcorrections occur when VOI configurations for the mouse brain lack WM and background regions; these regions were consequently included in the VOI-based PVEC approach of human brain in the current work, which yielded plausible values and changes after PVEC.

Present results indicate decreased mean changes of cortical SUV in amyloid tracer measurements after PVEC in HC and MCI subjects, whereas AD subjects showed an increase, no doubt due to their greater atrophy. Concerning the high white matter uptake of [^{18}F]-AV45, which is present irrespective of brain amyloidosis, this observation also makes sense; in most HC cases, the low cortical tracer uptake is located adjacent to the distinctly higher tracer uptake in underlying WM, which leads to net spill-in to GM in cases of low-moderate atrophy. A VOI-based PVEC approach including a white matter VOI accommodates this, and consequently leads to a decrease after PVEC in HCs. On the other hand, AD subjects show cortical [^{18}F]-AV45 uptake equal to or exceeding uptake in the subjacent WM, in conjunction with severe cortical atrophy. Therefore recovery of [^{18}F]-AV45 signal in GM necessarily increases after PVEC. MCI subjects fall somewhere between these two extremes, and in the present investigation showed as a group a small net decrease after PVEC, which is probably related to a quite high proportion of Aβ-negative cases (45%) in MCI groups.

Impact on longitudinal approaches

Longitudinal amyloid imaging recently showed high inter-subject variability when methodologically investigated with [^{11}C]-PiB PET to mean 2.5-year follow-up, and some AD subjects even showed an apparent decrease in amyloid burden over time when assessed by SUVR methods, which seems at odds with the known pathophysiology of disease progression (van Berckel et al., 2013). However, this seeming decrease was not apparent when using the full dynamic data set for calculation of binding potentials (BP_{ND}), suggesting blood flow changes between scans. We likewise observed high relative standard deviations in Δ%-COMP-SUVR for [^{18}F]-AV45 with two years of progression in our cohort. Two relevant effects were observed in our methodological comparison: First, the use of individual segmented GM VOIs (scaled by CBL and BST) showed higher Δ%-COMP-SUVR compared to full atlas VOIs, while additional performance of the PVEC substantially increased this effect, leading to a mean Δ%-COMP-SUVR of approximately 5.5% over two years for amyloid accumulation in HC and MCI subjects. Second, the use of the white matter reference distinctly lowered the relative standard deviations of diagnosis-related Δ%-COMP-SUVR group means, thus favoring the discrimination by group. The results of the hitherto largest longitudinal [^{11}C]-PiB PET study indicated an increase of 1.5% per year in HC and MCI subjects with low [^{11}C]-PiB levels and 3.1% per year in HC and MCI subjects with high [^{11}C]-PiB retention with cerebellar reference and without use of PVEC (Villemagne et al., 2013). In our analysis, the maximum mean [^{18}F]-AV45 increase observed in configurations without PVEC and with a cerebellar reference did not exceed 1.5% per year for amyloid accumulation in HC and MCI subjects. Taking into consideration the comparably high WM uptake of [^{18}F]-AV45, we can speculate that the detection of real longitudinal differences is suppressed by both spillover and progressing atrophy when PVEC is not performed. Thus, our findings lead us to conjecture that the fitness of quantitative longitudinal [^{18}F]-AV45 assessment over a brief two year period mainly benefits from the use of PVEC and secondarily from stabilization by the WM reference, which propagates to reduced inter-subject variability, i.e. improved precision. Although doubtless significant in planned intervention studies, we are aware that such differences are of minor relevance to routine clinical decisions with rating of Aβ-positivity/Aβ-negativity.

Table 5
Mean (± SD) % changes over two years in [^{18}F]-AV45 uptake when assessed by the ten different COMP-SUVR methodological configurations (column 1) for the three diagnosis groups (columns 2–4) for all subjects. Columns 5–7 display corresponding results when only Aβ-positive subjects were considered. Significant differences of all different configurations when contrasted against the COMP-SUVR-REF configuration are indicated by: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Configuration	All subjects			Aβ-positive subjects		
	HC (N = 88)	MCI (N = 148)	AD (N = 22)	HC (N = 26)	MCI (N = 66)	AD (N = 16)
2-year-Δ-COMP-SUVR-FULL$_{CBL}$	1.0 ± 5.5%***	0.6 ± 6.3%***	−0.4 ± 8.8%**	2.1 ± 7.2%***	0.9 ± 7.7%***	−1.2 ± 7.7%***
2-year-Δ-COMP-SUVR-FULL$_{BST}$	1.0 ± 4.3%	0.5 ± 4.5%*	−1.5 ± 4.3%	2.4 ± 5.0%	1.2 ± 4.8%	−1.1 ± 4.7%
2-year-Δ-COMP-SUVR-FULL$_{WM}$	0.9 ± 1.6%	1.2 ± 1.7%	1.1 ± 1.3%	1.0 ± 1.2%	1.6 ± 1.6%	1.4 ± 1.1%
2-year-Δ-COMP-SUVR-GM$_{CBL}$	1.6 ± 6.3%	1.3 ± 7.0%***	0.6 ± 9.6%	2.9 ± 8.3%	1.8 ± 8.5%***	−0.6 ± 8.5%*
2-year-Δ-COMP-SUVR-GM$_{BST}$	1.9 ± 4.5%	1.7 ± 4.7%	−0.1 ± 4.3%	3.5 ± 5.3%	2.9 ± 4.9%	0.6 ± 4.6%
2-year-Δ-COMP-SUVR-GM$_{WM}$	1.7 ± 3.0%	2.4 ± 3.4%	2.4 ± 2.9%	2.1 ± 2.5%	3.3 ± 3.5%	3.0 ± 2.4%
2-year-Δ-COMP-SUVR-PVEC$_{CBL}$	4.6 ± 9.9%***	4.2 ± 10.6%***	4.2 ± 12.7%*	6.5 ± 12.7%*	5.4 ± 12.2%***	2.7 ± 12.7%
2-year-Δ-COMP-SUVR-PVEC$_{BST}$	3.6 ± 7.5%	3.3 ± 7.1%*	1.6 ± 6.3%	5.6 ± 6.5%	5.2 ± 7.1%**	2.8 ± 6.0%
2-year-Δ-COMP-SUVR-PVEC$_{WM}$	3.4 ± 6.5%	4.0 ± 6.7%***	4.2 ± 5.9%	4.2 ± 4.2%	5.5 ± 6.3%*	5.1 ± 4.2%*
2-year-Δ-COMP-SUVR-REF	1.9 ± 5.4%	1.8 ± 6.3%	1.0 ± 8.2%	3.2 ± 7.5%	2.6 ± 7.8%	0.5 ± 7.7%

Limitations

Data included in this study derived from multiple centers. Therefore different quality of PET and MRI data was obtained and could only partially be balanced by image preprocessing. However it has to be considered that the same study conditions will likely occur in multicenter treatment trials. Furthermore, the 17 different PET scanners were included as a covariate to adjust for the possibility of instrumentation bias, which resulted in only small changes (<1%) between adjusted and non-adjusted data.

The detection of longitudinal reductions of amyloid burden, as seen with the SUVR method in some configurations, might suffer from interference from inadequate quantitation, since blood-flow changes between scans might not be properly accommodated. However, in large-scale studies, the use of SUVR is a convenient proxy for quantification of brain amyloid burden. Ultimately, the magnitude of this bias could only be answered by direct assessment of dynamic datasets in patients studied longitudinally, or through concomitant quantitative CBF studies. It seems premature to condemn SUVRs entirely, insofar as SUVRs using the three [^{18}F]-amyloid tracers have proven to correlate well with histopathological comparisons, as assessed in the development of these analyses.

Diagnosis groups of HC, MCI and AD were obtained by clinical assessment, with no histopathological gold standard information available. Therefore discriminatory analyses derive from detailed neuropsychiatric test batteries and biomarkers, with as yet no final validation. Furthermore, the detected longitudinal increases of amyloid burden over time could only be compared with previous findings and extrapolations of a large prospective PET study (Villemagne et al., 2013), since histopathologically verified amyloid accumulation over time is still lacking.

As not all extracortical regions in proximity to the brain (e.g. fat, bone, scalp) were included in the VOI-based algorithm used for PVEC, spillover from these sources may possibly contribute to the attribution of signal from brain VOIs; however, this potential bias seems likely to be of low magnitude since the distance from extracerebral structures is in the order of the FWHM. Our findings are for [^{18}F]-AV45, and may not generalize for other amyloid radioligands, which will have to be investigated in this context in further studies.

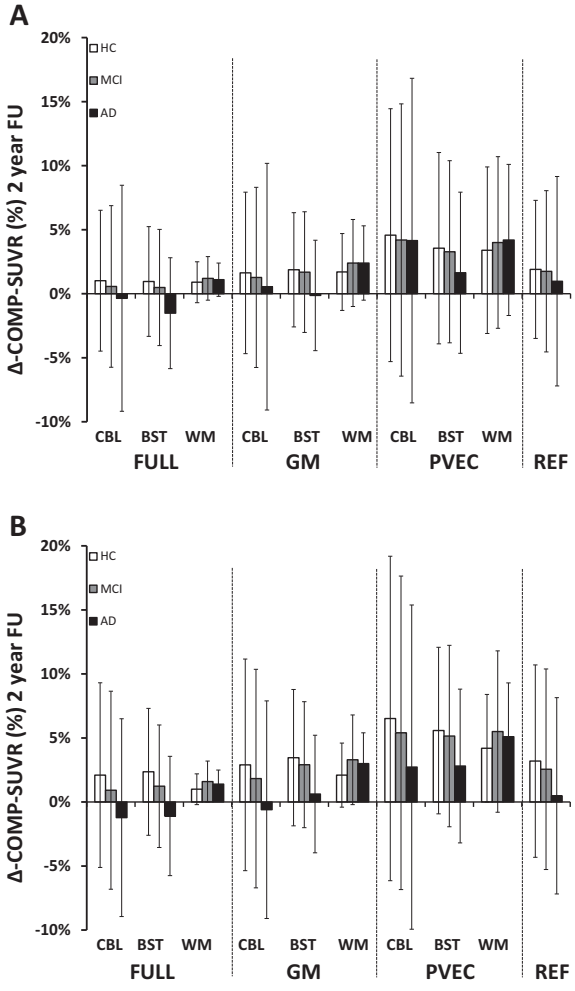

Fig. 5. Changes in longitudinal [^{18}F]-AV45 uptake. Longitudinal % changes in [^{18}F]-AV45 uptake (± SD) as assessed for HC (white; N = 88), MCI (gray; N = 148) and AD (black; N = 22) subjects and calculated for the ten COMP-SUVR configurations (A), as well as corresponding longitudinal % changes in [^{18}F]-AV45 uptake (± SD) when only Aß positive subjects were considered (B). Either full atlas VOIs (FULL), gray matter segmented VOIs (GM) or gray matter segmented VOIs after partial volume effect correction (PVEC) were used and scaled by either the cerebellum (CBL), the brainstem (BST) or the white matter (WM). An additional standard configuration consisted of GM segmented VOIs scaled by the whole cerebellum (REF).

Conclusion

We aimed to optimize the discriminative fitness of [^{18}F]-AV45-PET by methodological considerations of cortical atrophy/spill-over and comparison of reference regions. We found clear benefits from using VOI-based PVEC and defining the white matter or the brainstem rather than cerebellum as the reference region; both procedures increased the discriminatory power between MCI and HC, which are key target groups in therapy trials. Furthermore, longitudinal increases in amyloidosis during follow-up for two years were likely more accurately captured by application of the present optimized procedure with the white matter reference.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.11.055.

Acknowledgments

Data collection and sharing for this project was funded 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: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to Rev December 5, 2013 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The authors acknowledge Inglewood Biomedical Editing for professional editing of the manuscript.

References

Aanerud, J., Borghammer, P., Chakravarty, M.M., Vang, K., Rodell, A.B., Jonsdottir, K.Y., Moller, A., Ashkanian, M., Vafaee, M.S., Iversen, P., Johannsen, P., Gjedde, A., 2012. Brain energy metabolism and blood flow differences in healthy aging. J. Cereb. Blood Flow Metab. 32, 1177–1187.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Bohnen, N.I., Djang, D.S., Herholz, K., Anzai, Y., Minoshima, S., 2012. Effectiveness and safety of ^{18}F-FDG PET in the evaluation of dementia: a review of the recent literature. J. Nucl. Med. 53, 59–71.

Borghammer, P., Cumming, P., Aanerud, J., Forster, S., Gjedde, A., 2009. Subcortical elevation of metabolism in Parkinson's disease—a critical reappraisal in the context of global mean normalization. Neuroimage 47, 1514–1521.

Braak, H., Braak, E., 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 82, 239–259.

Brendel, M., Delker, A., Rotzer, C., Boning, G., Carlsen, J., Cyran, C., Mille, E., Gildehaus, F.J., Cumming, P., Baumann, K., Steiner, H., Haass, C., Herms, J., Bartenstein, P., Rominger, A., 2014. Impact of partial volume effect correction on cerebral beta-amyloid imaging in APP-Swe mice using [(18)F]-florbetaben PET. Neuroimage 84, 843–853.

Clark, C.M., Pontecorvo, M.J., Beach, T.G., Bedell, B.J., Coleman, R.E., Doraiswamy, P.M., Fleisher, A.S., Reiman, E.M., Sabbagh, M.N., Sadowsky, C.H., Schneider, J.A., Arora, A., Carpenter, A.P., Flitter, M.L., Joshi, A.D., Krautkramer, M.J., Lu, M., Mintun, M.A., Skovronsky, D.M., 2012. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study. Lancet Neurol. 11, 669–678.

Curiati, P.K., Tamashiro-Duran, J.H., Duran, F.L., Buchpiguel, C.A., Squarzoni, P., Romano, D.C., Vallada, H., Menezes, P.R., Scazufca, M., Busatto, G.F., Alves, T.C., 2011. Age-related metabolic profiles in cognitively healthy elders: results from a voxel-based [^{18}F]fluorodeoxyglucose-positron-emission tomography study with partial volume effects correction. AJNR Am. J. Neuroradiol. 32, 560–565.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845.

Dukart, J., Perneczky, R., Forster, S., Barthel, H., Diehl-Schmid, J., Draganski, B., Obrig, H., Santarnecchi, E., Drzezga, A., Fellgiebel, A., Frackowiak, R., Kurz, A., Muller, K., Sabri, O., Schroeter, M.L., Yakushev, I., 2013. Reference cluster normalization improves detection of frontotemporal lobar degeneration by means of FDG-PET. PLoS One 8, e55415.

Edison, P., Hinz, R., Ramlackhansingh, A., Thomas, J., Gelosa, G., Archer, H.A., Turkheimer, F.E., Brooks, D.J., 2012. Can target-to-pons ratio be used as a reliable method for the analysis of [11C]PIB brain scans? Neuroimage 60, 1716–1723.

Erlandsson, K., Buvat, I., Pretorius, P.H., Thomas, B.A., Hutton, B.F., 2012. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. Phys. Med. Biol. 57, R119–R159.

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19, 224–247.

Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691.

Johnson, K.A., Sperling, R.A., Gidicsin, C.M., Carmasin, J.S., Maye, J.E., Coleman, R.E., Reiman, E.M., Sabbagh, M.N., Sadowsky, C.H., Fleisher, A.S., Murali Doraiswamy, P., Carpenter, A.P., Clark, C.M., Joshi, A.D., Lu, M., Grundman, M., Mintun, M.A., Pontecorvo, M.J., Skovronsky, D.M., group, A.A.s, 2013. Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging. Alzheimers Dement. 9, S72–S83.

Joshi, A.D., Pontecorvo, M.J., Clark, C.M., Carpenter, A.P., Jennings, D.L., Sadowsky, C.H., Adler, L.P., Kovnat, K.D., Seibyl, J.P., Arora, A., Saha, K., Burns, J.D., Lowrey, M.J., Mintun, M.A., Skovronsky, D.M., Florbetapir, F.S.I., 2012. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. J. Nucl. Med. 53, 378–384.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., Macfall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436–443.

Landau, S.M., Thomas, B.A., Thurfjell, L., Schmidt, M., Margolin, R., Mintun, M., Pontecorvo, M., Baker, S.L., Jagust, W.J., 2014. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. Eur. J. Nucl. Med. Mol. Imaging 41, 1398–1407.

Lim, Y.Y., Ellis, K.A., Pietrzak, R.H., Ames, D., Darby, D., Harrington, K., Martins, R.N., Masters, C.L., Rowe, C., Savage, G., Szoeke, C., Villemagne, V.L., Maruff, P., Group, A.R., 2012. Stronger effect of amyloid load than APOE genotype on cognitive decline in healthy older adults. Neurology 79, 1645–1652.

Lim, Y.Y., Maruff, P., Pietrzak, R.H., Ellis, K.A., Darby, D., Ames, D., Harrington, K., Martins, R.N., Masters, C.L., Szoeke, C., Savage, G., Villemagne, V.L., Rowe, C.C., Group, A.R., 2014. Abeta and cognitive change: examining the preclinical and prodromal stages of Alzheimer's disease. Alzheimers Dement. 10, 743–751.

Rousset, O.G., Ma, Y., Evans, A.C., 1998. Correction for partial volume effects in PET: principle and validation. J. Nucl. Med. 39, 904–911.

Rousset, O.G., Collins, D.L., Rahmim, A., Wong, D.F., 2008. Design and implementation of an automated partial volume correction in PET: application to dopamine receptor quantification in the normal human striatum. J. Nucl. Med. 49, 1097–1106.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17, 87–97.

Thomas, B.A., Erlandsson, K., Modat, M., Thurfjell, L., Vandenberghe, R., Ourselin, S., Hutton, B.F., 2011. The importance of appropriate partial volume correction for PET quantification in Alzheimer's disease. Eur. J. Nucl. Med. Mol. Imaging 38, 1104–1119.

van Berckel, B.N., Ossenkoppele, R., Tolboom, N., Yaqub, M., Foster-Dingley, J.C., Windhorst, A.D., Scheltens, P., Lammertsma, A.A., Boellaard, R., 2013. Longitudinal amyloid imaging using ^{11}C-PiB: methodologic considerations. J. Nucl. Med. 54, 1570–1576.

Villemagne, V.L., Burnham, S., Bourgeat, P., Brown, B., Ellis, K.A., Salvado, O., Szoeke, C., Macaulay, S.L., Martins, R., Maruff, P., Ames, D., Rowe, C.C., Masters, C.L., 2013. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol. 12, 357–367.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., Morris, J.C., Petersen, R.C., Saykin, A.J., Schmidt, M.E., Shaw, L., Shen, L., Siuciak, J.A., Soares, H., Toga, A.W., Trojanowski, J.Q., 2013. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement. 9, e111–e194.

Wong, K.P., Wardak, M., Shao, W., Dahlbom, M., Kepe, V., Liu, J., Satyamurthy, N., Small, G.W., Barrio, J.R., Huang, S.C., 2010. Quantitative analysis of [^{18}F]FDDNP PET using subcortical white matter as reference region. Eur. J. Nucl. Med. Mol. Imaging 37, 575–588.

Yakushev, I., Landvogt, C., Buchholz, H.G., Fellgiebel, A., Hammers, A., Scheurich, A., Schmidtmann, I., Gerhard, A., Schreckenberger, M., Bartenstein, P., 2008. Choice of reference area in studies of Alzheimer's disease using positron emission tomography with fluorodeoxyglucose-F18. Psychiatry Res. 164, 143–153.

Ziegler-Graham, K., Brookmeyer, R., Johnson, E., Arrighi, H.M., 2008. Worldwide variation in the doubling time of Alzheimer's disease incidence rates. Alzheimers Dement. 4, 316–323.