Journal of Alzheimer's Disease xx (20xx) x–xx
DOI 10.3233/JAD-150814
IOS Press

Optimization of Statistical Single Subject Analysis of Brain FDG PET for the Prognosis of Mild Cognitive Impairment-to-Alzheimer's Disease Conversion

Catharina Lange[a], Per Suppa[a,b], Lars Frings[c], Winfried Brenner[a], Lothar Spies[b], Ralph Buchert[a,*] and for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*Department of Nuclear Medicine, Charité - Universitätsmedizin Berlin, Berlin, Germany*
[b]*Jung diagnostics GmbH, Hamburg, Germany*
[c]*Department of Nuclear Medicine, University of Freiburg, Freiburg, Germany*

Accepted 20 September 2015

Abstract.
Background: Positron emission tomography (PET) with the glucose analog F-18-fluorodeoxyglucose (FDG) is widely used in the diagnosis of neurodegenerative diseases. Guidelines recommend voxel-based statistical testing to support visual evaluation of the PET images. However, the performance of voxel-based testing strongly depends on each single preprocessing step involved.
Objective: To optimize the processing pipeline of voxel-based testing for the prognosis of dementia in subjects with amnestic mild cognitive impairment (MCI).
Methods: The study included 108 ADNI MCI subjects grouped as 'stable MCI' ($n=77$) or 'MCI-to-AD converter' according to their diagnostic trajectory over 3 years. Thirty-two ADNI normals served as controls. Voxel-based testing was performed with the statistical parametric mapping software (SPM8) starting with default settings. The following modifications were added step-by-step: (i) motion correction, (ii) custom-made FDG template, (iii) different reference regions for intensity scaling, and (iv) smoothing was varied between 8 and 18 mm. The t-sum score for hypometabolism within a predefined AD mask was compared between the different settings using receiver operating characteristic (ROC) analysis with respect to differentiation between 'stable MCI' and 'MCI-to-AD converter'. The area (AUC) under the ROC curve was used as performance measure.
Results: The default setting provided an AUC of 0.728. The modifications of the processing pipeline improved the AUC up to 0.832 ($p=0.046$). Improvement of the AUC was confirmed in an independent validation sample of 241 ADNI MCI subjects ($p=0.048$).
Conclusion: The prognostic value of voxel-based single subject analysis of brain FDG PET in MCI subjects can be improved considerably by optimizing the processing pipeline.

Keywords: Alzheimer's Disease Neuroimaging Initiative, F-18-fluorodeoxyglucose, intensity scaling, mild cognitive impairment, positron emission tomography, processing pipeline, prognosis, single subject analysis, statistical parametric mapping, template

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://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

*Correspondence to: Ralph Buchert, Department of Nuclear Medicine, Charité - Universitätsmedizin Berlin, Berlin, 10117, Germany. Tel.: +49 30 450627059; Fax: +49 30 450527912; E-mail: ralph.buchert@charite.de.

INTRODUCTION

Positron emission tomography (PET) with the glucose analog 2-[F-18]-fluoro-2-deoxy-D-glucose (FDG) is a well-established radionuclide imaging modality for non-invasive *in-vivo* assessment of synaptic function and dysfunction in the brain [1]. Patients with Alzheimer's disease (AD) show a characteristic pattern of cerebral hypoactivity including the posterior cingulate/precuneus area and parietotemporal association cortices not only in the dementia phase but already in the phase of mild cognitive impairment (MCI) [2–7]. Therefore, FDG PET is widely used for early diagnosis of AD and differentiation from neurodegenerative diseases with different characteristic FDG PET pattern [6, 8–12].

Revised criteria for the diagnosis of AD recommend biomarkers including brain FDG PET to complement clinical, i.e., symptom-based criteria with objective evidence of the underlying pathology [13–15], at least in research settings, although it has also been noted that synaptic dysfunction of the brain most likely is a down-stream consequence of amyloid-β pathology and, therefore, might be better considered a biomarker for staging and/or disease monitoring rather than a diagnostic marker [16]. Whereas the future role of FDG PET in the management of patients with suspected AD might not be clear yet, currently it is still widely used in clinically unclear cognitive impairment (CUCI) in everyday routine.

Interpretation of brain FDG PET is based on visual inspection of the reconstructed tomographic images. However, the quality of the interpretation can be improved by software support. Voxel-based statistical single subject analysis [17, 18], i.e., voxel-by-voxel statistical testing of the patient's FDG PET image against a database of normal brain FDG PETs, has been found particularly useful: it not only allows inexperienced readers to detect the AD pattern in FDG PET with the same accuracy (both sensitivity and specificity) as experts, but also results in small improvement of expert interpretation [19]. Thus, common practice guidelines for brain FDG PET recommend the use of voxel-based single subject analysis to support visual interpretation of brain FDG PET in patients with suspected AD [20, 21].

However, whereas there is general consensus *that* voxel-based single subject analysis should be used, there is much less consensus about *how* the analysis should be performed. This is a major limitation, because voxel-based testing requires several preprocessing steps, each of which can have strong impact on overall performance. The lack of standardization of voxel-based single subject analysis might result in the use of suboptimal protocols at some institutions so that the diagnostic and prognostic potential of brain FDG PET most likely is not fully exploited. The aim of the present study therefore was to optimize the processing pipeline of voxel-based single subject analysis for prediction of MCI-to-AD conversion within the framework of the freely available statistical parametric mapping software package (version SPM8) [22].

MATERIALS AND METHODS

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

MCI patients

Subjects with a baseline diagnosis of MCI, a follow-up time of at least 36 months and baseline FDG PET were downloaded from the ADNI database in March 2014. Subjects were categorized according to their diagnostic trajectory over 36 months: all subjects who did not decline, i.e., who remained MCI or changed between MCI and normal cognition, were included in the stable MCI group, whereas subjects whose diagnosis changed to AD (and then stayed AD) during the 3-year follow-up were regarded as MCI-to-AD converters. Conversion to non-AD dementia was an exclusion criterion. There were no further exclusion criteria, particularly no MCI patient was excluded based on limited quality of the PET image. Following this procedure, a total of 108 patients were included: 77 with stable MCI and 31 who had converted to AD dementia (ADD). FDG PET had been performed with 18 different scanners at 44 different ADNI centers. Subject demographics are given in Table 1. The ADNI participant roster ID (RID) of the included patients is given in the Supplementary Material.

Cognitively normal subjects and ADD patients

Thirty-two ADNI-normals (NC) and 32 ADNI-ADD patients with baseline FDG PET were included

Table 1
Baseline subject characteristics according to group. (NC, normal controls; MCI, mild cognitive impairment; AD, Alzheimer's disease; MMSE, Mini-Mental State Examination; FAQ, functional activities questionnaire; ABETA142, concentration of amyloid-β 1-42 peptide in cerebrospinal fluid; t-sum score for the following setting: motion correction, custom FDG template, parenchyma scaling, 12 mm smoothing)

Group	*n*	age* (y)	gender†	education* (y)	FAQ*	MMSE*	ABETA142*‡ (pg/ml)	t-sum score*
NC	32	73.8 ± 4.6	22/10	16.8 ± 2.7	0.56 ± 1.24	28.9 ± 1.2	n.a.	0 ± 8212
MCI stable	77	74.5 ± 7.7	23/54	16.0 ± 2.7	1.68 ± 2.26	27.7 ± 1.6	166.7 ± 63.6	14400 ± 17483
MCI converter	31	74.7 ± 6.4	12/19	15.8 ± 3.0	5.68 ± 5.10	27.1 ± 1.4	145.1 ± 42.8	37817 ± 20182
AD	32	74.0 ± 4.7	22/10	15.2 ± 2.8	13.34 ± 5.22	23.4 ± 2.2	142.6 ± 26.0	49020 ± 21897

*mean ± SD. †female/male. ‡$A\beta_{1\text{-}42}$ available in none of the NC subjects, 30 MCI stables, 17 MCI converters, and 5 AD subjects (ADNI table "UPENNBIOMK.csv").

as normal database for single subject analysis and for generation of an AD typical mask. The NC group was generated from all ADNI normals who (i) had baseline FDG PET, which (ii) had been acquired with a Philips Gemini TF PET/CT system (5 different centers), and (iii) had baseline MRI ($n = 38$). Four of these NC subjects were excluded because of abnormally enlarged inner cerebrospinal fluid space [RID: 4093, 5124, 5197, 5234]. Two further NC subjects were excluded because of at least one significant cluster of hypometabolism ($p \leq 0.001$) in leave-one-out voxel-based single subject analysis (default setting). The remaining 32 NC subjects are described in Table 1.

The ADD patients were selected to match the NC group by age and gender on a subject-by-subject base. In the included ADD patients, FDG PET had been acquired with 16 different scanners at 27 different centers. No attempt was made to restrict the ADD group to patients which also had been scanned with a Philips Gemini TF, since (i) this would have resulted in a considerably smaller sample of only 7 ADD patients and (ii) matching with respect to age and gender appeared more important to us.

FDG PET data

In 152 out of the total of 172 subjects, FDG PET had been acquired according to a dynamic protocol so that 6 frames of 5 min duration from 30 to 60 min post injection were available for analysis. The remaining 20 FDG PETs had been acquired as 30 min static emission scan starting 30 min post injection. Reconstructed dynamic (or static, if dynamic not available) PET data was downloaded in its original image format ("as archived", DICOM, Interfile, or ECAT) in order to guarantee that no preprocessing had been performed. Then, the original images were converted to Nifti, from DICOM and ECAT using SPM8, from Interfile using ImageConverter (version 1.1.5, download: http://www.turkupetcentre.net/programs/tpc_csharp.html).

Voxel-based single subject analysis

All image processing was performed using a custom-made pipeline for fully automated processing implemented in MATLAB and using routines (dicom import, ecat import, image calculator, smooth, realign, coregister, normalize, basic models, unified segmentation) of the freely available Statistical Parametric Mapping software package SPM (version SPM8, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, London, UK) [22, 23].

Several repeats of voxel-based single subject analysis were performed starting with a 'default' setting, which then was adapted by stepwise adding the following changes (as described below): (i) frame-by-frame motion correction of the dynamic PET sequences prior to summing to one static uptake image, (ii) custom-made tracer-specific FDG template generated from the NCs for stereotactical normalization, and (iii) different reference regions for scaling of voxel intensities. Finally, smoothing prior to voxel-based testing was varied. A summary of all settings is shown in Table 2.

The processing pipeline provides a batch mode utility so that all subjects from all groups, i.e., $n = 172$, were processed automatically in one batch for each setting of the single subject analysis.

Frame-by-frame motion correction

In dynamic FDG PETs, inter-frame motion was corrected using the 'realign' routine of SPM8. The first frame was used as reference. The magnitude of the motion was estimated as follows. Five reference points, which had been predefined in template space (located in precuneus, left/right parietotemporal and left/right lateral temporal cortex), were transferred to the first frame of the patient's dynamic scan by stereotactically

Table 2
Settings for single subject analysis

Setting	motion correction	template	intensity scaling	smoothing [mm]	comment
0	no	O-15-water	global scaling	12	SPM8 default
1	yes	O-15-water	global scaling	12	motion correction
2	yes	FDG	global scaling	12	custom FDG template
3a	yes	FDG	parenchyma	12	mean (or median)
3b	yes	FDG	iterative parenchyma	12	exclusion of hypo-voxels
3c	yes	FDG	Yakushev*	12	inclusion of hyper-voxels
3d	yes	FDG	pons	12	
4	yes	FDG	parenchyma	8:2:18	
5a	yes	FDG	parenchyma	12	ANCOVA: covariate = age
5b	yes	FDG	parenchyma	12	intensity scaling prior to smoothing

*based on [31].

normalizing the template to this frame. The motion between the first and any other frame was tracked for each reference point, and the distance (in mm) the point had moved was computed. The maximum distance over the 5 reference points was used as 'motion amplitude' to quantitatively characterize the motion between the first and the considered frame (independent of the direction of the motion). Frames with a motion amplitude >4 mm were discarded (rationale: 4 mm is about half the spatial resolution in the reconstructed images, which has been shown to be about the threshold for relevant errors by mismatch between PET and low-dose CT for attenuation correction [24]). A motion-corrected static uptake image was obtained by summing the remaining frames after realignment.

FDG brain template

The default PET template provided by SPM8 is based on [O-15]-water perfusion PET images and, therefore, might not be optimal to guide stereotactical normalization of brain FDG PET images [25]. Therefore, a tracer-specific FDG PET template was generated from the 32 NC FDG PETs. In detail, for each NC, the motion-corrected FDG PET was co-registered to its baseline MPRAGE MRI (the first of the two baseline MPRAGE scans was used in all cases; unpreprocessed MRI data was downloaded from ADNI). Then, the MRI was segmented and stereotactically normalized using SPM's unified segmentation algorithm [26]. Unified segmentation was guided by freely available tissue probability maps (TPM) with 1 mm isotropic resolution generated from a sample of 662 healthy elderly subjects [27]. The latter might provide better performance in the elderly patients with suspected neurodegenerative disease than the 2 mm TPM from healthy young adults provided by SPM [28]. A more detailed description of the MRI processing can be found in [29]. The optimal MRI transformation was applied to the co-registered FDG PET to transform it from native patient space into the anatomical space of the Montreal Neurological Institute (MNI) [22]. After stereotactical normalization, intensity scaling was performed by global scaling (described below). A preliminary FDG PET template was obtained by averaging the scaled FDG PETs over all 32 NC subjects.

In a second step, all NCs were stereotactically normalized to the preliminary FDG template (PET-based normalization), intensity scaled (global scaling), and averaged to create the final FDG PET template. PET-based stereotactical normalization reduced the voxel-by-voxel coefficient of variance (COV) over the stereotactically normalized and scaled NC FDG PET images (Fig. 1, rationale: "the lower the variability in the control group the higher the power of voxel-based single subject analysis for detection of disease-related alterations of FDG uptake").

Stereotactical normalization

Stereotactical normalization as part of preprocessing for voxel-based statistical testing was PET-based in all subjects, including MCI and ADD patients as well as NC subjects. The rationale for this was that PET-based stereotactical normalization appears more relevant clinically, since an individual (high resolution) T1-weighted MRI is not always available in routine patient care.

Each individual FDG PET image was stereotactically normalized into MNI space using the normalization routine of SPM8 and SPM's default [O-15]-water PET template or the new custom-made FDG template. The following settings were used: no tem-

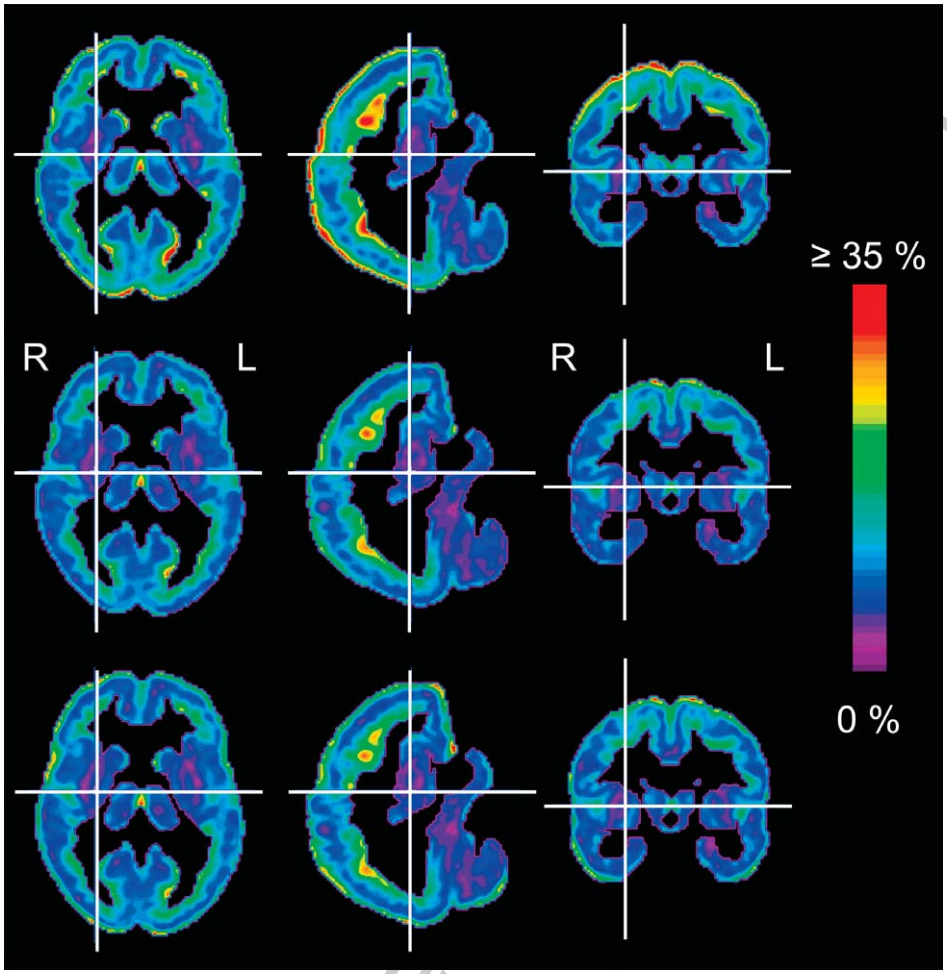

Fig. 1. Voxel-wise coefficient of variance (COV) over the 32 ADNI NC subjects for different methods of stereotactical normalization (comp. subsection 2.6). Top row: MRI-based stereotactical normalization using unified segmentation. Middle row: PET-based stereotactical normalization of the NCs using the FDG template as target. Bottom row: PET-based stereotactical normalization using an FDG template generated from the 32 ADNI ADD subjects as target. The stereotactically normalized PET images were scaled to the parenchyma mean before the COV was computed. The COV images were masked with the parenchyma mask for display purposes.

plate/source weighting, no template smoothing, source smoothing 8 mm, affine regularization to MNI, nonlinear frequency cut-off 25, nonlinear iterations 16, nonlinear regularization 1, preservation of concentration, trilinear interpolation and bounding box [–90 –126 –72; 90 90 108] mm with isotropic voxels of 2 mm edge length.

Smoothing

Stereotactically normalized images were smoothed by convolution with an isotropic 3-dimensional Gaussian kernel with full-width-at-half-maximum (FWHM) ranging from 8 mm to 18 mm in steps of 2 mm.

Intensity scaling

Intensity scaling was applied after smoothing as the last preprocessing step for voxel-based testing. The following scaling methods were implemented: conventional global scaling as implemented in SPM (‘proportional scaling’) [23, 30], parenchyma scaling, iterative parenchyma scaling (neglecting hypometabolic voxels by iterative parenchyma scaling), ’Yakushev‘ scaling (scaling factor based on hypermetabolic voxels after global scaling [31]), and scaling to the pons [32].

For conventional global scaling, the mean intensity M was computed over all voxels in the total image volume (including ‘air voxels’) and then the mean

intensity of all voxels with intensity ≥M/8 was used as reference value for scaling, i.e., each voxel value was divided by the reference value.

For parenchyma scaling, the reference value was computed as the mean voxel intensity within a mask that had been created by thresholding the custom FDG template at a voxel intensity value of 1.45 (Fig. 2). A similar mask has previously been created by the union of the a priori images of gray and white matter provided by SPM, each thresholded at a given probability [33]. Parenchyma scaling eliminates variability due to inter-subject variation of extracranial FDG uptake (scalp, nasopharyngeal space, etc.).

For iterative parenchyma scaling, brain regions with significant hypometabolism in voxel-based testing at the liberal significance level of $p \leq 0.01$ (uncorrected for multiple testing) in the i-th iteration were excluded from the computation of the reference value for the $(i + 1)$-th iteration [34]. The iteration was stopped when the relative change of the reference value dropped below 0.2% or after a maximum of 10 iterations (the latter stop criterion was not reached in any subject). Scaling of the NCs was adjusted during each iteration.

For pons scaling, the mean intensity within a predefined pons mask was used as reference value [32]. The pons mask was based on the pons region of interest (ROI) provided by the WFU PickAtlas (human atlas, TD lobes) [35]. Slight manual adjustment of the ROI was performed to adapt it to the customized FDG PET template. Four of the 108 MCI subjects were excluded from pons scaling, because the pons had not completely been within the field-of-view of the PET acquisition in these subjects.

Voxel-based testing

For each MCI subject, the scaled, smoothed, and stereotactically normalized FDG PET image was compared voxel-by-voxel against the group of NC subjects using the two-sample *t*-test [36] implemented in SPM with the following parameter settings: grand mean scaling = no, ANCOVA = no, no masking, no global calculation, no global normalization (age was used as covariate in setting 5a, Table 2). Scaling was turned off, since the images had been scaled during preprocessing (see below). For each setting of the single subject analysis, preprocessing of NC subjects was exactly the same as for MCI subjects.

T-sum score

The t-sum score as proposed by Herholz and coworkers was computed by summing the t-values from voxel-based testing of an MCI subject over all voxels within a binary 'ADD mask'. This ADD mask is intended to delineate the brain regions with AD-specific reduction of FDG uptake [37]. The ADD mask was generated by voxel-based group testing for reduced FDG uptake in the ADNI ADD patients versus the ADNI NC subjects included in the present study (uncorrected $p \leq 0.005$, cluster size ≥125 voxels = 1 ml). Since interactions between the ADD mask and other preprocessing steps cannot be ruled out (with stereotactical normalization, for example), the ADD mask was generated separately for each setting of the single subject analysis in order to avoid bias by a fixed predefined mask. A representative ADD mask is shown in Fig. 3.

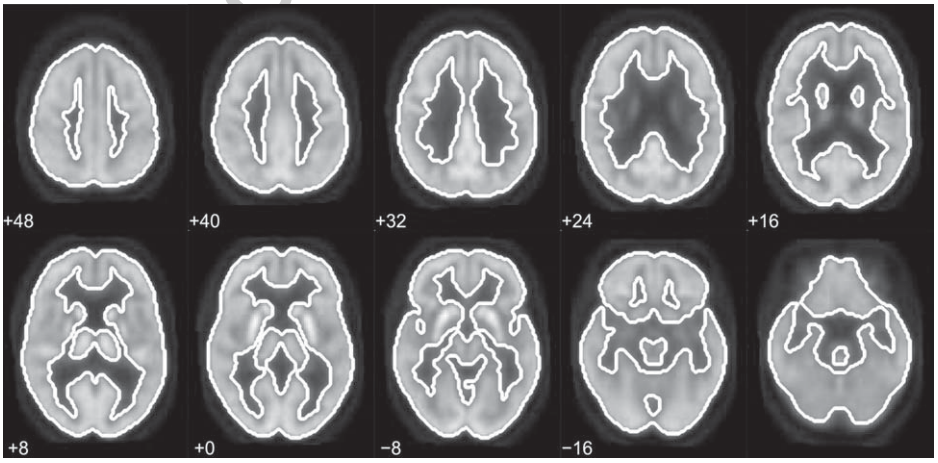

Fig. 2. Parenchyma mask overlaid to the FDG template.

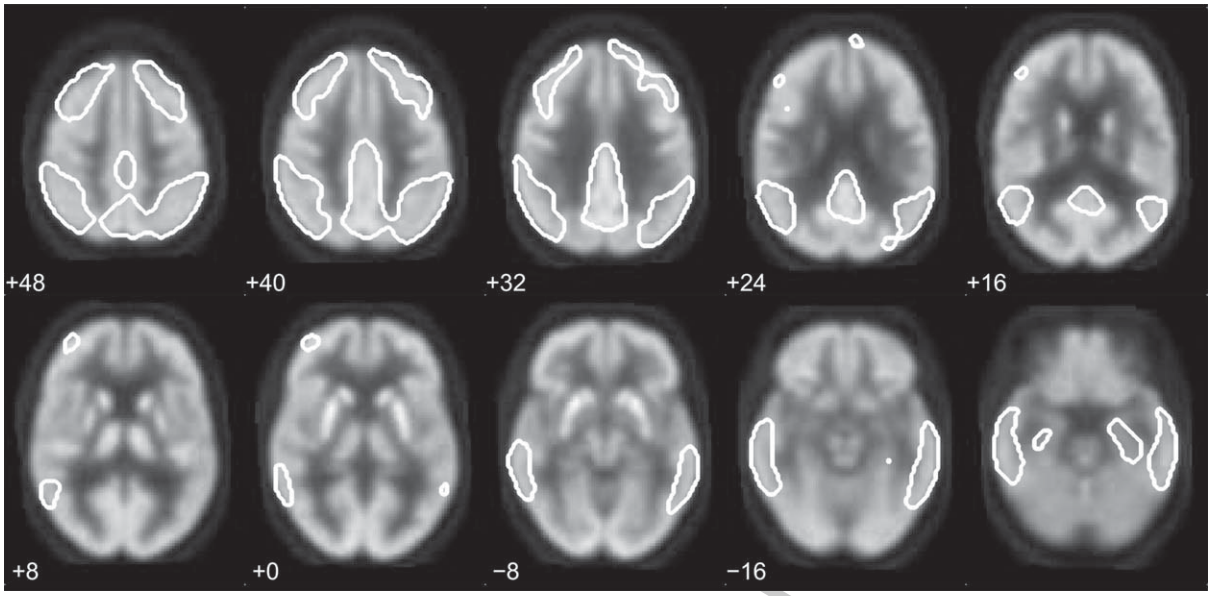

Fig. 3. Representative ADD mask (generated by ADD versus NC group testing with frame-by-frame motion correction, FDG template, parenchyma scaling, 12 mm smoothing) overlaid to the FDG template.

Receiver operating characteristic analysis

The power of the t-sum score for differentiation between 'MCI-to-AD converter' and 'MCI stable' was analyzed using receiver operating characteristic (ROC) analysis. The area AUC under the ROC curve was used as performance measure. The nonparametric DeLong test for paired samples was used for comparing the AUC between the t-sum ROC curves for different parameters settings [38].

The AUC does not require the selection of a cut-off and, therefore, is not affected by any limitations of the cut-off selection process, in contrast to sensitivity, specificity and predictive values. This also simplifies comparison of diagnostic or prognostic utility across methods and studies.

Head-to-head comparison against another method

For head-to-head comparison with optimized SPM8 single subject processing, the semi-quantitative brain FDG PET parameters of ADNI subjects made available by Foster and co-workers via the ADNI website (upload on March 17, 2015) were downloaded (on May 20, 2015). The following 6 semi-quantitative parameters derived by using routines from the Neurostat software package [17] are provided: (i) mean FDG uptake in the bilateral association cortices scaled to mean FDG uptake in the pons (denoted AVEASSOC by Foster et al.), (ii) mean FDG uptake in the frontal cortex scaled to mean FDG uptake in the pons (AVEFRONT), (iii) number of (hypometabolic) voxels $\geq$2 standard deviations and <3 standard deviations below the mean in the control group (X2SDSIGPXL), (iv) number of (hypometabolic) voxels $\geq$3 standard deviations below control mean (X3SDSIGPXL), (v) sum over all voxel z-scores $\geq$2 standard deviations below control mean (SUMZ2), and (vi) sum over all voxel z-scores $\geq$3 standard deviations below control mean (SUMZ3). These semi-quantitative parameters were available for 107 of the 108 ADNI MCI subjects included in the present study (see above).

Validation

Inclusion of the MCI subjects described (and used in the analyses described so far) was based on a search of the ADNI database in March 2014. For generation of an independent validation sample of ADNI MCI subjects, the search was repeated in August 2015 using exactly the same eligibility criteria. This resulted in a total of 241 additional MCI subjects who had completed the 3 years follow-up in the meanwhile (ADNI participant roster IDs are listed in the Supplementary Material). 181 of these MCI subjects had been cognitively stable for 3 years; the remaining 60 had converted to ADD. Subject demographics of the validation sample are given in Table 3.

Table 3
Baseline characteristics of the validation sample of ADNI MCI subjects. (MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; FAQ, functional activities questionnaire; ABETA142, concentration of amyloid-β 1-42 peptide in cerebrospinal fluid; t-sum score for the following setting: motion correction, custom FDG template, parenchyma scaling, 12 mm smoothing)

Group	*n*	age* (y)	gender†	education* (y)	FAQ*	MMSE*	ABETA142*‡ (pg/ml)	t-sum score*
MCI stable	181	70.5 ± 7.2	86/95	16.3 ± 2.6	1.58 ± 2.66	28.2 ± 1.6	143.8 ± 29.5	14604 ± 16754
MCI converter	60	73.7 ± 6.5	23/37	16.2 ± 2.7	5.44 ± 4.83	27.2 ± 1.7	152.5 ± 47.6	28356 ± 20085

*mean ± SD. †female/male. ‡ABETA142 available in 9 MCI stables and 17 MCI converters (ADNI table "UPENNBIOMK.csv").

Brain FDG PETs of the MCI subjects in the validation sample were processed as described above. The impact of the SPM8 parameter setting on the differentiation between 'MCI-to-AD converters' and 'MCI stables' was again assessed via comparison of the AUC under the ROC curve of the t-sum score.

In the validation sample, overall accuracy, sensitivity, specificity, and predictive values of the t-sum score were estimated in addition to the AUC. The cut-off was selected according to the Youden criterion [39], i.e., by maximizing the Youden index J = sensitivity + specificity – 1, which is symmetric in sensitivity and specificity and, therefore, imposes equal penalty on false positive and false negative classifications. Although maximization of the Youden index is a rather simple model, it might be affected by statistical noise. Thus, overfitting cannot be ruled out so that estimates of diagnostic accuracy measures are most likely overly optimistic. In order to correct for overfitting, 100 repeats of 20-fold cross-validation were performed. Estimating errors of accuracy estimates by variance across repeats of cross-validation is limited by the risk of duplicated training samples. We therefore used Equation (3) in [40] to estimate the 95% confidence interval of the accuracy measures.

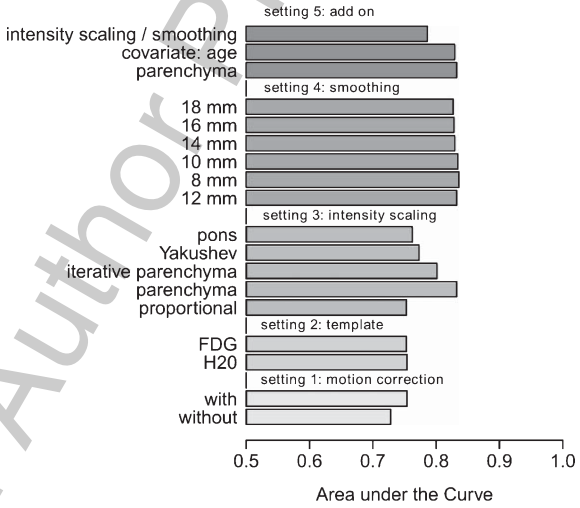

Fig. 4. Area under the ROC curve for the different settings of the SPM8 processing pipeline in the original sample of 108 MCI subjects.

RESULTS

Image processing worked properly in all subjects (according to visual inspection of stereotactically normalized images and statistical maps), i.e., there was no failure in any of the subjects (108 + 241 = 349 ADNI MCI subjects, 32 ADNI normals, and 32 ADNI ADD patients), although no subject was excluded based on technical constraints such as poor PET image quality. This demonstrates the robustness of the fully automatic SPM processing pipeline, which is an important prerequisite for use in everyday clinical routine. The processing time for single subject analysis was about 4 minutes on a standard PC, which is compatible with busy clinical workflow.

The results of the ROC analyses in the original sample of 108 MCI subjects are summarized in Fig. 4. With the SPM default setting for voxel-based single subject analysis, the t-sum score provided an AUC of 0.728 for the differentiation between 'MCI-to-AD converter' and 'MCI stable'. Frame-by-frame motion correction improved the AUC to 0.754. Whereas replacing SPM's [O-15]-water template by the custom FDG template did not further improve AUC (0.753), parenchyma scaling (instead of proportional scaling) resulted in considerable further improvement to AUC = 0.832. The total improvement from AUC = 0.728 for the default setting to AUC = 0.832 for the 'optimized' setting was statistically significant (two-sided $p = 0.046$).

'Simple', i.e., non-iterative parenchyma scaling performed better than all other scaling methods, including iterative parenchyma scaling. The degree of smoothing had negligible impact on the AUC, at least with parenchyma scaling. Reversed order of smoothing and intensity scaling, i.e., intensity scaling prior to smoothing, resulted in reduction of AUC (0.786). Taking into account the subjects' age as covariate in the statistical test did not further improve the AUC (0.829).

Among the 6 semi-quantitative brain FDG PET parameters provided by Foster and co-workers, the mean FDG uptake in the association cortices scaled to the mean pons uptake (AVEASSOC) achieved the highest AUC with a value of 0.745 (Fig. 5). The difference compared to AUC = 0.832 achieved with the optimized SPM8 processing showed a tendency towards statistical significance (two-sided $p = 0.080$).

ROC analysis of the SPM8 t-sum scores in the validation sample of 241 ADNI MCI subjects resulted in AUC of 0.675 and 0.746 with the default and with the optimized parameter setting, respectively. The difference was statistically significant (two-sided $p = 0.048$). Cross-validated overall accuracy, sensitivity, specificity, and predictive values are summarized in Table 4. All these measures were considerably larger for the optimized setting than for the default setting. The difference was highly significant statistically, as indicated by the fact that the 95% confidence intervals did not even overlap (except for the negative predictive value for which there was a small overlap).

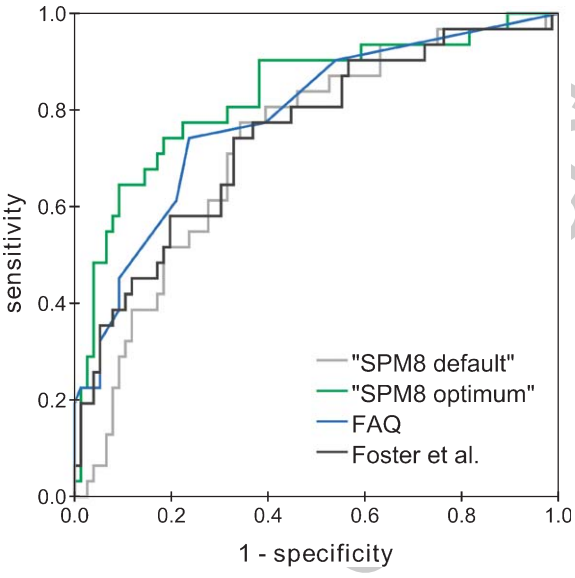

Fig. 5. ROC curves for prognosis of MCI-to-AD conversion in the original sample of 108 MCI subjects. "SPM8 default" and "SPM8 optimum" are for the t-sum score obtained with default and optimum SPM8 setting, respectively. "FAQ" is for the total score of the functional activity questionnaire. "Foster et al" is for the average FDG uptake in the association cortices scaled to mean FDG uptake in the pons (AVEASSOC) provided by Foster et al. on the ADNI website (subsection 2.13). All ROC curves are for the same 107 MCI subjects. The ROC curves presented in this figure use only 107 of the 108 MCI subjects included in the present study, since AVEASSOC was not available for one subject (RID 135).

DISCUSSION

The aim of this study was to optimize the parameter settings of voxel-based SPM single subject analysis for prediction of MCI-to-AD conversion within 3 years by brain FDG PET. The following aspects of the processing pipeline were considered: frame-by-frame motion correction, [O-15]-water versus FDG template, spatial smoothing, and intensity scaling.

The first step towards improved single subject analysis of brain FDG PET was motion correction. The majority of the ADNI brain FDG PETs included in the present study comprised 6 frames of 5 min duration from 30 to 60 min post injection. Motion correction was performed frame-by-frame by realigning frames 2 to 6 with the first frame. With modern PET/CT (and PET/MR systems), PET emission recording is in list mode which allows arbitrary framing of the acquired data during image reconstruction. Modern PET/CT (and PET/MR systems) also provide high sensitivity for the detection of radioactive decays so that adequate statistical image quality requires less than 30 min acquisition time (after injection of a standard dose of about 200 MBq FDG [20, 21]). In our department, we perform a 15-min acquisition 40 ± 5 min post injection which then is reconstructed into 15 frames of 1 min duration for frame-by-frame motion correction.

The second important factor was intensity scaling which has been found to have a large impact on the performance of single subject analysis of brain FDG PET also in previous studies [32, 41–44]. In the present study, direct voxel-wise scaling to the mean intensity in a predefined gray and white matter (parenchyma) mask provided the best performance. Compared to the widely used proportional scaling method implemented in SPM, the AUC increased from 0.754 to 0.832. This most likely is explained by elimination of extra variability associated with inter-subject differences of extracranial FDG uptake, for example in the scalp and in nasopharyngeal space. Proportional scaling typically averages the voxel intensity over all tissues with visually detectable FDG uptake including extracranial structures.

A limitation of simple scaling to the mean intensity in the fixed parenchyma mask is that this mask includes brain regions affected by reduced FDG uptake in patients with ADD and MCI due to AD, which results in underestimation of the true reference value. The latter causes overestimation of scaled FDG uptake which results in reduced power for the detection of hypometabolism (and spurious hypermetabolism) [45]. This effect can be avoided either by using a fixed

Table 4
Area (AUC) under the ROC curve, cut-off value on the t-sum score determined by the maximum Youden index, and accuracy measures for prediction of ADNI-MCI to ADD conversion within 36 months by the t-sum score computed by the SPM8 single subject processing pipeline with default or optimized setting in the validation sample of MCI subjects. All accuracy measures were cross-validated by 100 repeats of 20-fold cross-validation. 95% confidence intervals (CI) are given in brackets. The 95%-CI for the AUC was obtained as described in [38], the 95%-CIs for the accuracy measures were estimated according to [40]. The standard deviation of the cut-off is given in round brackets. (PPV, positive predictive value; NPV, negative predictive value)

Setting	AUC	cut-off	Cross validated				
			accuracy	sensitivity	specificity	PPV	NPV
default	0.675 [0.60–0.75]	21735 (8172)	0.57 [0.52–0.62]	0.58 [0.53–0.63]	0.56 [0.51–0.62]	0.31 [0.26–0.36]	0.80 [0.76–0.84]
optimized	0.746 [0.67–0.82]	18774 (1199)	0.68 [0.63–0.73]	0.70 [0.65–0.75]	0.68 [0.63–0.73]	0.42 [0.37–‘0.47]	0.87 [0.83–0.90]

anatomical reference region which is not affected by AD or by using data-driven techniques to automatically eliminate affected regions based on statistical criteria. Methods of both types were tested in the present study. The pons was used as AD-unaffected reference region, based on the finding of preserved pontine glucose metabolism in AD by Minoshima and co-workers [32]. Iterative parenchyma scaling and the Yakushev method [31] were used as data driven techniques. However, none of these methods performed better than simple parenchyma scaling. We hypothesize that this is related to statistical noise of the reference value: the larger the reference region the smaller the statistical noise of the reference value obtained by averaging the intensity over all voxels within the reference region. The results of the present study suggest that reduction of statistical noise by the large size of the parenchyma reference region overcompensates the impact of systematic underestimation of the reference value caused by AD-related hypometabolism in the parenchyma reference region, at least for prediction of MCI-to-AD conversion. With data-driven methods, the reference region varies between tests which might be considered a disadvantage in single subject analysis (inter-subject variability of test performance).

The mean of the voxel intensity over all voxels within the reference region was used as reference value to characterize the FDG uptake in the reference region. We also tested the median instead of the mean (results not shown). The rationale for this was that the median might be less sensitive to moderate (disease-related) intensity changes which primarily affect the intensity spectrum above the median and, therefore, do not change the median. However, using the median did not improve prognostic accuracy (for example, parenchyma scaling: AUC = 0.798 versus 0.832 with median and mean, respectively).

Pons scaling performed slightly worse than parenchyma scaling (AUC = 0.762 versus 0.832). In addition, when using the pons as reference region, it is mandatory to carefully check in each single subject whether the pons has been completely within the field-of-view of the PET acquisition. Failure to do so might result in false negative single subject analysis due to severe underestimation of pontine FDG uptake.

Concerning the brain template used to define the target space for stereotactical normalization, there was no difference with respect to MCI-to-AD prognosis between the [O-15]-water template provided by SPM and a custom-made tracer-specific template generated from FDG PETs of age-matched ADNI NC subjects. We made some attempts to improve the FDG PET template, for example by using the 32 ADNI ADD patients included in the present study rather than the ADNI NC subjects to generate the template. However, this ADD FDG template resulted in increased voxel-by-voxel coefficient of variance over the stereotactically normalized and parenchyma scaled NC FDG PET images (Fig. 1). Although this did not degrade the accuracy for prediction of MCI-to-AD conversion (0.831 versus 0.832 for ADD FDG and NC FDG template, respectively), the NC FDG PET template described in above was used for all analyses presented here.

It has been previously shown that the template can have a considerable impact on the performance of single subject analysis [46, 47]. That the impact was small in the present study might be explained by the fact that O-15-water and FDG PET provide rather similar images (both are considered surrogate of synaptic activity).

MRI-based stereotactical normalization of FDG PET was performed only during template generation (see Materials and Methods), although MRI-based stereotactical normalization has been shown to improve the power of voxel-based testing compared to PET-based stereotactical normalization [48]. However, in everyday clinical patient care, MRI is not available in all patients. Therefore, we recommend PET-based stereotactical normalization for clinical routine, in order to guarantee the same processing in all patients.

Fully consistent processing in all patients appears important in clinical routine to guarantee stable performance of statistical single subject analysis.

The amount of smoothing, too, had only very small impact on the prognosis of MCI-to-AD conversion, even though it was varied in the rather large range from 8 to 18 mm FWHM. The aim of spatial smoothing is (i) to cope with residual inter-subject variability after stereotactical normalization and (ii) to increase the signal-to-noise ratio for improved statistical power for detection of hypometabolic clusters. It has been suggested that spatial smoothing should match the spatial extent of the effect to be detected [49, 50]. Thus, one would expect rather strong smoothing to work best for the detection of the spatially rather extended AD-characteristic pattern of hypometabolism in FDG PET (typical volume of the ADD mask was about 370 ml, comp. Fig. 3). The fact that smoothing had only a very small effect in the present study might be explained by some interaction with the parenchyma mask used as reference region for intensity scaling. The parenchyma mask is rather narrow (Fig. 2) so that increasing the width of the Gaussian smoothing kernel beyond the radial width of the mask is expected to have only a small effect on voxel intensities within the parenchyma mask. In order to test this hypothesis, variation of the smoothing kernel was repeated in combination with proportional scaling. Proportional scaling typically includes the whole head as reference region and, therefore, should be more sensitive to smoothing than parenchyma scaling. This was confirmed: with proportional scaling, the AUC of the t-sum score increased with the amount of smoothing, from AUC = 0.749 at 8 mm kernel width to AUC = 0.767 at 14 mm to AUC = 0.782 at 18 mm. This indicates that the impact of spatial smoothing depends on the reference region for intensity scaling: the impact is large for proportional scaling, but small for parenchyma scaling. Stability of parenchyma scaling with respect to the amount of smoothing might be considered an advantage, particularly in multi-site and single-site/multi-camera settings in which the spatial resolution of the tested images depends also on camera-specific PET acquisition and reconstruction protocols.

It might be noted that smoothing with 8 mm FWHM provided greater AUC than smoothing with 12 mm FWHM (Fig. 4), although the difference was very small and far from being statistically significant. Nevertheless, we recommend 12 mm rather than 8 mm smoothing. The rationale for this is that 12 mm is better in compensating inter-scan variability in spatial resolution in the original brain FDG PET images. The variability of spatial resolution in ADNI PET images is rather small due to homogenization of the acquisition protocol across different PET scanners in the ADNI. Variability is expected to be larger in settings with less homogenized acquisition protocols. In these cases, 12 mm smoothing is more effective than 8 mm smoothing in reducing non-physiological inter-subject variability of FDG uptake.

Accounting for the subjects' age as covariate in the statistical testing did not improve the performance of FDG PET single subject analysis for the prognosis of MCI-to-AD conversion. Therefore, age correction does not appear mandatory for this task, at least as long as patients and control group for voxel-based testing are well matched with respect to age (all groups were very well matched with respect to age in the present study, Table 1). Age correction might have even detrimental effects, particularly if some of the older subjects in the control group suffer from preclinical AD. In this case, age correction will correct not only for effects of healthy aging on FDG uptake but, to some extent, also for AD-typical hypometabolism. The latter will reduce the power for detection of the AD pattern in patients to be diagnosed.

Finally, switching the order of image smoothing and intensity scaling, i.e., performing intensity scaling prior to smoothing, resulted in considerable deterioration of the prognostic power and, therefore, cannot be recommended.

Altogether, optimizing the parameter setting of the SPM processing pipeline improved the AUC of the t-sum score for differentiation between MCI-to-AD converters and MCI stable subjects by about 14% from 0.728 (SPM default settings) to 0.832 (Fig. 4, 5). The effect was statistically significant (two-sided $p = 0.046$). To put this into perspective, it might be noted that many studies suggest a capping of prognostic accuracy in MCI patients considerably below 100%, independent of the criteria and/or biomarkers used [28, 51–54]. Therefore, not only the relative improvement by 14%, but also the final absolute value of AUC = 0.832 appears rather remarkable, particularly as it can be achieved rather easily without extra costs, i.e., using standard FDG PET acquisition protocols (no dynamic imaging of the full time course of FDG concentration in tissue starting with i.v. injection required, no blood sampling, no tracer kinetic modeling) and the freely available SPM software package with only minor adaptions.

This finding was confirmed in an independent validation sample of 241 further ADNI MCI subjects. The relative improvement in AUC was about the same in

the original and in the validation sample: 14% and 11%, respectively. However, it should be noted that the absolute AUC values were lower in the validation sample: 0.675 versus 0.728 with default parameter settings, 0.746 versus 0.832 with optimized parameter settings of the SPM8 processing pipeline. We hypothesize that this is related to the fact that the original sample mainly included late MCI subjects from the ADNI-1 phase, whereas the validation sample included many subjects from ADNI-GO and ADNI-2 with early MCI in which prognosis is expected to be more difficult than in late MCI. To some extent this is reflected by the fraction of MCI-to-AD converters in both samples, as it is smaller in the validation sample (25% versus 29%).

The power of brain FDG PET for the prognosis of MCI-to-AD conversion has been investigated in several previous studies using different methods. Arbizu and coworkers, who evaluated a variant of the AD-related hypometabolic convergence index [55] for the prognosis of MCI-to-AD conversion in 121 ADNI MCI subjects reported an AUC of 0.804 for a multivariate model including the posterior cingulate index together with age, gender, MMSE, and ApoE4 status [51]. Morbelli and coworkers, who evaluated the AD t-sum score in 127 MCI patients from the European Alzheimer's Disease Consortium network reported an accuracy of 79.6% for prediction of MCI-to-AD conversion [54]. In the present study, maximum accuracy of the t-sum score was 83.3%.

In a recent study on multimodal prediction of MCI-to-AD conversion we found the sum score of the functional activity questionnaire (FAQ) to be the best single feature [56]. For the original $n = 108$ ADNI MCI sample included in the present study, ROC analysis of this sum score (FAQTOTAL) resulted in AUC = 0.786 (Fig. 5). Thus, the t-sum score from the single subject analysis of FDG PET performed better than the FAQ only after optimizing the processing protocol. This finding underpins the necessity of optimizing single subject analysis of brain FDG PET, since otherwise the additional benefit from FDG PET might be rather small, particularly when considering the cost-benefit ratio.

Concerning the parameter setting for single subject analysis of brain FDG PET within the SPM framework, Perani and colleagues optimized an SPM5-based processing pipeline with respect to differential diagnosis of neurodegenerative diseases including AD, frontotemporal lobar degeneration (FTLD), and dementia with Lewy bodies [57]. Visual interpretation of the statistical parametric maps improved the differentiation between AD and FTLD compared to visual interpretation of the raw FDG uptake images. The optimized SPM5 processing pipeline used PET-based stereotactical normalization (with very similar parameter settings as in the present study) to a dementia-specific FDG template, proportional intensity scaling followed by smoothing with an isotropic 3-dimensional Gaussian kernel of 8 mm FWHM. The impact of extracranial inter-subject variability of FDG uptake was taken into account by an explicit mask to restrict voxel-based testing to the brain. The results of this previous study are in good agreement with the results of the present study. Minor differences of the optimized processing pipeline between the two studies might be explained by the different task for which the processing was optimized: differential diagnosis of neurodegenerative diseases in the study by Perani and colleagues versus MCI-to-AD conversion in the present study. Visual interpretation of statistical parametric maps in the Perani study versus quantitative t-sum score analysis in the present study might also have contributed to the minor differences.

Limitations of the present study include the use of a fixed time interval for prediction (3 years) and that all analyses were strictly univariate. Future studies might use Kaplan-Meier analysis and/or multivariate Cox regression to better account for inter-subject variability of follow-up duration and time to conversion as well as to assess the incremental value of FDG PET over other features used for the diagnosis of AD.

conclusion

Optimizing SPM for voxel-based single subject analysis of brain FDG PET can provide considerable improvement of MCI-to-AD prediction. To achieve this we recommend: (i) reconstruction (of list mode data) into several frames of constant duration (1 to 5 min), (ii) frame-by-frame motion correction by realignment to the reference frame (chronologically closest to the low-dose CT for attenuation correction), (iii) discarding all frames with more than 4 mm displacement with respect to the reference frame in order to avoid attenuation artifacts (if the spatial mismatch with respect to the low-dose CT for attenuation correction can be corrected frame-by-frame during image reconstruction, this might be preferred), (iv) add the selected frames to generate one static FDG uptake image (5 min total duration provides sufficient statistical image quality in most cases), (v) 3-dimensional spatial smoothing with an isotropic Gaussian kernel with 12 mm FWHM, (vi) voxel-wise intensity scaling to the mean tracer uptake in brain parenchyma using a predefined mask in template space. These steps can

easily be implemented as a fully automatic processing pipeline.

ACKNOWLEDGMENTS

The authors L.S. and R.B. were supported by the European Regional Development Fund of the European Union (reference 10153407 and 10153463). P.S. and L.S. are employees of jung diagnostics GmbH.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/15-0814).

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: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.

The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://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.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-150814.

REFERENCES

[1] Phelps ME, Huang SC, Hoffman EJ, Selin C, Sokoloff L, Kuhl DE (1979) Tomographic measurement of local cerebral glucose metabolic rate in humans with (F-18)2-fluoro-2-deoxy-D-glucose: Validation of method. *Ann Neurol* **6**, 371-388.

[2] Minoshima S, Giordani B, Berent S, Frey KA, Foster NL, Kuhl DE (1997) Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Ann Neurol* **42**, 85-94.

[3] Herholz K (2010) Cerebral glucose metabolism in preclinical and prodromal Alzheimer's disease. *Expert Rev Neurother* **10**, 1667-1673.

[4] Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, Reiman EM, Holthoff V, Kalbe E, Sorbi S, Diehl-Schmid J, Perneczky R, Clerici F, Caselli R, Beuthien-Baumann B, Kurz A, Minoshima S, de Leon MJ (2008) Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *J Nucl Med* **49**, 390-398.

[5] Friedland RP, Budinger TF, Ganz E, Yano Y, Mathis CA, Koss B, Ober BA, Huesman RH, Derenzo SE (1983) Regional cerebral metabolic alterations in dementia of the Alzheimer type: Positron emission tomography with [18F]fluorodeoxyglucose. *J Comput Assist Tomogr* **7**, 590-598.

[6] Silverman DH (2004) Brain 18F-FDG PET in the diagnosis of neurodegenerative dementias: Comparison with perfusion SPECT and with clinical evaluations lacking nuclear imaging. *J Nucl Med* **45**, 594-607.

[7] Silverman DH, Small GW, Chang CY, Lu CS, Kung De Aburto MA, Chen W, Czernin J, Rapoport SI, Pietrini P, Alexander GE, Schapiro MB, Jagust WJ, Hoffman JM, Welsh-Bohmer KA, Alavi A, Clark CM, Salmon E, de Leon MJ, Mielke R, Cummings JL, Kowell AP, Gambhir SS, Hoh CK, Phelps ME (2001) Positron emission tomography in evaluation of dementia: Regional brain metabolism and long-term outcome. *JAMA* **286**, 2120-2127.

[8] Jagust W (2006) Positron emission tomography and magnetic resonance imaging in the diagnosis and prediction of dementia. *Alzheimers Dement* **2**, 36-42.

[9] Foster NL, Heidebrink JL, Clark CM, Jagust WJ, Arnold SE, Barbas NR, DeCarli CS, Turner RS, Koeppe RA, Higdon R, Minoshima S (2007) FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. *Brain* **130**, 2616-2635.

[10] Minoshima S, Foster NL, Sima AA, Frey KA, Albin RL, Kuhl DE (2001) Alzheimer's disease versus dementia with Lewy bodies: Cerebral metabolic distinction with autopsy confirmation. *Ann Neurol* **50**, 358-365.

[11] Foster NL, Wang AY, Tasdizen T, Fletcher PT, Hoffman JM, Koeppe RA (2008) Realizing the potential of positron emission tomography with 18F-fluorodeoxyglucose to improve the treatment of Alzheimer's disease. *Alzheimers Dement* **4**, S29-S36.

[12] Jagust WJ, Haan MN, Eberling JL, Wolfe N, Reed BR (1996) Functional imaging predicts cognitive decline in Alzheimer's disease. *J Neuroimaging* **6**, 156-160.

[13] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic

guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270-279.

[14] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.

[15] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[16] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, Gauthier S, Selkoe D, Bateman R, Cappa S, Crutch S, Engelborghs S, Frisoni GB, Fox NC, Galasko D, Habert MO, Jicha GA, Nordberg A, Pasquier F, Rabinovici G, Robert P, Rowe C, Salloway S, Sarazin M, Epelbaum S, de Souza LC, Vellas B, Visser PJ, Schneider L, Stern Y, Scheltens P, Cummings JL (2014) Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurol* **13**, 614-629.

[17] Minoshima S, Frey KA, Koeppe RA, Foster NL, Kuhl DE (1995) A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET. *J Nucl Med* **36**, 1238-1248.

[18] Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE (2007) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press.

[19] Burdette JH, Minoshima S, Vander Borght T, Tran DD, Kuhl DE (1996) Alzheimer disease: Improved visual interpretation of PET images by using three-dimensional stereotaxic surface projections. *Radiology* **198**, 837-843.

[20] Varrone A, Asenbaum S, Vander Borght T, Booij J, Nobili F, Nagren K, Darcourt J, Kapucu OL, Tatsch K, Bartenstein P, Van Laere K, European Association of Nuclear Medicine Neuroimaging, Committee (2009) EANM procedure guidelines for PET brain imaging using [18F]FDG, version 2. *Eur J Nucl Med Mol Imaging* **36**, 2103-2110.

[21] Waxman AD, Herholz K, Lewis DH, Herscovitch P, Minoshima S, Ichise M, Drzezga AE, Devous MD, Mountz JM (2009) *Society of Nuclear Medicine procedure guideline for FDG PET brain imaging*. Society of Nuclear Medicine, Reston, VA.

[22] Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner JT, Penny WD, eds. (2004) *Human Brain Function*, Academic Press, San Diego.

[23] Acton PD, Friston KJ (1998) Statistical parametric mapping in functional neuroimaging: Beyond PET and fMRI activation studies. *Eur J Nucl Med* **25**, 663-667.

[24] Andersson JL, Vagnhammar BE, Schneider H (1995) Accurate attenuation correction despite movement during PET imaging. *J Nucl Med* **36**, 670-678.

[25] Della Rosa PA, Cerami C, Gallivanone F, Prestia A, Caroli A, Castiglioni I, Gilardi MC, Frisoni G, Friston K, Ashburner J, Perani D, Consortium EADC-PET (2014) A standardized [18F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. *Neuroinformatics* **12**, 575-593.

[26] Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* **26**, 839-851.

[27] Lemaitre H, Crivello F, Grassiot B, Alperovitch A, Tzourio C, Mazoyer B (2005) Age- and sex-related effects on the neuroanatomy of healthy elderly. *Neuroimage* **26**, 900-911.

[28] Suppa P, Hampel H, Spies L, Fiebach JB, Dubois B, Buchert R (2015) Fully automated atlas-based hippocampus volumetry for clinical routine: Validation in subjects with mild cognitive impairment from the ADNI cohort. *J Alzheimers Dis* **44**, 183-193.

[29] Suppa P, Anker U, Spies L, Bopp I, Ruegger-Frey B, Klaghofer R, Gocke C, Hampel H, Beck S, Buchert R (2014) Fully automated atlas-based hippocampal volumetry for detection of Alzheimer's disease in a memory clinic setting. *J Alzheimers Dis* **44**, 183-193.

[30] Stamatakis EA, Glabus MF, Wyper DJ, Barnes A, Wilson JT (1999) Validation of statistical parametric mapping (SPM) in assessing cerebral lesions: A simulation study. *Neuroimage* **10**, 397-407.

[31] Yakushev I, Hammers A, Fellgiebel A, Schmidtmann I, Scheurich A, Buchholz HG, Peters J, Bartenstein P, Lieb K, Schreckenberger M (2009) SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. *Neuroimage* **44**, 43-50.

[32] Minoshima S, Frey KA, Foster NL, Kuhl DE (1995) Preserved pontine glucose metabolism in Alzheimer disease: A reference region for functional brain image (PET) analysis. *J Comput Assist Tomogr* **19**, 541-547.

[33] Wenzel F, Young S, Wilke F, Apostolova I, Arlt S, Jahn H, Thiele F, Buchert R (2010) B-spline-based stereotactical normalization of brain FDG PET scans in suspected neurodegenerative disease: Impact on voxel-based statistical single-subject analysis. *Neuroimage* **50**, 994-1003.

[34] Andersson JL (1997) How to estimate global activity independent of changes in local activity. *Neuroimage* **6**, 237-244.

[35] Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* **19**, 1233-1239.

[36] Muhlau M, Wohlschlager AM, Gaser C, Valet M, Weindl A, Nunnemann S, Peinemann A, Etgen T, Ilg R (2009) Voxel-based morphometry in individual patients: A pilot study in early Huntington disease. *AJNR Am J Neuroradiol* **30**, 539-543.

[37] Herholz K, Salmon E, Perani D, Baron JC, Holthoff V, Frolich L, Schonknecht P, Ito K, Mielke R, Kalbe E, Zundorf G, Delbeuck X, Pelati O, Anchisi D, Fazio F, Kerrouche N, Desgranges B, Eustache F, Beuthien-Baumann B, Menzel C, Schroder J, Kato T, Arahata Y, Henze M, Heiss WD (2002) Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage* **17**, 302-316.

[38] DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837-845.

[39] Youden WJ (1950) Index for rating diagnostic tests. *Cancer* **3**, 32-35.

[40] Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence (IJCAI)*.

[41] Buchert R, Wilke F, Chakrabarti B, Martin B, Brenner W, Mester J, Clausen M (2005) Adjusted scaling of FDG positron emission tomography images for statistical evaluation in patients with suspected Alzheimer's disease. *J Neuroimaging* **15**, 348-355.

[42] Herholz K, Perani D, Salmon E, Franck G, Fazio F, Heiss WD, Comar D (1993) Comparability of FDG PET studies in probable Alzheimer's disease. *J Nucl Med* **34**, 1460-1466.

[43] Yakushev I, Landvogt C, Buchholz HG, 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.

[44] Borghammer P, Jonsdottir KY, Cumming P, Ostergaard K, Vang K, Ashkanian M, Vafaee M, Iversen P, Gjedde A (2008) Normalization in PET group comparison studies–the importance of a valid reference region. *Neuroimage* **40**, 529-540.

[45] Borghammer P, Cumming P, Aanerud J, Gjedde A (2009) Artefactual subcortical hyperperfusion in PET studies normalized to global mean: Lessons from Parkinson's disease. *Neuroimage* **45**, 249-257.

[46] Della Rosa PA, Cerami C, Gallivanone F, Prestia A, Caroli A, Castiglioni I, Gilardi MC, Frisoni G, Friston K, Ashburner J, Perani D, Consortium E-P (2014) A standardized [(18)F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. *Neuroinformatics* **12**, 575-593.

[47] Gispert JD, Pascau J, Reig S, Martinez-Lazaro R, Molina V, Garcia-Barreno P, Desco M (2003) Influence of the normalization template on the outcome of statistical parametric mapping of PET scans. *Neuroimage* **19**, 601-612.

[48] Martino ME, de Villoria JG, Lacalle-Aurioles M, Olazaran J, Cruz I, Navarro E, Garcia-Vazquez V, Carreras JL, Desco M (2013) Comparison of different methods of spatial normalization of FDG-PET brain images in the voxel-wise analysis of MCI patients and controls. *Ann Nucl Med* **27**, 600-609.

[49] Worsley KJ, Marrett S, Neelin P, Evans AC (1996) Searching scale space for activation in PET images. *Hum Brain Mapp* **4**, 74-90.

[50] Rosenfeld A, Kak AC (1982) *Digital Picture Processing*, Academic Press, New York.

[51] Arbizu J, Prieto E, Martinez-Lage P, Marti-Climent JM, Garcia-Granero M, Lamet I, Pastor P, Riverol M, Gomez-Isla MT, Penuelas I, Richter JA, Weiner MW, Alzheimer's Disease Neuroimaging I (2013) Automated analysis of FDG PET as a tool for single-subject probabilistic prediction and detection of Alzheimer's disease dementia. *Eur J Nucl Med Mol Imaging* **40**, 1394-1405.

[52] Caroli A, Prestia A, Chen K, Ayutyanont N, Landau SM, Madison CM, Haense C, Herholz K, Nobili F, Reiman EM, Jagust WJ, Frisoni GB, Eadc-Pet Consortium N-D, Alzheimer's Disease Neuroimaging Initiative (2012) Summary metrics to assess Alzheimer disease-related hypometabolic pattern with 18F-FDG PET: Head-to-head comparison. *J Nucl Med* **53**, 592-600.

[53] Herholz K, Westwood S, Haense C, Dunn G (2011) Evaluation of a calibrated (18)F-FDG PET score as a biomarker for progression in Alzheimer disease and mild cognitive impairment. *J Nucl Med* **52**, 1218-1226.

[54] Morbelli S, Brugnolo A, Bossert I, Buschiazzo A, Frisoni GB, Galluzzi S, van Berckel BN, Ossenkoppele R, Perneczky R, Drzezga A, Didic M, Guedj E, Sambuceti G, Bottoni G, Arnaldi D, Picco A, De Carli F, Pagani M, Nobili F (2014) Visual versus semi-quantitative analysis of 18F-FDG-PET in amnestic MCI: An European Alzheimer's Disease Consortium (EADC) project. *J Alzheimers Dis* **44**, 815-826.

[55] Chen K, Ayutyanont N, Langbaum JB, Fleisher AS, Reschke C, Lee W, Liu X, Bandy D, Alexander GE, Thompson PM, Shaw L, Trojanowski JQ, Jack CR Jr, Landau SM, Foster NL, Harvey DJ, Weiner MW, Koeppe RA, Jagust WJ, Reiman EM, Alzheimer's Disease Neuroimaging, Initiative (2011) Characterizing Alzheimer's disease using a hypometabolic convergence index. *Neuroimage* **56**, 52-60.

[56] Ritter K, Schumacher J, Weygandt M, Buchert R, Allefeld C, Haynes J-D, for the Alzheimer's Disease Neuroimaging Initiative (2015) Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers. *Alzheimers Dement (Amst)* **1**, 206-215.

[57] Perani D, Della Rosa PA, Cerami C, Gallivanone F, Fallanca F, Vanoli EG, Panzacchi A, Nobili F, Pappata S, Marcone A, Garibotto V, Castiglioni I, Magnani G, Cappa SF, Gianolli L, Consortium E-P (2014) Validation of an optimized SPM procedure for FDG-PET in dementia diagnosis in a clinical setting. *Neuroimage Clin* **6**, 445-454.