



Data driven diagnostic classification in Alzheimer's disease based on different reference regions for normalization of PiB-PET images and correlation with CSF concentrations of A β species

Francisco Oliveira^a, Antoine Leuzy^b, João Castelhana^a, Konstantinos Chiotis^b, Steen Gregers Hasselbalch^c, Juha Rinne^d, Alexandre Mendonça^e, Markus Otto^f, Alberto Lleó^g, Isabel Santana^h, Jarkko Johanssonⁱ, Sarah Anderl-Straub^f, Christine Arnim^f, Ambros Beer^j, Rafael Blesa^g, Juan Fortea^g, Herukka Sanna-Kaisa^k, Erik Portelius^{l,m}, Josef Pannee^{l,m}, Henrik Zetterberg^{l,m,n,o}, Kaj Blennow^{l,m}, Ana P. Moreira^a, Antero Abrunhosa^a, Agneta Nordberg^{b,p}, Miguel Castelo-Branco^{a,*}

^a CIBIT, ICGAS, Institute for Biomedical Imaging in Life Sciences (IBILI), Faculty of Medicine, University of Coimbra, Portugal

^b Department of Neurobiology, Care Sciences and Society, Center for Alzheimer Research, Translational Alzheimer Neurobiology, Karolinska Institutet, Stockholm, Sweden

^c Danish Dementia Research Centre, Department of Neurology, Rigshospitalet, University of Copenhagen, Denmark

^d Division of Clinical Neurosciences, Turku University Hospital, University of Turku, Turku, Finland

^e Department of Neurology and Laboratory of Neurosciences, Faculty of Medicine, University of Lisbon, Lisbon, Portugal

^f Department of Neurology, University of Ulm, Germany

^g Neurology Department, Hospital de Sant Pau, Barcelona, Spain and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Barcelona, Spain

^h Centro Hospitalar e Universitário de Coimbra, Neurology Department, Faculty of Medicine, University of Coimbra, Portugal

ⁱ Turku University Hospital, Turku PET Centre, Turku, Finland

^j Department of Nuclear Medicine, Ulm University Hospital, Ulm, Germany

^k Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

^l Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Mölndal, Sweden

^m Clinical Neurochemistry Lab, Sahlgrenska University Hospital, Mölndal, Sweden

ⁿ Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, United Kingdom

^o UK Dementia Research Institute, London, United Kingdom

^p Department of Geriatric Medicine, Karolinska University Hospital Huddinge, Stockholm, Sweden

ABSTRACT

Positron emission tomography (PET) neuroimaging with the Pittsburgh Compound_B (PiB) is widely used to assess amyloid plaque burden. Standard quantification approaches normalize PiB-PET by mean cerebellar gray matter uptake. Previous studies suggested similar pons and white-matter uptake in Alzheimer's disease (AD) and healthy controls (HC), but lack exhaustive comparison of normalization across the three regions, with data-driven diagnostic classification.

We aimed to compare the impact of distinct reference regions in normalization, measured by data-driven statistical analysis, and correlation with cerebrospinal fluid (CSF) amyloid β (A β) species concentrations.

243 individuals with clinical diagnosis of AD, HC, mild cognitive impairment (MCI) and other dementias, from the Biomarkers for Alzheimer's/Parkinson's Disease (BIOMARKAPD) initiative were included. PiB-PET images and CSF concentrations of A β ₃₈, A β ₄₀ and A β ₄₂ were submitted to classification using support vector machines. Voxel-wise group differences and correlations between normalized PiB-PET images and CSF A β concentrations were calculated.

Normalization by cerebellar gray matter and pons yielded identical classification accuracy of AD (accuracy-96%, sensitivity-96%, specificity-95%), and significantly higher than A β concentrations (best accuracy 91%). Normalization by the white-matter showed decreased extent of statistically significant multivoxel patterns and was the only method not outperforming CSF biomarkers, suggesting statistical inferiority. A β ₃₈ and A β ₄₀ correlated negatively with PiB-PET images normalized by the white-matter, corroborating previous observations of correlations with non-AD-specific subcortical changes in white-matter. In general, when using the pons as reference region, higher voxel-wise group differences and stronger correlation with A β ₄₂, the A β ₄₂/A β ₄₀ or A β ₄₂/A β ₃₈ ratios were found compared to normalization based on cerebellar gray matter.

* Corresponding author.

E-mail address: mcbresco@fmed.uc.pt (M. Castelo-Branco).

<https://doi.org/10.1016/j.nicl.2018.08.023>

Received 24 March 2018; Received in revised form 1 July 2018; Accepted 17 August 2018

Available online 19 August 2018

2213-1582/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Positron emission tomography (PET) imaging with the ^{11}C -Pittsburgh Compound B (PiB) tracer is currently used in many nuclear medicine imaging centers to visualize in vivo amyloid plaques in the brain, which represent a core molecular feature of Alzheimer's disease (AD) (Hardy & Selkoe, 2002).

The binary assessment of PiB-PET images, abnormal (amyloid-positive) versus normal (amyloid-negative), can be done by examining tracer uptake in cortical regions of interest. While the most commonly used approach, from a clinical standpoint, is the visual assessment of summated concentration images, quantitative approaches can also be applied; the most common of these is the standardized uptake value ratio (SUVR), which consists in normalizing uptake within target regions to that within a reference region. A global cut-off can then be applied to determine whether the PiB image is positive or negative. Quantitative assessment increases the accuracy and confidence of the visual readings and is also useful for longitudinal studies and clinical trials. The cerebellar gray matter has been widely used as reference region since its amyloid accumulation has been demonstrated to bear no significant differences between healthy controls (HC) and AD patients (Price et al., 2005; Klunk et al., 2004). Other biomarker extensively used in the clinical diagnosis of AD is the cerebrospinal fluid (CSF) concentration of amyloid- β (Blennow et al., 2012; Ros en et al., 2013). It is well known that, in AD patients, the concentration of amyloid- β_{42} ($\text{A}\beta_{42}$) in the CSF is generally decreased (Olsson et al., 2016) concurrently with elevated brain retention of amyloid tracers, such as PiB or ^{18}F -florbetapir (Leuzy et al., 2016; Johnson et al., 2013; Mattsson et al., 2014).

Mild cognitive impairment (MCI) may often represent a prodromal stage of AD, with a conversion rate to dementia due to AD of about 10% to 25% per year while healthy elderly progress at a rate of approximately 1% to 2% per year (Grand et al., 2011). MCI patients who are PiB amyloid-positive are very likely cases of prodromal AD, while patients with MCI who are PiB amyloid-negative are less likely to represent a prodromal stage and to undergo conversion to AD (Wolk et al., 2009; Okello et al., 2009; Jack et al., 2010).

Our main goal was to assess, using multivariate approaches, if the cerebellar gray matter is the best choice, from a clinical point of view, to be used as reference region when compared with the pons or sub-cortical white matter, since these two areas were also found to have similar PiB retentions in AD patients and HC subjects (Klunk et al., 2004). To decide which approach that would be the best option, we here investigated: 1) the ability to discriminate clinically diagnosed patients using voxel-wise statistical analysis, 2) the voxel-wise correlation with the CSF $\text{A}\beta$ concentrations (providing both clinical and biological agreement), and 3) the accuracy in data-driven classification between clinically defined AD patients and HC or patients with other non AD dementias.

A secondary goal was to compare, using data driven classification methods, the classification accuracy of PiB using the SUVR against the accuracy achieved using the CSF concentrations of $\text{A}\beta_{38}$, $\text{A}\beta_{40}$ and $\text{A}\beta_{42}$ and their normalized values as assessed by the $\text{A}\beta_{42}/\text{A}\beta_{38}$ and $\text{A}\beta_{42}/\text{A}\beta_{40}$ ratios determined in the same central laboratory.

2. Methods

2.1. Dataset

The dataset used in this study has been described elsewhere (Leuzy et al., 2016) and is summarized in Table 1. It consists of 243 subjects from seven European academic centers belonging to the Biomarkers for Alzheimer's and Parkinson's Disease (BIOMARKAPD) initiative. It contains five groups of subjects: HC, patients with AD, patients with MCI, patients with frontotemporal dementia (FTD) and patients with vascular dementia (VaD). PiB-PET acquisitions protocols varied across

Table 1

Summary of demographics, clinical and locally measured biomarkers according to the diagnostic group.

	AD (n = 122)	MCI (n = 81)	FTD (n = 20)	VaD (n = 7)	HC (n = 13)
Age, years	65 (59, 72)	64 (58, 71)	64 (59, 73)	61 (52, 74)	67 (58, 71)
Sex, M:F	50:72	37:44	9:11	3:4	6:7
MMSE, points	23 (20, 26)	27 (26, 28)	23 (20, 27)	26 (20, 29)	29 (28, 30)
PiB visual, positive	113	50	3	0	1
Ab42, positive	96	46	8	5	1
CSF-PiB, months	2.4 (0.7, 5.2)	4.0 (1.8, 8.4)	2.0 (1.1, 4.0)	3.5 (2.8, 6.1)	1.8 (1.3, 7.4)

Age, MMSE and CSF-PiB are reported as median (quartile 1, quartile 3), CSF-PiB is the time between the CSF collection and the PiB-PET exam.

sites. In all cases a late summation was considered, being the post injection intervals: 40 to 60 min ($n = 101$), 40 to 70 min ($n = 31$), 50 to 70 min ($n = 24$) and 60 to 90 min ($n = 87$). PiB-PET images were classified locally by a nuclear medicine physician as either positive (abnormal) if there was high binding in cortical regions, or negative (normal) if there was a predominantly white matter binding. All PiB-PET images had a isotropic voxel size of 2 mm. Local $\text{A}\beta_{42}$ values were classified as positive (abnormal) or negative (normal) using an optimal cut-off of 557 pg/ml (Zwan et al., 2016). Local $\text{A}\beta_{42}$ concentrations were measured using commercially available sandwich ELISA (INNO-TEST, Fujirebio-Europe) and with similar protocol. Concerning central harmonization of measures (used in this study), see below.

Patients were assessed according to standard local clinical routines, and all diagnoses were made by a multidisciplinary team using a consensus-based approach. Patients with AD fulfilled the 1984 National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable AD dementia (McKhann et al., 1984), MCI patients were diagnosed according to the Petersen et al. (Petersen et al., 1999) criteria, FTD patients were diagnosed according to the Neary et al. (Neary et al., 1998) criteria, and finally, the VaD patients were diagnosed according to the National Institute of Neurological Disorders and Stroke - Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria for vascular dementia (Rom an et al., 1993).

The HC subjects were recruited from relatives and caregivers of patients. Inclusion criteria were absence of memory or other cognitive complaints; independence in basic and instrumental daily life activities; and no discernible neurological or psychiatric disease.

All participants, or caregivers, when appropriate, gave written informed consent to participate in the research, which was conducted according to the Declaration of Helsinki and subsequent revisions. Ethical approval was obtained from local regional ethics committees.

CSF concentration values used in this study were centrally obtained. $\text{A}\beta_{42}$ concentrations were obtained using the reference measurement procedure (RMP) by liquid chromatography (LC) tandem mass spectrometry (MS) (MS-RMP) while $\text{A}\beta_{38}$ and $\text{A}\beta_{40}$ concentrations were obtained by a fully validated LC-MS method (Leinenbach et al., 2014). $\text{A}\beta_{38}$, $\text{A}\beta_{40}$ and $\text{A}\beta_{42}$ were also analyzed using the MSD V-PLEX $\text{A}\beta$ Peptide Panel 1 (4G8) kit (Meso Scale Diagnostics, Rockland, MD, USA), following the manufacturer's protocol. Samples from the local centers were sent for analysis to Clinical Neurochemistry Laboratory, Gothenburg University, M lndal, Sweden. Technical measurement protocols are described elsewhere (Leuzy et al., 2016).

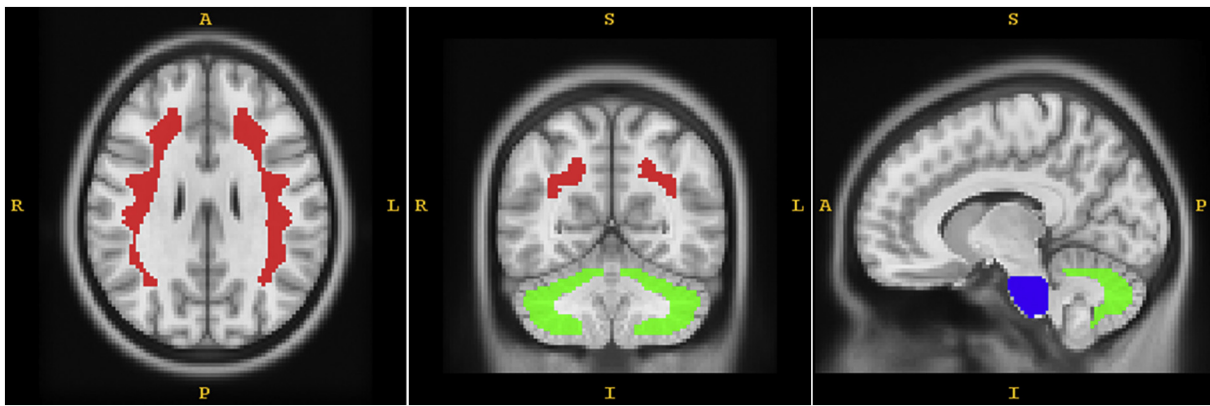


Fig. 1. Illustration of the reference regions used. Subcortical white matter is painted red, cerebellar gray is painted green and pons is painted blue.

2.2. PiB-PET image pre-processing

Before further processing, all images were non-linearly spatially normalized to the Montreal Neurological Institute (MNI) T1 MRI template using Statistical Parametric Mapping 8 (SPM8), as described elsewhere (Leuzy et al., 2016). The spatial normalization was made uniquely based on the PiB-PET images. All spatially normalized images were visually inspected and consequently the registration was fine tuned when necessary.

The SUVR was computed at the voxel level for all images, using three different reference regions: cerebellar gray matter, pons and subcortical white matter; which we defined as $SUVR_{CER}$, $SUVR_{PONS}$ and $SUVR_{WM}$, respectively. All three masks were defined on the T1 MRI template ICBM152 and then shrunk at least 4 mm all around to diminish the influence of the partial volume effects and imperfections of the image registration process. The cerebellar gray matter is essentially the cerebellum without the cerebellar peduncles. Fig. 1 illustrates the masks used as reference region. Note that, since the PiB-PET images were spatially normalized to the MNI space then the pons, cerebellum and white matter are also in the MNI space and the masks defined in the T1 MRI template can be directly applied to the spatially normalized PiB-PET images.

2.3. Voxel-wise assessment of the SUVR differences

Voxel-wise group differences were evaluated using analysis of variance (ANOVA) using Statistical Parametric Mapping 12 (SPM12) following smoothing with a Gaussian kernel with a full width at half maximum (FWHM) of 12 mm. Post hoc pairwise comparisons were made using the Student *t*-test. To address the multiple comparisons issue, significance was only ascribed to regions with voxel-level $p < .001$.

2.4. Voxel-wise correlation between SUVR and CSF A β concentrations

Correlations between CSF A β_{38} , A β_{40} , A β_{42} , A $\beta_{42}/A\beta_{38}$, A $\beta_{42}/A\beta_{40}$ concentrations (measured with MSD and MS-RMP) and voxel-wise $SUVR_{CER}$, $SUVR_{PONS}$ and $SUVR_{WM}$ were computed after smoothing the SUVR images with a Gaussian kernel with a FWHM of 12 mm. Parametric correlation maps (positive and negative) were computed on the full cohort of subjects together. Correction for multiple comparisons was assessed as in the previous section.

2.5. Comparison of the automatic classification accuracies

Regarding the assessment of the ability to differentiate between clinically defined AD (given that a postmortem neuropathological golden standard was not available across sites) and HC or other

dementias OD, five sets of features were extracted from the data: (1) A β_{38} , A β_{40} , A β_{42} , A $\beta_{42}/A\beta_{38}$, A $\beta_{42}/A\beta_{40}$ based on MSD; (2) A β_{38} , A β_{40} , A β_{42} , A $\beta_{42}/A\beta_{38}$, A $\beta_{42}/A\beta_{40}$ based on MS-RMP; (3) voxel-wise $SUVR_{CER}$; (4) voxel-wise $SUVR_{PONS}$; and (5) voxel-wise $SUVR_{WM}$. The goal was to set up an automatic classification approach to decide if a subject data belongs to the AD group or not.

Since the HC, FTD and VaD groups included a small number of individuals comparatively to the AD group and accumulation of amyloid plaques in the brain is not a feature of these three groups of individuals, we opted to join them together in just one dataset referred to as HC/OD. Thus, the binary classification is AD versus HC/OD.

We used support vector machines (SVM) (Chang & Lin, 2011) as a classification technique. This technique can be divided in two steps: in the first (learning/training step) a mathematical model, i.e. a decision function, that best separates the training dataset is built using optimization techniques; in the second step (test) the model built in the first step is used to classify new data. Based on the patient's features, the decision function gives a score to the patient.

The leave one out cross-validation (LOOCV) technique was used to assess the performance of the classifiers. The LOOCV technique is a cross validation method that is used to estimate the performance of a classifier. In this case, it uses the data of a subject to be classified, while the remaining subjects' data are used to train the classifier. This procedure is repeated until all subjects' data have been classified once based on the classifier built with the remaining subjects' data. Then, based on the results obtained on the successive classification tests, the accuracy, sensitivity and specificity are computed.

Since the two groups of subjects are unbalanced, in the optimization process is given more weight to the HC/OD than to the AD cases. The weight of the HC/OD is 122/40 times the weight of the AD. Thus, the optimizer instead to converge to the maximal accuracy tends to converge to the maximal balanced accuracy.

Since there is a high correlation among neighboring voxels, before use the voxel-wise SUVR in the classifier, the SUVR images were re-sampled into 8 mm isotropic voxels. For classification, we considered only the uptake of PiB in the brain cortex, defining an anatomical mask to select only the voxels that belong to this region. The resampled voxels were then used as features (voxel-as-feature approach) (Oliveira & Castelo-Branco, 2015).

Statistical comparison of classifiers accuracy was done using Cochran's Q test followed by the McNemar test as a post-hoc procedure (IBM SPSS Statistics 20).

3. Results

3.1. Voxel-wise differences of the PiB SUVR among groups

Voxel-wise ANOVA showed a statistically significant difference

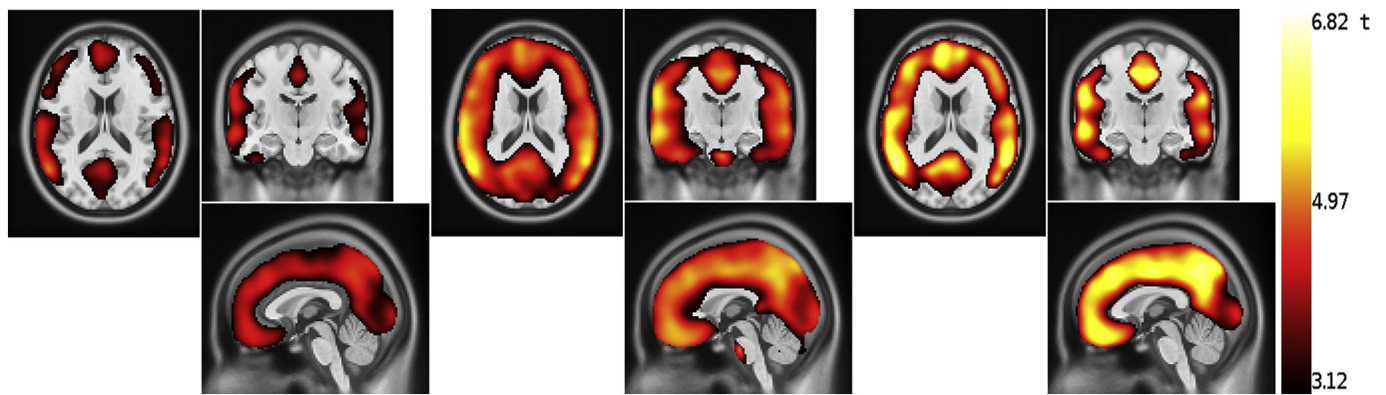


Fig. 2. Regions where the SUVR of the MCI patients is significantly higher than the SUVR of the HC subjects. From the left to the right, voxel-wise t -value obtained using the $SUVR_{CER}$, $SUVR_{PONS}$ and $SUVR_{WM}$. Note that the latter misses large clusters of cortical regions, in particular in occipitoparietal and temporal regions, with a similar pattern for $SUVR_{CER}$. Only the $SUVR_{PONS}$ captures the whole cortical mantle.

(voxel-level $p < .001$) in the PiB uptake in all cortical areas among the defined four groups of subjects using the $SUVR_{CER}$ and $SUVR_{PONS}$ and in most of the cortical mantle using the $SUVR_{WM}$. Some regions, however, were not being detected by this method. Post hoc t -tests showed that, using any of the three reference regions, there was a statistically significant difference between the AD and HC groups, and between AD and OD across almost all cortical voxels. In the comparison AD vs MCI, the differences were slightly higher using $SUVR_{CER}$ and $SUVR_{PONS}$, compared to $SUVR_{WM}$. In the comparison of MCI and HC, statistically significant differences were observed in a larger cluster of areas in the statistical maps using the $SUVR_{PONS}$ than using the $SUVR_{CER}$ and $SUVR_{WM}$, which failed to capture occipitoparietal regions (Fig. 2). Although the brain areas with significant differences are smaller using the $SUVR_{WM}$ than using the $SUVR_{PONS}$, there are clusters with higher t -value using the $SUVR_{WM}$. Also in Fig. 2, a small cluster can be observed in the pons using the $SUVR_{PONS}$ for group comparison. Since it is in the border between regions with very different uptake and the t -values are not very high, this cluster is very likely a false positive, representing a typical border effect.

Images for the other comparisons are available as supplementary figures.

3.2. Voxel-wise correlation between CSF $A\beta$ concentrations and PiB SUVR

We performed a voxel-wise correlation analysis between the SUVR images of all subjects and the CSF $A\beta$ concentrations and their ratios. Table 2 presents a summary of the observed patterns of correlation. In general, the correlations were slightly stronger using the concentrations measured by the MSD than measured by the MS-RMP methods.

When comparing the whole brain correlations as function of the reference region used for normalization of the PiB-PET images, we found a weak positive correlation between the CSF $A\beta$ concentration and the $SUVR_{CER}$ and $SUVR_{WM}$ in the ventricles and/or brainstem but not with the $SUVR_{PONS}$. $A\beta_{42}$, the $A\beta_{42}/A\beta_{38}$ and $A\beta_{42}/A\beta_{40}$ ratios showed a moderate to strong negative (as expected) correlation with $SUVR_{CER}$ and $SUVR_{PONS}$ in all cortical regions, and in most but not all the cortical mantle with the $SUVR_{WM}$, suggesting that the latter is indeed less sensitive. $A\beta_{38}$ and $A\beta_{40}$ correlated significantly and negatively (albeit with a small effect size) with the $SUVR_{WM}$ in part of the parietal lobe, while they did not significantly correlate with the $SUVR_{CER}$ and $SUVR_{PONS}$.

Fig. 3 shows a comparison of the voxel-wise statistically significant correlation between $A\beta_{42}/A\beta_{40}$ measured by the MSD and the SUVR for all three reference regions. Images for the other correlations are available as supplementary figures.

3.3. Classification accuracy

Results from the assessment of the classification accuracies (taking into account the limitation that it is not possible to use a neuropathological gold standard, but just the clinical diagnosis) using the LOOCV are depicted in Table 3 and Fig. 4. Cochran's Q test showed that there was a statistically significant difference ($p < .001$) among the accuracies achieved on the differentiation of clinical AD from HC/OD using the all sets of features. Post hoc tests were made using the McNemar test. P -value results are shown in Table 4. It can be observed that the classification accuracies obtained using the $SUVR_{CER}$ or $SUVR_{PONS}$ are significantly higher than the accuracies obtained using the CSF concentration features. The classification accuracy obtained using the $SUVR_{WM}$ was inferior only at a trend level to the classification accuracies obtained using the $SUVR_{PONS}$ or $SUVR_{CER}$.

$SUVR_{PONS}$ and $SUVR_{CER}$ provided exactly the same accuracies, correctly classifying 155 out of 162 cases (95.7%) (Fig. 4). Regarding the seven misclassified cases, four were clinically diagnosed as AD but all five classifiers indicate they are not AD patients, suggesting that future work should focus on neuropathological validation. In two cases, the patients were diagnosed as FTD but all five classifiers indicated that they are more likely to have AD, again suggesting that gold standard and clinical discrimination issues remain to be solved. Finally, the last case was clinically diagnosed as AD and as AD by the classifiers based on the CSF concentrations but classified as non-AD by the classifiers based on SUVR. This was the only case where the classifiers based on CSF concentrations classified accordingly as the clinical diagnosis while the classifiers based on the SUVR did not, suggesting that the latter is usually more consistent with clinical assessment.

On this dataset of AD and HC/OD, 113 of the 122 AD were visually classified as PiB positive and 36 of the 40 HC/OD were visually classified as PiB negative (Table 1). This represents a sensitivity of 92.6%, a specificity of 90% and accuracy of 92.0%, which is inferior to the accuracy found using the cerebellar gray matter or pons as reference region (one-tailed McNemar test, $p = .035$).

3.4. Comparison of amyloid burden and CSF data in MCI patients

The five classifiers built were applied to the data from the MCI patients with the goal to assess if each patient data is more close to AD than to HC/OD. The rate of MCI patients classified as AD-like was similar for all classifiers and varied between 63% and 65%. The better agreement was between the classifiers based on $SUVR_{CER}$ and $SUVR_{PONS}$ (agreement 80/81, Cohen's Kappa .973), and the worst agreement between the classifier based on CSF concentrations computed with the MS-RMP and the classifier based on $SUVR_{PONS}$ (agreement 75/81, Cohen's Kappa 0.839).

Table 2

Summary of the statistically significant correlation patterns found between the CSF A β concentrations and the PiB SUVR normalized by the three reference regions. NS - not significant correlation or just in small cluster (less than 100 voxels), WC - weak correlation ($0.2 < |r| \leq 0.4$), MD - moderate correlation ($0.4 < |r| \leq 0.7$), SC - strong correlation ($0.7 < |r| \leq 0.9$), (+) - positive correlation and (–) - negative correlation.

		SUVR _{cer}	SUVR _{pons}	SUVR _{wm}
MSD	A β ₃₈	(+)WC: ventricles and brainstem	(+)WC: ventricles	(+)WC: ventricles
		(–)NS	(–)NS	(–)WC: parietal lobe
	A β ₄₀	(+)WC: ventricles and brainstem	(+)WC: part of ventricles	(+)WC: part of ventricles
		(–)NS	(–)NS	(–)WC: part of parietal lobe
	A β ₄₂	(+)WC-MC: brainstem	(+)NS	(+)WC-MC: brainstem
		(–)WC: all brain cortex	(–)MC: all brain cortex	(–)WC-MC: all brain cortex
A β ₄₂ /A β ₃₈	(+)WC: brainstem	NS	(+)MC: brainstem	
A β ₄₂ /A β ₄₀	(–)MC: all brain cortex	(–)MC-SC: all brain cortex	(–)MC: all brain cortex	
	(+)WC: brainstem	(+)NS	(+)MC: brainstem	
	(–)MC: all brain cortex	(–)MC-SC: all brain cortex	(–)MC-SC: all brain cortex	
	(+)WC: ventricles and brainstem	(+)WC: ventricles	(+)WC: part of ventricles	
MS-RMP	A β ₃₈	(–)NS	(–)NS	(–)WC: part of parietal lobe
		(+)WC: ventricles and brainstem	(+)NS	(+)NS
	A β ₄₀	(–)NS	(–)NS	(–)WC: part of parietal lobe
		(+)WC: brainstem	(+)NS	(+)WC: brainstem
	A β ₄₂	(–)WC: all brain cortex	(–)MC: all brain cortex	(–)WC-MC: all brain cortex
		(+)WC: brainstem	(+)NS	(+)MC: brainstem
A β ₄₂ /A β ₃₈	(–)MC: all brain cortex	(–)MC: all brain cortex	(–)WC-MC: all brain cortex	
	(+)WC: brainstem	(+)NS	(+)MC: brainstem	
A β ₄₂ /A β ₄₀	(–)MC: all brain cortex	(–)MC-SC: all brain cortex	(–)WC-MC: all brain cortex	
	(+)WC: brainstem	(+)NS	(+)MC: brainstem	

Now, comparing the classification made by the classifier based on SUVR_{PONS} with the classifications based on the PiB visual assessment and locally measured A β ₄₂ using the optimal cut-off (Zwan et al., 2016), an agreement of 75/81 and 61/81 was obtained, respectively. Similar results were obtained comparing with the SUVR_{CER} based classifier.

4. Discussion

In this data-driven multivariate study we investigated the impact of PiB SUVR normalization (cerebellar gray matter, pons or white matter) on overall statistical classification of clinical diagnostic categories, and a comparison with CSF A β measures. To test the relative value of these options to differentiate patient groups we used an automatic data classification framework based on a dataset acquired in multiple European Centers. Importantly, we also tested which of the A β ₃₈, A β ₄₀ and A β ₄₂ individual values or ratios correlated best with PiB-PET SUVR images.

The results showed that the classification accuracy of clinically defined AD versus HC or OD based on the SUVR_{CER} and SUVR_{PONS} images are equal and significantly higher than the accuracies obtained using the CSF concentrations. Thus, this allows us to conclude that the PiB-PET SUVR seems to be a promising solution to be used in multivariate classification when compared the CSF concentrations of multiple A β species, although this needs future confirmation with the

Table 3

Cross-validation classification results from the differentiation between clinically defined AD and HC/OD using the SVM classifiers. Values of accuracy, sensitivities, specificities and balanced accuracy are given in percentage. Please note that all CSF measures were taken into account as classification features.

	Accuracy	Sensitivity	Specificity	Balanced accuracy
CSF measured with MS-RMP	88.3	91.8	77.5	84.7
CSF measured with MSD	90.7	93.4	82.5	88.0
SUVR _{WM}	93.8	95.1	90.0	92.5
SUVR _{CER}	95.7	95.9	95.0	95.5
SUVR _{PONS}	95.7	95.9	95.0	95.5

neuropathological gold standard. It is however possible that adding Tau levels might increase CSF performance. In fact, only in one case the classifiers based on CSF concentrations classified accordingly to the clinical diagnosis while the classifiers based on SUVR_{CER} or SUVR_{PONS} did not. This means that in clinical practice the use of the CSF concentrations needs reappraisal when compared to the classification using the SUVR_{CER} or SUVR_{PONS} alone. Our finding does not mean the CSF concentrations should not be measured, we only conclude that to perform just the differential diagnosis of AD, the CSF concentrations may

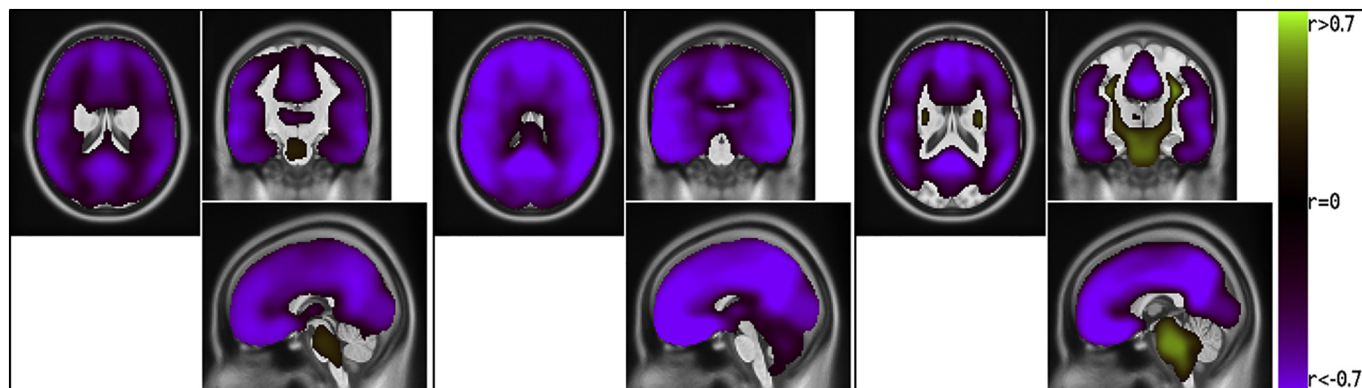


Fig. 3. Voxel-wise statistically significant correlation between MSD A β ₄₂/A β ₄₀ and SUVR_{CER}, SUVR_{PONS} and SUVR_{WM}, respectively. Correlation was computed for the entire dataset. Note that parts of the SUVR_{WM} maps lack a correlation pattern.

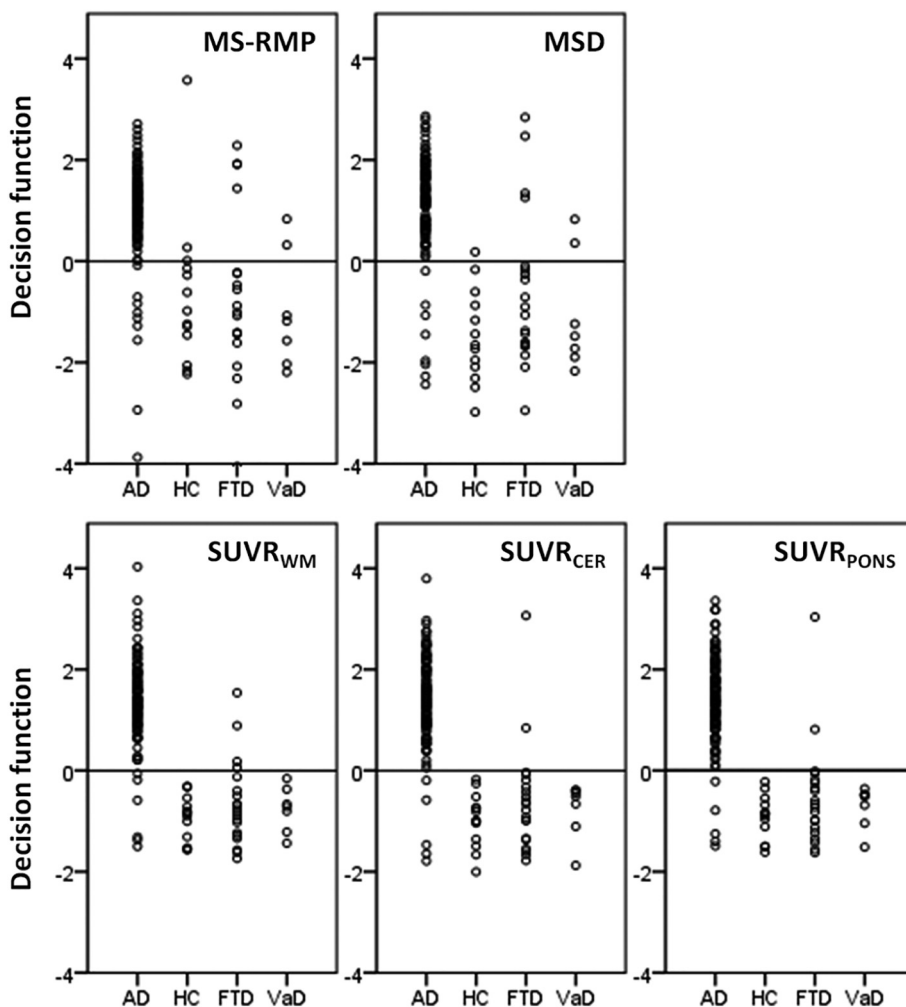


Fig. 4. Values of the decision functions obtained from the SVM classifiers. Values were obtained during the accuracy assessment using the LOOCV strategy. In all these cases, a positive value means that the case is more compatible with the AD patients than the other conditions. A negative value means the opposite.

Table 4

P-values for the post hoc pairwise accuracies comparison using the McNemar test. Please note that all CSF measures were taken into account as classification features.

	CSF measured with MSD	SUVR _{WM}	SUVR _{CER}	SUVR _{PONS}
CSF measured with MS-RMP	0.289	0.035	0.002	0.002
CSF measured with MSD		0.227	0.021	0.021
SUVR _{WM}			0.250	0.250
SUVR _{CER}				1

be possibly redundant if a PiB-PET acquisition is available; however the CSF concentrations contain complementary biological information, for instance Tau biomarkers, that may be relevant to the physician (Rosén et al., 2013). This tenet will however remain controversial without neuropathological validation.

Although the qualitative PiB-PET visual evaluation is often used to help the physicians in the diagnosis of the patients, the accuracies found using the cerebellar gray matter or pons as reference regions are higher than the accuracy found using the PiB-PET visual evaluation.

It is important to stress out that the PiB-PET images were acquired using different platforms and scanning windows, which we view as a strength, given the positive results identified in this study. On the other hand, while the CSF Aβ concentrations were measured centrally with

the same assay procedures, the samples were collected at seven different clinical centers, which may have introduced variability due to differences in pre-analytical protocols (Bjerke, et al., 2010). This shows the robustness of both PiB-PET imaging and CSF biomarkers also in the multicenter setting.

Our finding partially contradicts the results of Mattsson et al. (Mattsson et al., 2014), where the authors found that CSF Aβ₄₂ and florbetapir-PET did not differ in terms of area under the curve (AUC) in the classification of the AD versus HC. In this study we have used more than one thousand of features (the resampled SUVR voxels) to represent the PiB-PET image, which contains more information than a single value (global or regional PiB), as used in Mattsson et al. (Mattsson et al., 2014). Moreover, we have used a set of CSF biomarkers as features, which allows increasing the classification accuracy comparatively if just one Aβ feature was used at a time. Previous studies (Leuzy et al., 2016; Janelidze et al., 2016) have shown the ratios Aβ₄₂/Aβ₄₀ and Aβ₄₂/Aβ₃₈ originate higher classification accuracy than using only the Aβ₄₂. When we compared the classification results obtained using the CSF biomarkers from the MSD and MS-RMP methods, we found no significant difference.

When the classifiers were applied to the MCI patients, a good agreement among all classifiers was found. In the worst case (CSF MS-RMP based classifier versus SUVR_{PONS} based classifier) there was a disagreement in 6 out of 81 patients. Depending on the classifier, 63% to 65% of the MCI patients were classified as AD-like, which may lead to different diagnosis/prognostic for these patients in comparison with

the other who are classified as non AD-like. Which is the best classifier to predict the conversion from MCI to AD is a question that only a subsequent follow-up study can answer. It is important to stress out that the classification based on the $SUVR_{PONS}$ and locally measured $A\beta_{42}$ disagree in 20 out of 81 cases, which is a very substantial difference.

The $A\beta_{42}/A\beta_{38}$ and $A\beta_{42}/A\beta_{40}$ ratios gave higher (negative) voxel-wise correlation with PiB-PET SUVR than the $A\beta_{42}$ concentration alone. These higher correlations with PiB-PET SUVR may explain why the $A\beta_{42}$ ratios provided better classification results than the $A\beta_{42}$ concentration (Leuzy et al., 2016; Janelidze et al., 2016). Our findings suggest that the voxel-wise correlation of the SUVR with the $A\beta_{42}/A\beta_{38}$ and $A\beta_{42}/A\beta_{40}$ ratios is slightly stronger if the pons is used to normalize the uptake than if one uses the cerebellar gray matter or the brain white matter (Fig. 3). This provides a strong biological argument in favor of this reference region.

We found that $A\beta_{38}$ and $A\beta_{40}$ correlate weakly and negatively with the $SUVR_{WM}$ in part of the parietal lobe, while they do not significantly correlate with the $SUVR_{CER}$ and $SUVR_{PONS}$. These findings may be of particular biological significance in terms of specificity. Future studies should examine how they relate with the observations of Janelidze et al. (Janelidze et al., 2016) who found that $A\beta_{38}$, $A\beta_{40}$ (as well as $A\beta_{42}$) correlate with non-AD-specific subcortical changes such as larger lateral ventricles and white matter lesions.

We also found that, in general, the voxel-wise SUVR differences between groups of patients are higher (greater F value and larger areas) using the cerebellar gray matter and pons as reference region than using the white matter. This suggests that the latter has less power in detecting the cortical extent of early damage. Also, the difference between MCI and HC is higher (larger areas) using $SUVR_{PONS}$ than using the other two SUVR.

The results we obtained are consistent with the ones obtained using other amyloid ligands. For instance, using ^{18}F -Florbetapir, Habert et al. (Habert et al., 2017) found that when they used an association of the whole cerebellum and pons as reference region they obtained the best discrimination between HC and AD. Unfortunately they did not compare the pons against cerebellum, which precludes direct comparisons. Using the amyloid ligand ^{18}F -flutemetamol, Thurfjell et al. (Thurfjell et al., 2014) found the best discrimination accuracy using the pons as reference region, comparatively to the whole cerebellum or only the cerebellar gray matter, on a dataset of autopsy confirmed AD. This slight superiority of the pons against the cerebellum or cerebellar gray matter are also in agreement with the results of Klunk et al. (Klunk et al., 2004) where the authors found that the relative difference of the PiB uptake between AD and HC is smaller in the pons than in the cerebellum, which means the PiB uptake is more stable in the pons than in the cerebellum.

The main limitation of this study is the lack of an anatomical brain image per patient. Thus, the image registration process, i.e. normalization to the MNI space, was done based on the PiB-PET image only. Consequently, the accuracy of the registration process is inferior to what could be achieved if a structural image like MRI was available. For this reason, we reduced the size of the masks used to ensure as much as possible that, for each patient, each mask contains only voxels of the target brain area. Other consequence was our option to exclude the striatal region from the mask used to extract the SUVR values used in the automated classification process. Note that in elderly patients where a dilatation of the ventricle is common, if the striatal region was included in the mask used for classification it may happen that in some patients we would collect the values of the SUVR from the ventricles rather than from the striatal region.

We have used a linear SVM as classifier model due to its simplicity, wide acceptance and proved good ability for many common classification problems using multivariate medical data (Oliveira & Castelo-Branco, 2015; Oliveira et al., 2018; Duarte et al., 2014; Moradi et al., 2015).

As final remarks, both PiB $SUVR_{PONS}$ and $SUVR_{CER}$ are well suitable

to be used in the differential diagnosis of AD, even if further studies also with postmortem neuropathological gold standard will be important for final validation of diagnostic accuracy. Although $SUVR_{PONS}$ and $SUVR_{CER}$ led to similar classifications accuracies, the $SUVR_{PONS}$ generally showed a higher t -value and larger extent of voxel-wise differences between patient groups. This suggests that the normalization of the PiB-PET uptake images by the pons may be a better option than the normalization by the cerebellar gray matter, as corroborated by studies using other ligands.

Acknowledgements

This study was supported by the JPND networks BiomarkAPD (01ED1203F) and PreFrontAls (01ED1512), the German Federal Ministry of Education and Research (FTLDC O1GI1007A), FAIR-PARK I1633190, the Foundation of the State Baden-Württemberg (D.3830), Boehringer Ingelheim Ulm University BioCenter (D.5009), Thierry Latran Foundation. This study was part of BIOMARKAPD, EU Joint Programme–Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of JPND (www.jpnd.eu): Stockholm (A.L., K.C. and A.N.), the Swedish Research Council (projects 529-2012-14 and 05817), the Karolinska Institutet Strategic Neuroscience program, the Stockholm Country Council-Karolinska Institutet regional agreement on medical training and clinical research (ALF grant), Swedish Brain Power, the Swedish Brain Foundation, Swedish Alzheimer Foundation, Gun and Bertil Stohnes Foundation, Demensfonden, the Alzheimer Foundation in Sweden, the Foundation for Old Servants, the Swedish Foundation for Strategic Research (SSF), and the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n_o HEALTH-F2-2011-278850 (INMiND); Gothenburg (E.P., J.P., H.Z., and K.B.) the Swedish Research Council (project 529-2012-14), the Gamla Tjänarinnor foundation, ERC (681712), the Wolfson Foundation, the Knut and Alice Wallenberg Foundation, Frimurarestiftelsen and the Alzheimer Foundation, the Torsten Söderberg foundation, Hjärnfonden and the Swedish Alzheimer Foundation; Barcelona (A.Leo', R.B. and J.F.), Instituto de Salud Carlos III (PI11/03035-BIOMARKAPD, PI11/02425 and PI14/01126, PI13/01532, PI14/01561), jointly funded by Fondo Europeo de Desarrollo Regional (FEDER), European Union, 'Una manera de hacer Europa' and 'Marato' TV3' grant 0142610; Turku (JOR and SKH), Academy of Finland (decision no. 263193), Sigrid Juselius Foundation, Turku University Hospital Clinical Grants; Ulm (M.O., S.A.S., C.A.F.V.A., and A.B.), BMBF (Ministry of Science and Technology): Competence net neurodegenerative dementias (project: FTLDC), the JPND networks for standardisation of biomarkers (SOPHIA) and the JPND project, PreFrontals, the foundation of the state of Baden-Wuerttemberg and The Thierry Latran Foundation and BIU (Boehringer IngelheimUlm University BioCentre). Contract grant sponsor: "Projecto Operacional Regional do Centro"–BIGDATIMAGE; CENTRO-01-0145-FEDER-000016 and MEDPersystPOCI-01-0145-FEDER-016428; Contract grant sponsor: FCT; Contract grant number: UID/NEU/04539/2013; Contract grant sponsor: COMPETE; Contract grant number: POCI-01-0145-FEDER-007440.

Conflicts of interest

K.B. reports personal fees from IBL International, Roche Diagnostics and Eli Lilly. C.A.F.V.A. reports personal fees from Desitin Arzneimittel GmbH, Dr. Willmar Schwabe GmbH & Co, personal fees and non-financial support from Nutricia GmbH, Lilly Deutschland GmbH, and grants from Roche Diagnostics GmbH, Biologische Heilmittel Heel GmbH, and ViaMed GmbH. J.O.R. serves as a consultant neurologist for Clinical Research Services Turku (CRST) Ltd. A.N. has received grants from GE Healthcare and Bayer Healthcare, served on the scientific advisory boards of GE Healthcare, Avid, and Eli Lilly, and received speaker honorarium from GE Healthcare, Piramal, Novartis, and Bayer

Healthcare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2018.08.023>.

References

- Blennow, K., Zetterberg, H., Fagan, A.M., 2012. Fluid biomarkers in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 2, a006221. <https://doi.org/10.1101/cshperspect.a006221>.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (27), 1. <https://doi.org/10.1145/1961189.1961199>.
- Duarte, J.V., Ribeiro, M.J., Violante, I.R., Cunha, G., Silva, E., Castelo-Branco, M., 2014. Multivariate pattern analysis reveals subtle brain anomalies relevant to the cognitive phenotype in neurofibromatosis type 1. *Hum. Brain Mapp.* 35, 89–106. <https://doi.org/10.1002/hbm.22161>.
- Grand, J.H.G., Caspar, S., MacDonald, S.W.S., 2011. Clinical features and multi-disciplinary approaches to dementia care. *J. Multidiscip. Healthc.* 4, 125–147. <https://doi.org/10.2147/JMDH.S17773>.
- Habert, M.-O., Bertin, H., Labit, M., Diallo, M., Marie, S., Martineau, K., et al., 2017. Evaluation of amyloid status in a cohort of elderly individuals with memory complaints: validation of the method of quantification and determination of positivity thresholds. *Ann. Nucl. Med.* <https://doi.org/10.1007/s12149-017-1221-0>.
- Hardy, J., Selkoe, D.J., 2002. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 297, 353–356. <https://doi.org/10.1126/science.1072994>.
- Jack, C.R., Wiste, H.J., Vemuri, P., Weigand, S.D., Senjem, M.L., Zeng, G., et al., 2010. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain* 133, 3336–3348. <https://doi.org/10.1093/brain/awq277>.
- Janelidze, S., Zetterberg, H., Mattsson, N., Palmqvist, S., Vanderstichele, H., Lindberg, O., et al., 2016. CSF A β 42/A β 40 and A β 42/A β 38 ratios: better diagnostic markers of Alzheimer disease. *Ann. Clin. Transl. Neurol.* 3, 154–165. <https://doi.org/10.1002/acn3.274>.
- Johnson, K.A., Sperling, R.A., Gidycz, C.M., Carmasin, J.S., Maye, J.E., Coleman, R.E., et al., 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. <https://doi.org/10.1016/j.jalz.2012.10.007>.
- Klunk, W.E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D.P., et al., 2004. Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound-B. *Ann. Neurol.* 55, 306–319. <https://doi.org/10.1002/ana.20009>.
- Leinenbach, A., Pannee, J., Dülffer, T., Huber, A., Bittner, T., Andreasson, U., et al., 2014. Mass spectrometry-based candidate reference measurement procedure for quantification of amyloid- β in cerebrospinal fluid. *Clin. Chem.* 60, 987–994. <https://doi.org/10.1373/clinchem.2013.220392>.
- Leuzy, A., Chiotis, K., Hasselbalch, S.G., Rinne, J.O., Mendonça, A., Otto, M., et al., 2016. Pittsburgh compound B imaging and cerebrospinal fluid amyloid- β in a multicentre European memory clinic study. *Brain*. <https://doi.org/10.1093/brain/aww160>.
- Mattsson, N., Insel, P.S., Landau, S., Jagust, W., Donohue, M., Shaw, L.M., et al., 2014. Diagnostic accuracy of CSF A β 42 and florbetapir PET for Alzheimer's disease. *Ann. Clin. Transl. Neurol.* 1, 534–543. <https://doi.org/10.1002/acn3.81>.
- 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, 939–944.
- Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Initiative, AsDN, 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 104, 398–412. <https://doi.org/10.1016/j.neuroimage.2014.10.002>.
- Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., et al., 1998. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 51, 1546–1554.
- Okello, A., Koivunen, J., Edison, P., Archer, H.A., Turkheimer, F.E., Nägren, K., et al., 2009. Conversion of amyloid positive and negative MCI to AD over 3 years: an ¹¹C-PIB PET study. *Neurology* 73, 754–760. <https://doi.org/10.1212/WNL.0b013e3181b23564>.
- Oliveira, F.P.M., Castelo-Branco, M., 2015. Computer-aided diagnosis of Parkinson's disease based on [¹²³I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines. *J. Neural Eng.* 12, 026008. <https://doi.org/10.1088/1741-2560/12/2/026008>.
- Oliveira, F.P.M., Faria, D.B., Costa, D.C., Castelo-Branco, M., Tavares, J.M.R.S., 2018. Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [¹²³I]FP-CIT SPECT images. *Eur. J. Nucl. Med. Mol. Imaging* 45, 1052–1062. <https://doi.org/10.1007/s00259-017-3918-7>.
- Olsson, B., Lautner, R., Andreasson, U., Öhrfelt, A., Portelius, E., Bjerke, M., et al., 2016. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol.* 15, 673–784. [https://doi.org/10.1016/S1474-4422\(16\)00070-3](https://doi.org/10.1016/S1474-4422(16)00070-3).
- 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. <https://doi.org/10.1001/archneur.56.3.303>.
- Price, J.C., Klunk, W.E., Lopresti, B.J., Lu, X., Hoge, J.A., Ziolkowski, S.K., et al., 2005. Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh Compound-B. *J. Cereb. Blood Flow Metab.* 25, 1528–1547. <https://doi.org/10.1038/sj.jcbfm.9600146>.
- Román, G.C., Tatemichi, T.K., Erkinjuntti, T., Cummings, J.L., Masdeu, J.C., Garcia, J.H., et al., 1993. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology* 43, 250–260.
- Rosén, C., Hansson, O., Blennow, K., Zetterberg, H., 2013. Fluid biomarkers in Alzheimer's disease – current concepts. *Mol. Neurodegener.* 8. <https://doi.org/10.1186/1750-1326-8-20>.
- Thurfjell, L., Lilja, J., Lundqvist, R., Buckley, C., Smith, A., Vandenberghe, R., et al., 2014. Automated quantification of ¹⁸F-flutemetamol PET activity for categorizing scans as negative or positive for brain amyloid: concordance with visual image reads. *J. Nucl. Med.* 55, 1623–1628. <https://doi.org/10.2967/jnumed.114.142109>.
- Wolk, D.A., Price, J.C., Saxton, J.A., Snitz, B.E., James, J.A., Lopez, O.L., et al., 2009. Amyloid imaging in mild cognitive impairment subtypes. *Ann. Neurol.* 65, 557–568. <https://doi.org/10.1002/ana.21598>.
- Zwan, M.D., Rinne, J.O., Hasselbalch, S.G., Nordberg, A., Lleó, A., Herukka, S.-K., et al., 2016. Use of amyloid-PET to determine cutpoints for CSF markers: a multicenter study. *Neurology* 86, 50–58. <https://doi.org/10.1212/WNL.0000000000002081>.