Validation of an optimized SPM procedure for FDG-PET in dementia diagnosis in a clinical setting

Daniela Perani[a,b,c,d,1,*], Pasquale Anthony Della Rosa[d,1], Chiara Cerami[a,e,1], Francesca Gallivanone[d], Federico Fallanca[c], Emilia Giovanna Vanoli[c], Andrea Panzacchi[c], Flavio Nobili[f], Sabina Pappatà[g], Alessandra Marcone[f], Valentina Garibotto[h], Isabella Castiglioni[d], Giuseppe Magnani[i], Stefano F. Cappa[e,j], Luigi Gianolli[c], EADC-PET Consortium[2]

[a]Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy
[b]Vita-Salute San Raffaele University, Milan, Italy
[c]Nuclear Medicine Unit, San Raffaele Hospital, Milan, Italy
[d]Istituto di Bioimmagini e Fisiologia Molecolare, CNR, Segrate, Italy
[e]Clinical Neurosciences Department, San Raffaele Hospital, Milan, Italy
[f]Dept of Neuroscience (DINOGMI), University of Genoa, Genoa, Italy
[g]Institute of Biostructure and Bioimaging, CNR, Naples, Italy
[h]Department of Medical Imaging, Geneva University, Geneva, Switzerland
[i]Neurology Department, San Raffaele Hospital, Milan, Italy
[k]Istituto Universitario degli Studi Superiori, Pavia, Italy

ARTICLE INFO

ABSTRACT

Diagnostic accuracy in FDG-PET imaging highly depends on the operating procedures. In this clinical study on dementia, we compared the diagnostic accuracy at a single-subject level of a) Clinical Scenarios, b) Standard FDG Images and c) Statistical Parametrical (SPM) Maps generated via a new optimized SPM procedure. We evaluated the added value of FDG-PET, either Standard FDG Images or SPM Maps, to Clinical Scenarios. In 88 patients with neurodegenerative diseases (Alzheimer's Disease—AD, Frontotemporal Lobar Degeneration—FTLD, Dementia with Lewy bodies—DLB and Mild Cognitive Impairment—MCI), 9 neuroimaging experts made a forced diagnostic decision on the basis of the evaluation of the three types of information. There was also the possibility of a decision of normality on the FDG-PET images. The clinical diagnosis confirmed at a long-term follow-up was used as the gold standard. SPM Maps showed higher sensitivity and specificity (96% and 84%), and better diagnostic positive (6.8) and negative (0.05) likelihood ratios compared to Clinical Scenarios and Standard FDG Images. SPM Maps increased diagnostic accuracy for differential diagnosis (AD vs. FTD; beta 1.414, p = 0.019). The AUC of the ROC curve was 0.67 for SPM Maps, 0.57 for Clinical Scenarios and 0.50 for Standard FDG Images. In the MCI group, SPM Maps showed the highest predictive prognostic value (mean LOC = 2.46), by identifying either normal brain metabolism (exclusionary role) or hypometabolic patterns typical of different neurodegenerative conditions.

1. Introduction

The correct identification of dementia subtypes represents a major challenge for clinicians, since an early differential diagnosis may be difficult on clinical grounds only. The use of biomarkers in the different neurodegenerative processes (e.g., Alzheimer's Disease (AD), Frontotemporal Lobar Degeneration (FLTD) or Dementia with Lewy bodies (DLB)) can help obtain more accurate diagnosis, also in the prodromal stages of the diseases (Anchisi et al., 2005; Galluzzi et al., 2013). Distinctive topographic and pathophysiological markers have been thus included in the new proposed research and clinical criteria for dementias (Sperling et al., 2011; Albert et al., 2011; Jack

* Corresponding author: Vita-Salute San Raffaele University, Nuclear Medicine Department, San Raffaele Hospital, Division of Neuroscience, San Raffaele Scientific Institute, Via Olgettina 60, Milan 20132, Italy. Tel: +39 02 26432224 or 26432223; fax: +39 02 26415202.
    E-mail address: perani.daniela@hsr.it (D. Perani).
  [1] These authors contributed equally to the work.
  [2] Alexander Drzezga and Robert Perneczky [Munich], Mira Didic and Eric Guedj [Marseilles], Bart N. Van Berckel and Rik Ossenkoppele [Amsterdam], Silvia Morbelli [Genoa], Giovanni Frisoni and Anna Caroli [Brescia].

et al., 2011; McKhann et al., 2011; Dubois et al., 2010; Dubois et al., 2014; Gorno-Tempini et al., 2011; Rascovsky et al., 2011; McKeith et al., 2005), comprising changes in Aβ and tau protein values in the cerebrospinal fluid, structural brain changes visible on brain Magnetic Resonance Imaging (MRI) and metabolic changes seen with 18-fluorodeoxyglucose positron emission tomography (FDG-PET).

In AD, the pathophysiological process starts decades before the clinical onset of cognitive impairments in at risk individuals (Bateman et al., 2012; Mosconi et al., 2014). Therefore, the concept of the "AD pathophysiological process" has been teased apart from that of "AD dementia" (Dubois et al., 2013)". The current research criteria for AD encourage translating the results from biomarker studies into guidelines for diagnosis (Sperling et al., 2011; Albert et al., 2011; Jack et al., 2011; McKhann et al., 2011; Dubois et al., 2010; Dubois et al., 2014). As far as imaging markers of AD are concerned, there is a widespread evidence for an increase in diagnostic accuracy for metrics-based assessment (Frisoni et al., 2013). Among the topographic imaging markers, the best MRI marker in AD, even at the prodromal stage, i.e. hippocampal volume, might be influenced by several potential confounders, such as aging (Fjell et al., 2013), concomitant pathologies affecting the medial temporal lobe (e.g., diabetes or sleep apnoea) (Fotuhi et al., 2012), and the presence of hippocampal sclerosis or other dementia types that might also show hippocampal atrophy (de Souza et al., 2013). FDG-PET instead shows high sensitivity in detecting typical and reliable patterns of brain metabolic dysfunction, also in the early AD phase (Anchisi et al., 2005; Herholz et al., 2002; Prestia et al., 2013). It has been considered to be more sensitive than MRI in the typical AD pathological cascade (Jack et al., 2010), since the pathological phenomena leading to neuronal synaptic dysfunction affect glucose consumption prior to causing cell death and detectable atrophy (Bateman et al., 2012; Chételat et al., 2008; Perani, 2014).

The typical AD pattern represented by hypometabolism in temporo-parietal regions, precuneus and posterior cingulate cortex is closely related to cognitive impairment (Perani, 2008) and allows the early identification of the downstream neuronal degeneration, in prodromal phase before the full onset of dementia (Mosconi et al., 2008; Herholz, 2010), and many years before in at risk individuals (McKeith et al., 2005; Bateman et al., 2012). Moreover, on the basis of the specific patterns of topographic distribution of the metabolic changes, FDG-PET can also help to recognize and differentiate other dementia types (Teune et al., 2010; Bohnen et al., 2012). Therefore, FDG-PET imaging has been included also in the supportive criteria for non-AD dementias, such as FTLD and DLB (Gorno-Tempini et al., 2011; Rascovsky et al., 2011; McKeith et al., 2005).

The PET imaging recommendations for dementia diagnosis are largely referring to FDG-PET literature based on subjective methods for evaluating metabolic FDG-PET changes (i.e., visual inspection). These approaches greatly depend on the observer's experience and may reduce sensitivity and specificity especially in those centres where advanced expertise in image reading is unavailable, due to the lack of an objective cut-off between normal and pathological findings. A lower diagnostic confidence may arise, especially for the earliest disease stages when only subtle metabolic abnormalities may be present. To overcome these limitations, FDG-PET images have been assessed using statistical image analysis, such as AD t-sum (Herholz et al., 2002), Stereotactic Surface Projection (SSP) statistics (Neurostat®) (Foster et al., 2007; Minoshima et al., 2001) and Statistical Parametrical Mapping (SPM) (Signorini et al., 1999; Patterson et al., 2010). Other methods have been reported and compared in accuracy (Caroli et al., 2012).

AD t-sum score provides a measure of scan abnormality associated with a preset threshold for discrimination between AD patients and controls. This method, however, based on the sum of t-values in predefined regions typically affected by AD, does not allow differential diagnosis in dementia conditions and the identification of whole-brain patterns of hypometabolism (Herholz et al., 2002; Prestia et al., 2013) at difference with SPM procedures (Signorini et al., 1999; Patterson et al., 2010).

Neurostat® (Minoshima et al., 2001) is an atlas-based method allowing a whole-brain parametric analysis of FDG-uptake, yielding z-scores as a result of the voxel by voxel comparison between an individual subject and a predefined control group. The Neurostat® (Minoshima et al., 2001) package uses surface projection technology to generate and display an entire statistical map on a projection of brain surfaces in different views. Neurostat® (Minoshima et al., 2001) uses four different intensity normalization methods, including global intensity normalization, as well as normalization to the thalamus, cerebellum and pons. A limit is that the individual FDG-PET scan is normalized to the Talairach and Tournoux atlas space and 3D rendering of statistical map of glucose hypometabolism are visualized through a cortical surface projection technology. In addition, despite the possibility of statistically inferring upon the pattern of FDG metabolism through a colour bar, p-values for single voxels or clusters of voxels are not available and this procedure does not allow the detection of metabolic changes in deep brain structures (i.e., basal ganglia or midbrain).

SPM analysis of FDG-PET images has been also used in the evaluation of brain metabolic changes in neurodegenerative conditions showing specific topographic patterns associated with cognitive decline and dementia (Anchisi et al., 2005; Signorini et al., 1999; Patterson et al., 2010; Yakushev et al., 2009). In brief, each individual FDG-PET scan is first warped in the standard MNI space using a template image for spatial normalization and subsequently smoothed with a 3D Gaussian kernel. Parametric analysis of FDG-uptake in SPM is obtained using voxel-level statistical parametric mapping at the whole-brain level, in the framework of the general linear model by means of a two-sample t-test, comparing each subject against images pertaining to a reference control group. The comparison between each individual and the reference group yields a contrast t-map testing for areas with relative decreases in metabolism (i.e., hypometabolism) compared to the controls. Significance values from the voxel-wise t-test are finally reported.

One of the main advantages of the SPM approach resides in the use of metrics evaluated at a predefined significance statistical threshold to define the topography of hypometabolism, thus drastically reducing the chance for false positives and increasing specificity (Silverman et al., 2003). The presence of false positives is particularly detrimental at the single-subject level, when it is crucial to exclude neurodegenerative conditions and to effectively discriminate between the different dementia conditions. These patterns can also be easily read even by intermediate-skilled readers.

The aim of this study was to assess and validate at a single-subject level the sensitivity and specificity of a new SPM procedure compared to subjective visual inspection of FDG-uptake distribution maps and to the sole clinical information. Voxel-based SPM hypometabolism maps were produced by means of spatial normalization to a new standardized FDG-PET template that showed to increase reliability and accuracy of estimated brain metabolic patterns (Della Rosa et al., 2014). Statistical analysis was based on a large normal dataset (112 control scans) for single-subject comparisons. A group of neuroimaging experts was individually asked to provide a forced diagnosis and indicate their level of confidence on the basis of a) summary of clinical data, b) visual inspection of Standard FDG Images, and c) assessment of SPM t-maps, in a large series of patients with clinical diagnosis of neurodegenerative disorders (i.e., AD, FTLD and DLB), and in subjects with Mild Cognitive Impairment—MCI.

2. Material and methods

2.1. Subjects

A series of 88 subjects was retrospectively pooled from the population database of the Neurology Centres for Cognitive Disorders in the San Raffaele Hospital (Milan, Italy). Medical history and neurological examination were provided by neurologists and neuropsychological assessment by expert neuropsychologists within the above centres.

The clinical information (medical history, neurological examination and neuropsychological assessment) was evaluated by three neurologist experts in dementia diagnosis (AM, SFC, GM) that examined the whole medical records related to the entire clinical course of each subject in order to correctly assign the clinical diagnosis at the follow-up, which was used as gold-standard. Only subjects with the clinical follow-up and complete diagnostic agreement by the three expert neurologists (AM, SFC, GM) were included.

According to the diagnosis at the clinical follow-up (27.6 ± 4.1 months; range $22-35$ months) (gold-standard), we identified patients fulfilling consensus criteria for AD (Jack et al., 2011; McKhann et al., 2011; Dubois et al., 2010; Dubois et al., 2014) (n = 27; men 16, female 11; mean age = 67.40 ± 6.68; MMSE < 22; CDR global score range $1-$ 2), DLB (McKeith et al., 2005) (n = 9; men 3, female 6; mean age = 74.11 ± 6.90; MMSE < 23; CDR global score range $1-2$) and for the FTLD spectrum (n = 24; men 11, female 13; mean age = 68 ± 6). FTLD patients included 10 behavioural variant of frontotemporal dementia (bvFTD) (Rascovsky et al., 2011) (MMSE < 28; CDR global score range $0.5-1$), 5 primary progressive aphasias (PPA) (Gorno-Tempini et al., 2011) (MMSE range $23-29$; CDR global score range $0.5 -1$), and 9 corticobasal degeneration syndrome (CBS) (Armstrong et al., 2013) (MMSE < 21; CDR global score range $1-2$). In addition, 28 subjects were classified at the baseline as MCI (Petersen et al., 2009) (amnestic single-domain, non-amnestic single domain and multidomain; men 16, female 12; mean age = 71.32 ± 5.67; MMSE range $25-28$; CDR global score 0.5). Among these, 7 subjects progressed to dementia at follow-up (27.6 ± 4.1 months); while 6 reverted to cognitive normal condition and 15 remained stable.

We selected 93 cognitively normal subjects by the European Alzheimer Disease Consortium (EADC)-PET dataset (http://www.eadc.info/) and 19 cognitively normal subjects that were previously acquired in the Nuclear Medicine Dept of the San Raffaele Hospital. All of them were included in the SPM procedure for statistical comparisons. Cognitive health was established in each PET centre by means of a structured clinical and a neuropsychological battery and subjects were followed up for more than a year, as specified in a previous paper (Morbelli et al., 2012).

The Institutional Ethical Committee at the University-Hospital San Raffaele of Milan approved this study.

2.2. FDG-PET image acquisition procedure

FDG-PET acquisitions of the whole patient group and of 112 cognitively normal subjects were performed according to the guidelines of the European Association of Nuclear Medicine (EANM) (Morbelli et al., 2012; Varrone et al., 2009).

All FDG PET images of the patients were acquired at the Nuclear Medicine Dept., San Raffaele Hospital (Milan, Italy), with a Discovery STE (GE Medical Systems, Milwaukee, WI) multi-ring PET tomography (PET-CT) system (time interval between injection and scan start = 45 min; scan duration = 15 min).

2.3. SPM-FDG-PET image pre-processing

In order to obtain voxel-based statistical parametric hypometabolic maps, each FDG-PET brain image scan was pre-processed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm/software/SPM5/), running in Matlab 7.6 (MathWorks Inc., Sherborn, MA).

Functional normalization procedure allowed to place the images in the standard MNI space using a new FDG-PET dementia-specific template for spatial normalization, based on images derived from both neurological patients (sample of scans representative of the various forms of dementia in the population) and age-matched controls, developed by Della Rosa et al. (2014), available to download in the "Templates" section (on the SPM official website http://www.fil.ion.ucl.ac.uk/spm/). We created this new [^{18}F]-FDG PET population-specific template for

spatial normalization, since the standard spatial normalization requires a non-linear registration of PET images to a [^{15}O]-H$_2$O template provided with the SPM software. FDG uptake (i.e., the pattern of signal intensities in the [^{18}F]-FDG PET image) may not be represented properly by the pattern of signal intensities in the standard [^{15}O]-H$_2$O PET template which is based on blood flow.

This Dementia-Specific FDG-PET template (Della Rosa et al., 2014) was built by averaging and smoothing (with an 8-mm FWHM Gaussian filter) a number of 100 intensity and spatially normalized FDG-PET images (50 controls and 50 patients). This template was used to spatially transform FDG-PET patient images to the MNI (Montreal Neurological Institute) reference space, as it has been shown to provide a higher degree of accuracy for spatial normalization of FDG-PET scans and a higher statistical sensitivity at the single-subject level in SPM5, useful for clinical purposes (Della Rosa et al., 2014).

Prior to normalization, we first performed approximate manual image re-orientation and positioning to MNI space of each subject FDG-PET image. To spatially normalize the FDG-PET patient images to the Dementia-Specific FDG-PET template, we used the normalization algorithm provided by SPM5 with the following parameter settings including 12-parameters' affine transformation, $7 \times 8 \times 7$ discrete cosine transform basis functions, no template and source weighting; discrete cosine transform cut-off: 25 mm; 16 non-linear iterations and the non-linear regularization term set to 1. No modulation ("preserve concentrations") and trilinear interpolation were used during final reslicing.

Normalized images were written in the default SPM5 bounding box with an isotropic voxel size of 2 mm. Visual inspections of normalized images allowed to ensure registration quality and convergence of the normalization procedure. Spatially normalized images of all subjects were subsequently smoothed with an isotropic 3D Gaussian kernel of 8 mm FWHM before entering statistical analysis.

2.4. SPM-FDG-PET image single-subject analysis

A single-subject analysis for comparison to a cognitively normal control group (n = 112; age range 50–80 years) was performed using SPM5 for all the included subjects.

In the application of SPM approach, usually the number of images included as control reference is very small ($\leq$20) (Signorini et al., 1999), and limited by ethical constraints dealing with the PET acquisition of normal controls. It has been shown however, that increasing the sample size of control database for comparison, improves the diagnostic performances (Chen et al., 2008).

The control database used for comparison in our new approach amounts to 112 images, which allows to use higher and more conservative significance thresholds at both the voxel (p = 0.05 FWE-corrected) and cluster level (extent > 100 voxels) and, even more importantly, to correct for multiple comparisons (i.e., FWE: family-wise error correction). This reduces the false positive rate and allows for more robust inference on FDG patterns of hypometabolism when assessing neurodegenerative diseases.

Furthermore, each FDG-PET scan of the healthy control group (n = 112) underwent an intensity rescaling and global count intensity normalization (Della Rosa et al., 2014) to have the same mean intensities (Friston et al., 1994; Buchert et al., 2005) in order to standardize the magnitude of all voxel values in every image, thus accounting for potential sources of variability both between scanners (i.e., centre-specific image scaling) or between individuals (i.e., patient weight, cardiac output, the amount of injected radioactivity). Then, each FDG-PET scan of the control group was normalized to the Dementia-Specific FDG-PET template and tested for normality in a jack-knife approach, where every normalized FDG PET scan was evaluated with respect to the remaining sample via the two sample t-test in SPM5 (Della Rosa et al., 2014).

Each FDG-PET patient image scan was tested for relative 'hypometabolism' by comparison with the reference group of 112 controls on a voxel-by-voxel basis using the general linear model, by means of the two sample t-test design of SPM5. Age was included as a covariate. Due to the lack of any significant difference in metabolic activity of male and female demented patients (Minoshima et al., 1997), gender was not controlled in the analysis. The measurements were assumed to be independent and have unequal variance between levels. Global normalization of voxel values used proportional scaling to a mean voxel value of 6.5 mg/100 mL/min (see Signorini et al., 1999 for details) to minimize inter-subject variability. Proportional scaling basically scales each image according to a reference count, which is the global brain activity to a physiologically realistic reference value of 6.5 mg/100 mL/min.

The threshold was left at the default 0.8 value (i.e., the mean brain intensity was computed from only those voxels with intensity above 0.8 of the mean over the entire scan). Voxel-wise comparisons were made using an explicit FDG-PET mask (Ridgway et al., 2009). This mask was created using the SPM masking toolbox to produce an average binary mask, where the voxels from which to determine the FDG-metabolism parameter estimates were restricted to an explicit mask. The latter resulted from optimal thresholding of voxels in each image based on their correlation with an average image (i.e., the average of FDG-PET images from the 112 healthy control scans) with voxels not meeting the optimality criterion set to zero (masked). This mask was applied (i.e., explicit masking option in SPM5-GLM models) to restrict subsequent single-subject statistical analyses only to within-brain voxels in order to eliminate variance due to inter-subject variation and noise from outside the brain (Spence et al., 2006).

The SPM comparison between each single FDG-PET scan and the healthy control group of scans essentially provides regional differences in relative glucose metabolism by means of a t-statistic for each voxel (SPM-t maps). Clusters of decreased metabolism were considered significant when they met a significance level of p = 0.05, corrected for multiple comparisons with the family-wise error (FWE) option at the voxel level, and contained more than 100 voxels.

2.5. Rating

Nine neuroimaging experts with an extensive experience in the dementia diagnosis (DP, FF, EGV, AP, FN, SP, CC, VG, LG) were presented with clinical and neuropsychological information ("Clinical Scenarios") (Fig. 1A), standard clinical display of FDG-PET images ("Standard FDG Images") (Fig. 1B), and voxel-based statistical parametric hypometabolic maps ("SPM Maps") (see Fig. 1C). Raters were all experts in the field of dementia research and diagnosis. They had a long lasting experience in reporting of FDG-PET scans (i.e., visual assessment of Standard FDG Images and SPM Maps) together with the clinical information. They were informed that the study included patients who had at follow-up a clinically confirmed diagnosis of AD, FTLD spectrum (including different subtypes), DLB, as well as MCI subjects.

2.6. Rating materials

2.6.1. Clinical and neuropsychological information (Clinical Scenarios)

An anonymized score sheet grid was built for each patient at baseline (near and before the FDG-PET scan). It summarized general clinical information (onset, disease duration, age, education, comorbidity), neurological signs, cognitive symptoms (impairments of memory, executive functions, language, visuo-spatial, orientation, praxia), behavioural changes (positive and negative symptoms, as well as sleep disorders), global cognitive efficiency (i.e., Mini Mental State Examination (MMSE) score), and functionality (i.e., Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) scores) (Fig. 1A). Results of a detailed neuropsychological battery were also provided for each subject

on a separate sheet. In particular, the administered tests were Rey Auditory Verbal Learning Test, Rey's figure recall test, verbal and visual digit span tasks, attentive matrices, phonological and semantic fluency, Token test, Aachener Aphasie Test (AAT) or "Batteria per l'analisi dei deficit afasici" (BADA) subtests, Rey's figure copy and Raven's progressive matrices scores.

2.6.2. Standard display of FDG-uptake (Standard FDG Images)

A standard anonymous display of FDG-PET distribution was generated for each subject. Each image was reoriented to the AC–PC line and then displayed in a 5×5 matrix of transaxial images (i.e., 25 axial slices) from rostral to ventral brain sections in a radiological convention (i.e., left = right; right = left). Each axial slice covered 4 mm on the z-axis. Images were shown as relative metabolic rates with the highest and lowest pixel values in the scan placed at the highest value (100% red–pink) or lowest value (0% blue–purple) according to a rainbow colour scale.

2.6.3. Voxel-based SPM hypometabolic maps (SPM Maps)

We provided raters with SPM-t coloured anonymous displays of FDG-PET hypometabolic overlaid on a canonical MRI T1-weighted structural brain scan in the MNI space. Each SPM map included the same matrix (i.e., 5×5 matrix of transaxial images) used for the visual inspection of the Standard FDG Images. SPM Maps were displayed in a neurological convention (i.e., left = left; right = right). Each axial slice covered 4 mm on the z-axis ranging from −40 to +56. Images showed statistically significant reductions of metabolism with the significant t-values in yellow/red scale.

2.7. Rating procedure

2.7.1. Step 1

Each rater independently evaluated one type of information for each subject: a) Clinical Scenarios; or b) Standard FDG Images; or c) SPM Maps. Thus, the raters were presented with the Standard FDG Images or SPM Maps independently for any given subject and in both situations were blinded to any clinical data. Raters were asked to make a forced diagnosis of AD, FTLD spectrum, or DLB, and also of negative PET scan, and to indicate the degree of diagnostic confidence on a 3-point scale (1—poor confidence, 2—medium confidence or 3—high confidence). The diagnostic labels assigned by each rater were compared to the clinical diagnosis at follow-up (the gold standard).

All raters were then informed that all patients included in this study had a follow-up clinically confirmed diagnosis of AD or DLB, FTLD subtypes, stable MCI or MCI conversion or reversion, but they did not know the proportion of subjects with each diagnosis.

Based solely upon the Standard FDG Images or the SPM Maps, raters were first asked to grade the level of overall scan abnormality as normal, uncertain or abnormal. Second, raters were instructed to report more extensive details on brain metabolism, encompassing the involved brain lobes, (i.e., frontal, temporal, parietal, occipital) and they had to decide whether they were hypometabolic or not and /or if the pattern of hypometabolism was bilateral or distributed in asymmetrical way (i.e., >left hemisphere or >right hemisphere). Third, raters were asked to focus on more specific brain areas critical to the differential diagnosis of specific neurodegenerative diseases. Hypometabolism in specific brain regions was rated, namely in the anterior cingulate cortex (ACC), medial frontal cortex, orbito-frontal cortex, frontal operculum; in the inferior and superior parietal lobules, parietal operculum, precuneus, posterior cingulate cortex; in the temporal lateral cortex, temporo-medial cortex and temporal pole and; in the medial and lateral occipital cortices. Subcortical structures, namely the thalamus and basal ganglia, as well as the midbrain were also included in the list of specific areas to be rated. As in the former case, raters had to decide whether these areas were hypometabolic or not and/or if the pattern of hypometabolism was bilateral or asymmetrical.
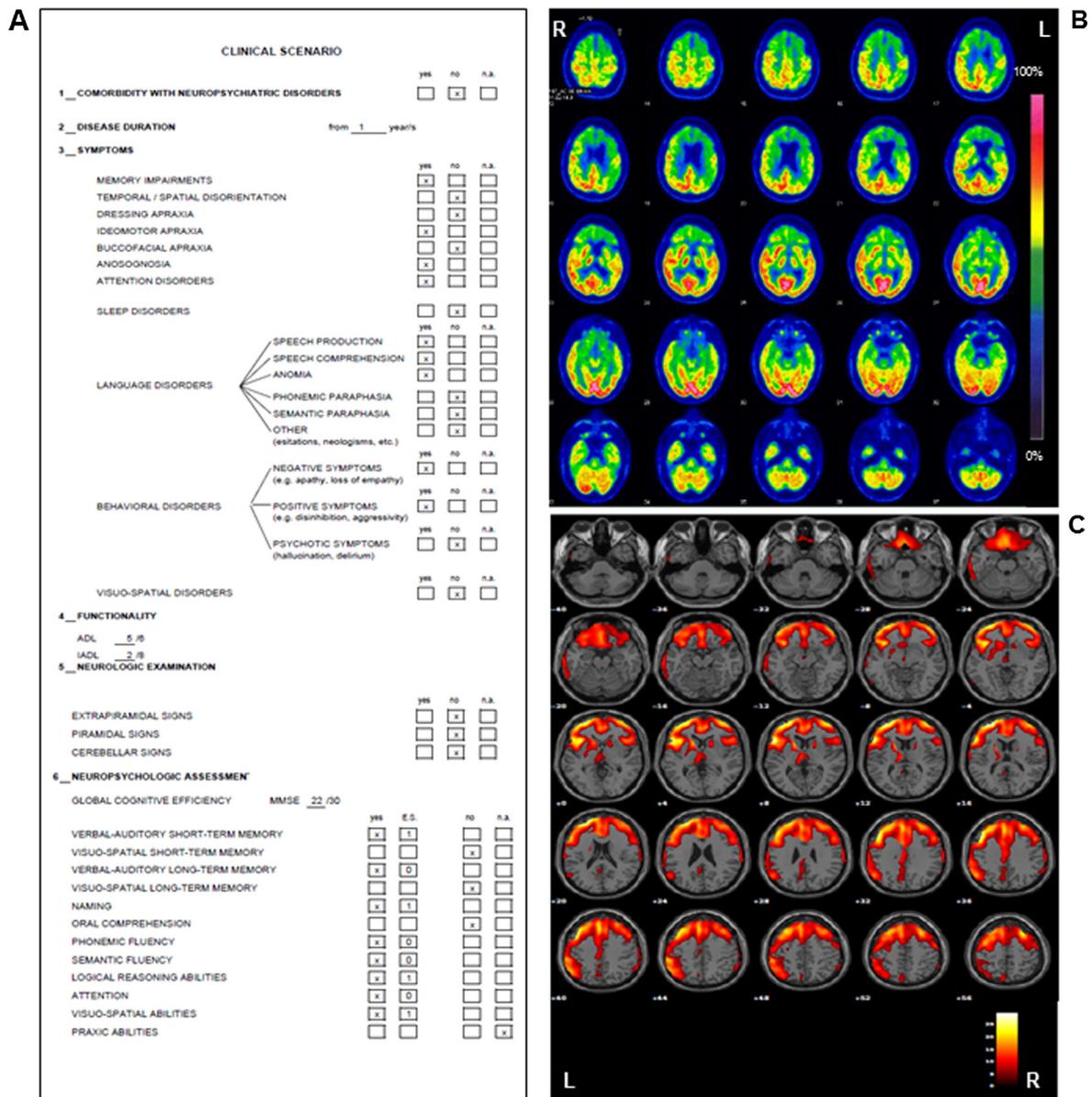
Fig. 1. Example of the material provided for each subject: A) a Clinical Scenario; B) a Standard FDG Image and C) an SPM Map of a patient affected by behavioural variant of frontotemporal dementia. L = Left; R = right.

2.7.2. Step 2

The procedure for evaluation of the Clinical Scenarios, the Standard FDG Images and the SPM Maps was identical as those outlined above for step 1. The only difference was that each rater saw for each subject the Clinical Scenarios first and initially entered their diagnosis and degree of diagnostic certainty based exclusively upon this information. Second, the rater could be presented with the Standard FDG Images or SPM Maps along with the Clinical Scenarios and was asked to evaluate the combination of both types of information, either Clinical Scenarios plus Standard Images or Clinical Scenarios plus SPM Maps, asking them for assessments of overall degree of scan normality, the presence of hypometabolism in more extensive or specific regions and of a potential asymmetrical pattern. Finally, they were asked to make a forced diagnosis among the predefined disease categories added with their degree of confidence independently from the one indicated solely on the basis of the clinical information, thus maybe resulting in a potentially different diagnostic choice. The aim of step 2 was to assess the differential value of adding qualitative (Standard FDG Images) or quantitative (SPM Maps) information to Clinical Scenarios as a support for both diagnostic accuracy and confidence.

Clinical Scenarios, Standard FDG Images and SPM Maps for steps 1 and 2 were sent independently to raters on separate dates (i.e., a 6 month time gap) and had different alphanumeric codes. In this way, raters could not associate or compare information across subjects and between the two steps.

2.8. Study design

The three types of information (i.e., Clinical Scenarios, Standard FDG Images, and SPM Maps) were associated to each subject in a factorial manner and arranged according to six possible permutations (i.e., Clinical Scenarios—Standard FDG Images—SPM Maps, Clinical Scenarios—SPM Maps—Standard FDG Images, Standard FDG Images—Clinical Scenarios—SPM Maps, Standard FDG Images—SPM Maps—Clinical Scenarios, SPM Maps—Clinical Scenarios—Standard FDG Images, and SPM Maps—Standard FDG Images—Clinical Scenarios). For

each diagnostic group, a comparable number of subjects was assigned to each permutation (no. of subjects/no. of permutations), in order for each permutation to be equally represented among considered dementia conditions. The sequence according to which each permutation was assigned to each successive subject in each diagnostic group was randomized. This design was implemented in order to have the same amount of subjects rated for each of the three types of information, and to test agreement between different pairs of raters on the same information and between the same pair of raters on different types of information.

Overall rater consensus was also evaluated independently for each type of information by computing a contingency coefficient measuring the association between the follow-up diagnostic category and the rated disease category as indicated in the study.

A set of twelve subjects was used as a consistency sample in order to assess reliability of the rating procedure.

2.9. Statistical analysis

2.9.1. Step 1

The diagnostic labels assigned by each rater were evaluated according to the clinical diagnosis at follow-up (gold standard).

In order to assess the diagnostic power of each type of information, sensitivity and specificity were computed independently for Clinical Scenarios, Standard FDG Images and SPM Maps information and expressed as percentages, averaged across the nine raters and weighted by the number of subjects evaluated by each rater for each type of information.

Moreover, for each type of information, we calculated both positive and negative likelihood ratios (LR+ and LR−), which provide the ratio between the probability of positive or negative test outcome in patients and the probability of positive or negative test outcome in true negatives. LR+ ≥ 5 and LR− ≤ 0.2 are agreed upon as being diagnostically useful (Frisoni et al., 2013).

The inter-rater agreement was calculated by rater for all possible rater pairs evaluating the same subjects for each type of information. An average phi-value between rater pairs and a chi-transform was computed independently for Clinical Scenarios, Standard FDG Images and SPM Maps and tested for significance. Inter-rater reliability by subjects was assessed using k-statistics calculated on the basis of the agreement between the two raters judging each subject, independently from the designated pair among all the possible rater pairs. The level of agreement based on the k-statistics was classified as fair (k = 0.20–0.39), moderate (k = 0.40–0.59), substantial (k = 0.60–0.79), and almost perfect (k = 0.80–1.00) (Foster et al., 2007).

Diagnostic accuracy was also assessed for Clinical Scenarios, Standard FDG Images and SPM Maps taking in consideration the percentage of rated false negatives (FN%) for the two major dementia categories here represented (i.e., AD and FTLD), and by means of a concordance index using an unanimity rule for the rater pair (number of correctly classified subjects when the rater pair is concordant). Three independent logistical models were fit to a categorical variable representing an AD or FTLD diagnosis at follow-up, using as predictor variable only the correctly and concordantly classified AD or FTLD subjects by the rater pair through Clinical Scenarios, Standard FDG Images or SPM Maps in order to estimate a concordance index for each type of information on the basis of diagnostic accuracy. Predicted group classification probabilities were then used for ROC curves corresponding to Clinical Scenarios, Standard FDG Images and SPM Maps in order to assess the average performance of the three types of information. The area under the curve (AUC) was computed in order to test the accuracy of each information type.

2.9.2. Step 2

Clinical information combined with the Standard FDG Images (Clinical Scenarios plus Standard FDG Image) or SPM Maps (Clinical Scenarios plus SPM Maps) was evaluated for each subject. The randomized procedure for evaluation of the three types of information was identical for both steps 1 and 2.

In order to measure the level of agreement between the raters for Clinical Scenarios plus Standard FDG Images or Clinical Scenarios plus SPM Maps on consistency subjects, we computed a generalized kappa coefficient and used the z-statistic to test the null hypothesis, given the equal number of ratings. We then compared in an ANOVA model the level of confidence (LOC) in making a diagnostic choice only for correctly classified subjects in each disease category. In addition, considering only AD and FTLD groups, i.e. the two major dementia categories here represented, we assessed the strength of Standard FDG Images or SPM Maps in terms of correctly classifying those AD or FTLD subjects, which were misclassified based solely upon Clinical Scenarios.

For both combinations (i.e., Clinical Scenarios plus Standard FDG Images or Clinical Scenarios plus SPM Maps), raters were also asked to specify the hypometabolic brain regions as well as the whole-brain distribution of the pattern (i.e., bilateral or asymmetrical).

In order to quantify the contribution of the identification of selected hypometabolic patterns through Standard FDG Images or SPM Maps to the correct classification of AD and FTLD spectrum, we computed an index, expressed as the differential % of incidence (DI) of localized hypometabolism in disease-specific areas between Standard FDG Images and SPM Maps for either AD or FTLD. Only the AD and FTLD subjects that were correctly and concordantly classified by each rater pair were taken in consideration. The % difference between the number of times an area was rated as hypometabolic through Standard FDG Images and SPM Maps was calculated for each specific area weighted by the total number of correct classifications for either AD or FTLD, independently for Standard FDG Images and SPM Maps. Only areas with an index above 10% were considered.

Finally, LOC for correctly evaluating MCI was also compared between Standard FDG Images (Clinical Scenarios plus Standard FDG Images) or SPM Maps (Clinical Scenarios plus SPM Maps).

3. Results

3.1. Step 1

Raters' performance for diagnostic accuracy and confidence levels is illustrated in Figs. 2 and 3 (see also the figure on the supplementary materials). Analysis of raters' performances for diagnostic accuracy indicated higher sensitivity and specificity values for SPM Maps (96% and 84%), compared to Clinical Scenarios (91% and 40%) and Standard FDG Images (78% and 50%), considering the total sample of 88 subjects. Diagnostic LR+ was better for SPM Maps (6.08) than for Standard FDG Images (1.55) or Clinical Scenarios (1.52), exceeding the LR+ accuracy cut-off (≥5) only for SPM Maps. LR−values were best for SPM Maps (0.05) and very poor for Standard FDG Images (0.45) and Clinical Scenarios (0.22), with only SPM Maps fell below LR-cut off (≤0.2) (see Fig. 4A and 4B).

The chi-transforms of average phi-values measuring inter-rater agreement between each rater pair were higher for SPM Maps (χ^2 = 173.53, p < 0.001) than for Clinical Scenarios (χ^2 = 126.04, p = 0.01) and lowest for Standard FDG Images (χ^2 = 79.29, p = 0.9).

The k-statistic measuring inter-rater reliability by subjects independently of the rater pair was substantial for SPM Maps (k = 0.6), moderate for Clinical Scenarios (k = 0.4) and fair for Standard FDG Images (k = 0.3).

The contingency coefficient (C) measuring general association between diagnosis at follow-up and rated category for all disease categories and overall rater consensus was higher for SPM Maps (C = 0.722, p < 0.0001) than for Clinical Scenarios (C = 0.673, p < 0.0001) or Standard FDG Images (C = 0.678, p < 0.0001).

When considering only AD and FTLD, the average false-negative rate for SPM Maps was 0%, for Clinical Scenarios 2% and for Standard FDG
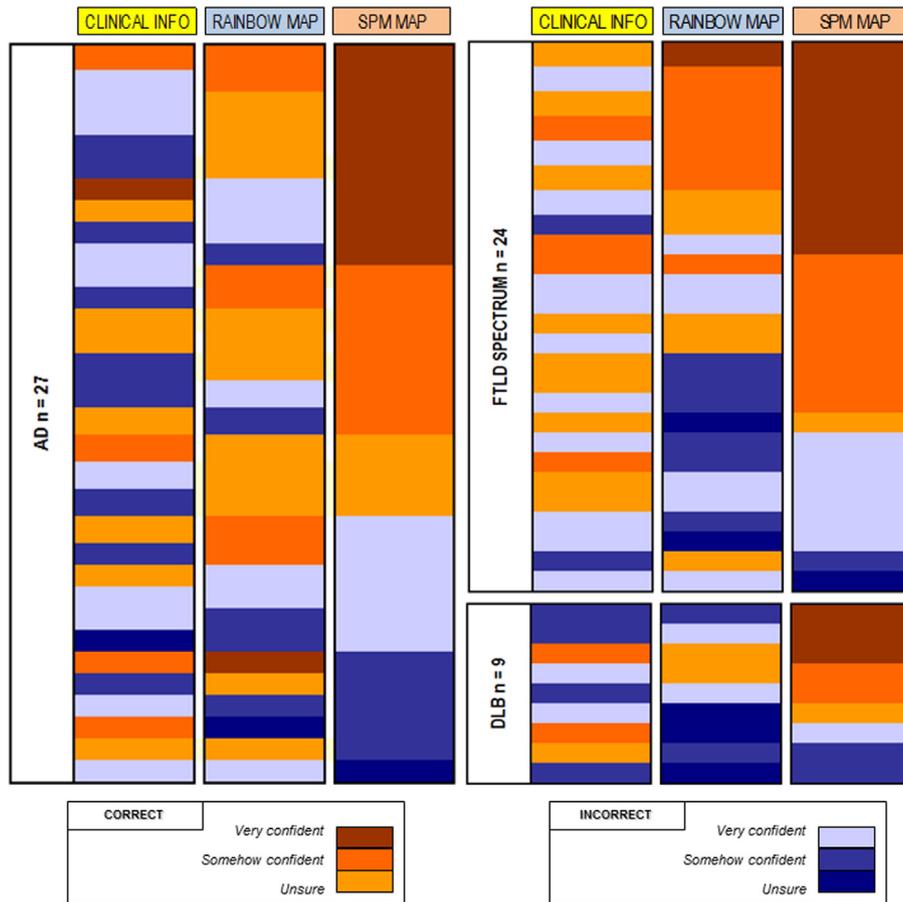
Fig. 2. Mean raters' performance for diagnostic accuracy and confidence level in dementia patients. Each horizontal bar represents the ratings in a single subject for each of the three information (Clinical Scenarios, Standard FDG Images and SPM Maps). Clinical diagnoses judged to be correct are shown in shades of red (red = very confident, orange = somehow confident, pale orange = unsure). Incorrect diagnoses are shown in shades of blue (dark blue = very confident, azure = somehow confident, sky blue = unsure).
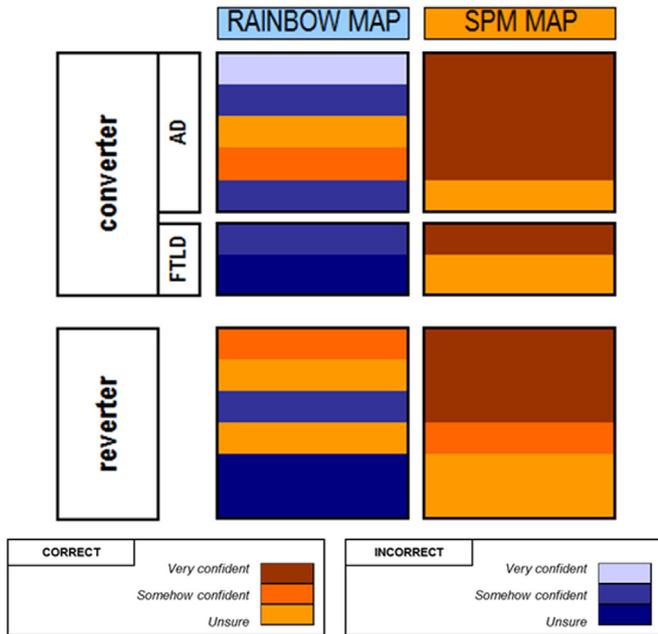
Fig. 3. Mean raters' performance for diagnostic accuracy and confidence level in MCI subjects who converted to dementia or reverted to normal condition. Each horizontal bar represents the ratings in a single subject for Standard FDG Image (left column) and SPM Map (right column) information. Clinical diagnoses judged to be correct are shown in shades of red (red = very confident, orange = somehow confident, pale orange = unsure). Incorrect diagnoses are shown in shades of blue (dark blue = very confident, azure = somehow confident, sky blue = unsure).

Images 16%. Logistic regressions tested whether SPM Maps is a more robust metric than Clinical Scenarios or Standard FDG Images in terms of both concordance and accuracy for the classification of FTLD with respect to AD. It revealed that the SPM Maps model (beta = 1.414; p = 0.019) correctly classified a significantly higher number of concordant subjects in the FTLD group, compared to the Clinical Scenarios (beta = 0.671; p = 0.291) or Standard FDG Images (beta = −0.041; p = 0.945), considering AD classification as the reference. ROC curves measuring of the goodness-of-fit of the SPM Maps, Clinical Scenarios and Standard FDG Images models indicated that the AUC was 0.67 for SPM Maps, 0.57 for Clinical Scenarios and 0.50 for Standard FDG Images. The asymptotic significance was inferior to 0.05 only for the SPM Maps model (p = 0.039; Clinical Scenarios: p = 0.416; Standard FDG Images: p = 0.955). This means that the logistic regression using the SPM Maps information classified correctly and concordantly the FTLD group significantly better than chance (see Fig. 5).

3.2. Step 2

The overall inter-rater diagnostic agreement based on a generalized kappa coefficient for consistently rated subjects was substantial for Clinical Scenarios plus SPM Maps combination (k = 0.6; z = 10.57, p < 0.0001) and only moderate for Clinical Scenarios plus Standard FDG Images (k = 0.4; z = 5.99, p < 0.0001). Viewing the Clinical Scenarios plus SPM Maps combination (mean LOC = 2.4) significantly increased the diagnostic confidence as compared to Clinical Scenarios plus Standard FDG Images (mean LOC = 2.07) (p = 0.003). In the case of the FTLD spectrum only a trend was present in the same
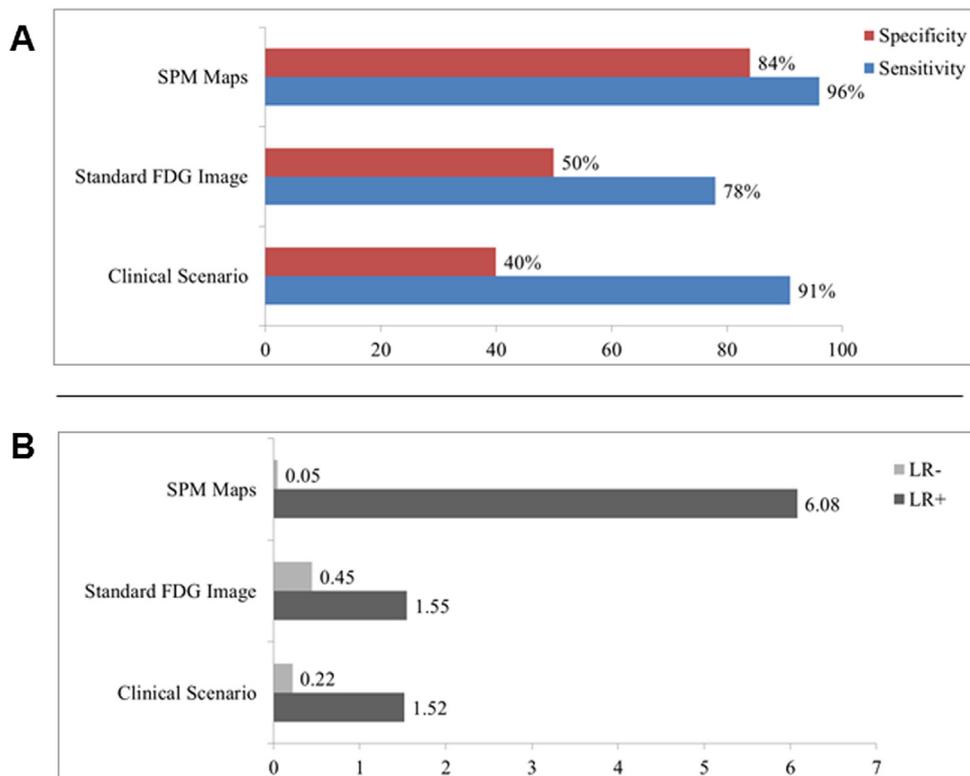
Fig. 4. A) Sensitivity (blue) and specificity (red) values of Clinical Scenarios, Standard FDG Image and SPM Maps. B) Positive (LR+) (dark grey) and negative (LR−) likelihood (pale grey) ratio for correct classification of patients, broken down by type of information.

direction (Clinical Scenarios plus SPM Maps mean LOC = 2.26; Clinical Scenarios plus Standard FDG Images mean LOC = 1.97, p = 0.078).

In the subjects with a correct diagnosis of AD, raters found hypometabolism in the temporo-parietal (DI = +26%) and posterior cingulate regions (DI = +38%) more frequently with SPM Maps than Standard FDG Images. In the subjects with correct diagnosis of FTLD, the anterior cingulate cortex (DI = +16%), the medial frontal cortex (DI = +12%) and the superior anterior temporal cortex (DI = +18%) in the left hemisphere were identified as hypometabolic much more frequently with SPM Maps than with Standard FDG Images.

3.3. MCI evaluation

Within the MCI group, the 7 subjects who progressed to dementia (i.e., 5 AD and 2 FTLD) were all correctly classified by SPM Maps at baseline. Only 2 MCI subjects who progressed to AD were correctly classified by the Standard FDG Images. All MCI subjects classified as AD at follow-up showed the typical hypometabolic pattern suggestive for AD (Anchisi et al., 2005; de Souza et al., 2013; Herholz et al., 2002; Teune et al., 2010) and characterized by bilateral temporo-parietal hypometabolism, and also involving precuneus/posterior cingulate
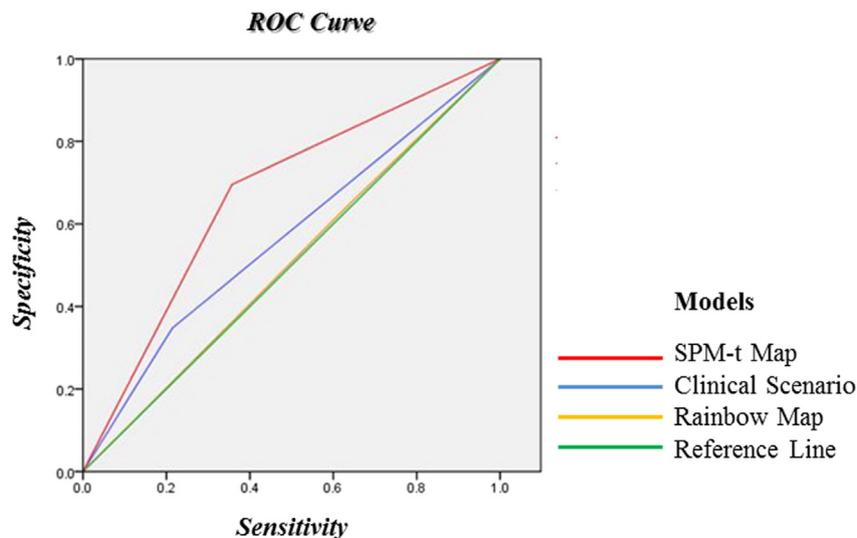

Fig. 5. ROC curve for Clinical Scenarios (area under the ROC curve (AUC) = 0.57), Standard FDG Images (AUC = 0.50) and SPM Maps (AUC = 0.67) models, showing the better correct classification through SPM Maps.

cortex. One subject classified at baseline as executive non-amnestic MCI presented frontal and anterior temporal hypometabolism with prevalent involvement of ventro-medial prefrontal cortex, a pattern suggestive of bvFTD (Rascovsky et al., 2011; Teune et al., 2010; Salmon et al., 2003). One additional subject, clinically classified as amnestic single-domain MCI at baseline showed predominant anterior temporal polar hypometabolism, typically described in association with the semantic variant of primary progressive aphasia (svPPA) (Gorno-Tempini et al., 2011). The latter two subjects both fulfilled criteria for respectively bvFTD and svPPA at follow-up (Gorno-Tempini et al., 2011; Rascovsky et al., 2011). See also Fig. 3 for details.

SPM analysis identified also specific hypometabolic patterns suggestive for different neurodegenerative conditions (i.e., AD, DLB, FTLD) in 15 MCI subjects who did not convert within the time frame of the present clinical study. A longer follow-up would be necessary in these subjects to obtain values in the prediction of disease progression.

The SPM Maps of six MCI individuals (3 amnestic single domain, 2 amnestic multi-domain and 1 visuo-spatial non-amnestic) showed normal brain metabolism at baseline. On the basis of this information they were all classified as negative by each rater. At follow-up, all these subjects reverted to normal cognition.

Compared to Clinical Scenarios plus Standard FDG Images, the effect of adding SPM Maps to Clinical Scenarios on ratings was significantly more beneficial for correctly classifying both positive MCI who progressed and the negative MCI category (Clinical Scenarios plus SPM Maps mean LOC = 2.46; Clinical Scenarios plus Standard FDG Images mean LOC = 1.81, p < 0.001).

4. Discussion

The main result of this single-subject study is that the SPM-based tool for the analysis of FDG-PET imaging improved diagnostic accuracy in dementia and pre-dementia conditions compared to visual inspection of FDG-uptake distribution, providing also additional value to clinical information. Our data add solid evidence to the importance of metrics in the clinical setting (Frisoni et al., 2013; Foster et al., 2007; Perani et al., 2014).

In summary, analysis of performances showed very high sensitivity and specificity for SPM Maps (96% and 84%) as compared to Clinical Scenarios (91% and 40%) and Standard FDG Images (78% and 50%) (Fig. 4A). Independent measures of diagnostic accuracy (LR+ and LR−) were better for SPM Maps than Standard FDG Images with only SPM Maps exceeding the proposed cut-off value for dementia diagnosis (Fig. 4B).

Noteworthy, SPM Maps showed higher inter-rater agreement compared to the other types of information. All this results in a substantial reliability of SPM analysis for the diagnostic classification, confirming that the assessment of SPM disease-specific hypometabolic patterns is less influenced by the specific reader expertise than visual inspection.

In the case of the differential diagnosis between AD and FTLD, there is evidence of a high percentage of false-negative with visual inspection (16%). On the contrary, SPM analysis yielded no false negatives. ROC curves showed that SPM information is a very robust metric to correctly and concordantly differentiate among diagnostic groups (see Fig. 5). In particular, the ROC area under the curve (AUC) proves that SPM-t maps classified FTLD subtypes better than visual inspection, underlining the importance of voxel-based analysis for differential diagnosis, especially in the case of dementia subtypes with more heterogeneous genetic and pathological signatures, such as those within the FTLD spectrum.

Moreover, SPM single subject information provides more diagnostic strength to clinical information than Standard FDG Images. Noteworthy, viewing the SPM Maps in combination with Clinical Scenarios significantly increased diagnostic confidence as compared to the combination of Clinical Scenarios and Standard FDG Images.

Finally, SPM analysis allowed a better delineation of the anatomical signatures specific for AD and FTLD spectrum, as shown by the disease differential indexes. Crucially, hypometabolism in the temporo-parietal and posterior cingulate for AD subjects, and the anterior cingulate, medial frontal and superior anterior temporal cortices for bvFTD subjects was more frequently identified by SPM Maps than Standard FDG Images.

The assessment of measured level of confidence revealed that diagnostic confidence of clinical information was lower in FTLD variants, in DLB and in some AD patients with atypical presentation. This could be partially ascribed to the fixed format of the clinical scenario that might have neglected some crucial clinical details. Nevertheless, the use of SPM Maps added a further value to clinical evaluation providing the clinician a higher degree of accuracy compared to Standard FDG Images.

The results of raters' performances on the MCI group proved that SPM Maps allow identifying with significantly stronger confidence distinct patterns of hypometabolism underlying MCI condition at the baseline, which predicted the further progression of cognitive decline to different dementia conditions at the clinical follow-up. According to previous findings, heterogeneous hypometabolic profiles may be recognized using FDG-PET in MCI subjects developing into different dementia diseases (Mosconi et al., 2008; Yakushev et al., 2009). In addition, the MCI individuals who reverted to normal cognition at the follow-up were all negative on SPM analysis at the baseline, thus supporting the role of our new SPM tool as an exclusionary test. It has been highlighted that a cognitively impaired subject with a negative PET scan has a low chance in the progression of cognitive disorders towards dementia (Silverman et al., 2008).

It is now known that different operating procedures of imaging biomarkers (visual inspection vs. semi-quantitative/quantitative method) can be responsible for the heterogeneous levels of their estimated diagnostic and prognostic accuracy (see Frisoni et al., 2013 and Perani et al., 2014 for meta-analyses). Within the different operating procedures for FDG-PET imaging, voxel-based analyses at a single-subject level are the most accurate, thus mandatorily calling for an objective statistical analysis of FDG-PET brain images that allows to reach the highest sensitivity and specificity (Perani et al., 2014).

Given that accuracy in dementia diagnosis highly depends not only on which marker (topographic or pathological) is used, but also on how it is measured (qualitative or semi-quantitative/quantitative) (Frisoni et al., 2013), in this context, we showed that SPM Maps specifically improve FDG-PET imaging in dementias.

The main limitation of this study that resides in the lack of pathological confirmation needs to be acknowledged. Further studies are necessary and some are in due course, in order to replicate the results in multi-site trials.

Appendix A. Supplementary data

Supplementary material for this article can be found online at http://dx.doi.org/10.1016/j.nicl.2014.10.009.

References

Albert, M.S., DeKosky, S.T., Dickson, D., et al., 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. Alzheimer's & Dementia: the Journal of the Alzheimer's Association 7 (3), 270–279. http://dx.doi.org/10.1016/j.jalz.2011.03.00821514249.

Anchisi, D., Borroni, B., Franceschi, M., et al., 2005. Heterogeneity of brain glucose metabolism in mild cognitive impairment and clinical progression to Alzheimer disease. Archives of Neurology 62 (11), 1728–1733. http://dx.doi.org/10.1001/archneur.62.11.172816286547.

Armstrong, M.J., Litvan, I., Lang, A.E., et al., 2013. Criteria for the diagnosis of corticobasal degeneration. Neurology 80 (5), 496–503. http://dx.doi.org/10.1212/WNL.0b013e31827f0fd123359374.

Bateman, R.J., Xiong, C., Benzinger, T.L., et al., 2012. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. New England Journal of Medicine 367 (9), 795–804. http://dx.doi.org/10.1056/NEJMoa120275322784036.

Bohnen, N.I., Djang, D.S., Herholz, K., Anzai, Y., Minoshima, S., 2012. Effectiveness and safety of ^{18}F-FDG PET in the evaluation of dementia: a review of the recent literature. Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine 53 (1), 59–71. http://dx.doi.org/10.2967/jnumed.111.09657822173840.

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. Journal of Neuroimaging: Official Journal of the American Society of Neuroimaging 15 (4), 348–355. http://dx.doi.org/10.1177/10512284052801691625440.

Caroli, A., Prestia, A., Chen, K., et al., 2012. Summary metrics to assess Alzheimer disease-related hypometabolic pattern with ^{18}F-FDG PET: head-to-head comparison. Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine 53 (4), 592–600. http://dx.doi.org/10.2967/jnumed.111.09494622343502.

Chen, W.P., Samuraki, M., Yanase, D., et al., 2008. Effect of sample size for normal database on diagnostic performance of brain FDG PET for the detection of Alzheimer's disease using automated image analysis. Nuclear Medicine Communications 29 (3), 270–276. http://dx.doi.org/10.1097/MNM.0b013e3282f3fa7618349798.

Chételat, G., Desgranges, B., Landeau, B., et al., 2008. Direct voxel-based comparison between grey matter hypometabolism and atrophy in Alzheimer's disease. Brain: A Journal of Neurology 131 (1), 60–71. http://dx.doi.org/10.1093/brain/awm28818063588.

De Souza, L.C., Chupin, M., Bertoux, M., et al., 2013. Is hippocampal volume a good marker to differentiate Alzheimer's disease from frontotemporal dementia? Journal of Alzheimer's Disease: JAD 36 (1), 57–66. http://dx.doi.org/10.3233/JAD-12229323542864.

Della Rosa, P.A., Cerami, C., Gallivanone, F., et al., 2014. A standardized [^{18}F]FDG-PET template for spatial normalization in statistical parametric mapping analysis pf brain images in dementia. Neuroinformatics http://dx.doi.org/10.1007/s12021-014-9235-424952892.

Dubois, B., Epelbaum, S., Santos, A., et al., 2013. Alzheimer disease: from biomarkers to diagnosis. Revue Neurologique 169 (10), 744–751. http://dx.doi.org/10.1016/j.neurol.2013.07.016.

Dubois, B., Feldman, H.H., Jacova, C., et al., 2014. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet. Neurology 13 (6), 614–629. http://dx.doi.org/10.1016/S1474-4422(14)70090-024849862.

Dubois, B., Feldman, H.H., Jacova, C., et al., 2010. Revising the definition of Alzheimer's disease: a new lexicon. Lancet. Neurology 9 (11), 1118–1127. http://dx.doi.org/10.1016/S1474-4422(10)70223-420934914.

Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B., Alzheimer's Disease Neuroimaging Initiative, 2013. Brain changes in older adults at very low risk for Alzheimer's disease. Journal of Neuroscience: the Official Journal of the Society for Neuroscience 33 (19), 8237–8242. http://dx.doi.org/10.1523/JNEUROSCI.5506-12.201323658162.

Foster, N.L., Heidebrink, J.L., Clark, C.M., et al., 2007. FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. Brain: A Journal of Neurology 130 (10), 2616–2635. http://dx.doi.org/10.1093/brain/awm17717704526.

Fotuhi, M., Do, D., Jack, C., 2012. Modifiable factors that alter the size of the hippocampus with ageing. Nature Reviews. Neurology 8 (4), 189–202. http://dx.doi.org/10.1038/nrneurol.2012.2722410582.

Frisoni, G.B., Bocchetta, M., Chételat, G., et al., 2013. Imaging markers for Alzheimer disease: which vs how. Neurology 81 (5), 487–500. http://dx.doi.org/10.1212/WNL.0b013e31829d86e823897875.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping 2 (4), 189–210. http://dx.doi.org/10.1002/hbm.460020402.

Galluzzi, S., Geroldi, C., Amicucci, G., et al., 2013. Supporting evidence for using biomarkers in the diagnosis of MCI due to AD. Journal of Neurology 260 (2), 640–650. http://dx.doi.org/10.1007/s00415-012-6694-023070466.

Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., et al., 2011. Classification of primary progressive aphasia and its variants. Neurology 76 (11), 1006–1114. http://dx.doi.org/10.1212/WNL.0b013e31821103e621325651.

Herholz, K., 2010. Cerebral glucose metabolism in preclinical and prodromal Alzheimer's disease. Expert Review of Neurotherapeutics 10 (11), 1667–1673. http://dx.doi.org/10.1586/ern.10.13620977325.

Herholz, K., Salmon, E., Perani, D., et al., 2002. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. Neuroimage 17 (1), 302–306. http://dx.doi.org/10.1006/nimg.2002.120812482085.

Jack Jr, C.R., Albert, M.S., Knopman, D.S., et al., 2011. Introduction to the recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia: the Journal of the Alzheimer's Association 7 (3), 257–262. http://dx.doi.org/10.1016/j.jalz.2011.03.00421514247.

Jack Jr, C.R., Knopman, D.S., Jagust, W.J., et al., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet. Neurology 9 (1), 119–128. http://dx.doi.org/10.1016/S1474-4422(09)70299-620083042.

McKeith, I.G., Dickson, D.W., Lowe, J., et al., 2005. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. Neurology 65 (12), 1863–1872. http://dx.doi.org/10.1212/01.wnl.0000187889.17253.b116237129.

McKhann, G.M., Knopman, D.S., Chertkow, H., et al., 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. Alzheimer's & Dementia: the Journal of the Alzheimer's Association 7 (3), 263–269. http://dx.doi.org/10.1016/j.jalz.2011.03.00521514250.

Minoshima, S., Foster, N.L., Sima, A.A., Frey, K.A., Albin, R.L., Kuhl, D.E., 2001. Alzheimer's disease versus dementia with Lewy bodies: cerebral metabolic distinction with autopsy confirmation. Annals of Neurology 50 (3), 358–365. http://dx.doi.org/10.1002/ana.113311558792.

Minoshima, S., Giordani, B., Berent, S., Frey, K.A., Foster, N.L., Kuhl, D.E., 1997. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. Annals of Neurology 42 (1), 85–94. http://dx.doi.org/10.1002/ana.4104201149225689.

Morbelli, S., Drzezga, A., Perneczky, R., et al., 2012. Resting metabolic connectivity in prodromal Alzheimer's disease. A European Alzheimer Disease Consortium (EADC) project. Neurobiology of Aging 33 (11), 2533–2550. http://dx.doi.org/10.1016/j.neurobiolaging.2012.01.00522365486.

Mosconi, L., Murray, J., Tsui, W.H., et al., 2014. Brain imaging of cognitively normal individuals with 2 parents affected by late-onset AD. Neurology 82 (9), 752–760. http://dx.doi.org/10.1212/WNL.00000000000018124523481.

Mosconi, L., Tsui, W.H., Herholz, K., et al., 2008. Multicenter standardized ^{18}F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine 49 (3), 390–398. http://dx.doi.org/10.2967/jnumed.107.04538518287270.

Patterson, J.C., Lilien, D.L., Takalkar, A., Pinkston, J.B., 2010. Early detection of brain pathology suggestive of early AD using objective evaluation of FDG-PET scans. International Journal of Alzheimer's Disease 2011. http://dx.doi.org/10.4061/2011/94659020858966.

Perani, D., 2014. FDG-PET and amyloid-PET imaging: the diverging paths. Current Opinion in Neurology 27 (4), 405–413. http://dx.doi.org/10.1097/WCO.00000000000010924927239.

Perani, D., 2008. Functional Neuroimaging of Cognition. Handbook of Clinical Neurology 88, pp. 61–111.

Perani, D., Schillaci, O., Padovani, A., et al., 2014. A survey of FDG and amyloid PET imaging in dementia and GRADE analysis. BioMed Research International 2014, 785039.

Petersen, R.C., Roberts, R.O., Knopman, D.S., et al., 2009. Mild cognitive impairment: ten years later. Archives of Neurology 66 (12), 1447–1455. http://dx.doi.org/10.1001/archneurol.2009.26620008648.

Prestia, A., Caroli, A., van der Flier, W.M., et al., 2013. Prediction of dementia in MCI patients based on core diagnostic markers for Alzheimer disease. Neurology 80 (11), 1048–1056. http://dx.doi.org/10.1212/WNL.0b013e318287283023390179.

Rascovsky, K., Hodges, J.R., Knopman, D., et al., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134 (9), 2456–2477.

Ridgway, G.R., Omar, R., Ourselin, S., Hill, D.L., Warren, J.D., Fox, N.C., 2009. Issues with threshold masking in voxel-based morphometry of atrophied brains. Neuroimage 44 (1), 99–111. http://dx.doi.org/10.1016/j.neuroimage.2008.08.04518848632.

Salmon, E., Garraux, G., Delbeuck, X., et al., 2003. Predominant ventromedial frontopolar metabolic impairment in frontotemporal dementia. Neuroimage 20 (1), 435–440. http://dx.doi.org/10.1016/S1053-8119(03)00346-X14527604.

Signorini, M., Paulesu, E., Friston, K., et al., 1999. Rapid assessment of regional cerebral metabolic abnormalities in single subjects with quantitative and nonquantitative [^{18}F]FDG PET: a clinical validation of statistical parametric mapping. Neuroimage 9 (1), 63–80. http://dx.doi.org/10.1006/nimg.1998.03819918728.

Silverman, D.H., Mosconi, L., Ercoli, L., Chen, W., Small, G.W., 2008. Positron emission tomography scans obtained for the evaluation of cognitive dysfunction. Seminars in Nuclear Medicine 38 (4), 251–261. http://dx.doi.org/10.1053/j.semnuclmed.2008.02.00618514081.

Silverman, D.H., Truong, C.T., Kim, S.K., et al., 2003. Prognostic value of regional cerebral metabolism in patients undergoing dementia evaluation: comparison to a quantifying parameter of subsequent cognitive performance and to prognostic assessment without PET. Molecular Genetics and Metabolism 80 (3), 350–355. http://dx.doi.org/10.1016/S1096-7192(03)00139-214680983.

Spence, J.S., Carmack, P.S., Gunst, R.F., Schucany, W.R., Woodward, W.A., Haley, R.W., 2006. Using a white matter reference to remove the dependency of global signal on experimental conditions in SPECT analyses. Neuroimage 32 (1), 49–53. http://dx.doi.org/10.1016/j.neuroimage.2006.03.02516651010.

Sperling, R.A., Aisen, P.S., Beckett, L.A., et al., 2011. Toward defining the preclinical stage of Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia: the Journal of the Alzheimer's Association 7 (3), 280–292. http://dx.doi.org/10.1016/j.jalz.2011.03.00321514248.

Teune, L.K., Bartels, A.L., de Jong, B.M., et al., 2010. Typical cerebral metabolic patterns in neurodegenerative brain diseases. Movement Disorders: Official Journal of the Movement Disorder Society 25 (14), 2395–2404. http://dx.doi.org/10.1002/mds.2329120669302.

Varrone, A., Asenbaum, S., Vander Borght, T., et al., 2009. European association of nuclear medicine neuroimaging committee. EANM procedure guidelines for PET brain imaging using [^{18}F]FDG, version 2. European Journal of Nuclear Medicine and Molecular Imaging 36 (12), 2103–2110. http://dx.doi.org/10.1007/s00259-009-1264-019838705.

Yakushev, I., Hammers, A., Fellgiebel, A., et al., 2009. SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. Neuroimage 44 (1), 43–50. http://dx.doi.org/10.1016/j.neuroimage.2008.07.01518691659.