Visual Versus Semi-Quantitative Analysis of $^{18}$F-FDG-PET in Amnestic MCI: An European Alzheimer’s Disease Consortium (EADC) Project

Silvia Morbellia, Andrea Brugnolob, Irene Bossertc, Ambra Buschiazzoa, Giovanni B. Frisonic, Samantha Galluzzic, Bart N.M. van Berckele, Rik Ossenkoppelee, Robert Perneckzy, Alexander Drzezgai, Mira Didicj, Eric Guedjk, Gianmario Sambucetia, Gianluca Bottonia, Dario Arnaldib, Agnese Picco, Fabrizio De Carlil, Marco Paganm, and Flavio Nobilih

aNuclear Medicine, Department of Health Sciences (DISSAL), University of Genoa, IRCCS AOU San Martino–IST, Genoa, Italy
bClinical Neurology, Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, and Mother-Child health (DINOGMI), University of Genoa, IRCCS AOU San Martino–IST, Genoa, Italy
cLENITEM Laboratory of Epidemiology and Neuroimaging-IRCCS S. Giovanni di Dio-FBB, Brescia, Italy
dUniversity Hospitals and University of Geneva, Geneva, Switzerland
*eDepartment of Nuclear Medicine & PET Research, VU University Medical Center, Amsterdam, The Netherlands
fNeuropediatrics and Ageing Research Unit, School of Public Health, Faculty of Medicine, The Imperial College of Science, Technology, and Medicine, London, UK
fWest London Cognitive Disorders Treatment and Research Unit, London, UK
iDepartment of Psychiatry and Psychotherapy, Technische Universität, Munich, Germany
jDepartment of Nuclear Medicine, Technische Universität, Munich, Germany
kAPHM, CHU Timone, Service de Neurologie et Neuropsychologie, Aix-Marseille University, INSERM U 1106, Marseille, France
lAPHM, CHU Timone, Service de Médecine Nucléaire; Aix-Marseille University, CERIMED, INT CNRS UMR7289, Marseille, France
mInstitute of Bioimaging and Molecular Physiology, Consiglio Nazionale delle Ricerche (CNR), Genoa, Italy
nInstitute of Cognitive Sciences and Technologies, Consiglio Nazionale delle Ricerche (CNR), Rome, Italy
oDepartment of Nuclear Medicine, Karolinska University Hospital, Stockholm, Sweden

Accepted 25 September 2014

Abstract. We aimed to investigate the accuracy of FDG-PET to detect the Alzheimer’s disease (AD) brain glucose hypometabolic pattern in 142 patients with amnestic mild cognitive impairment (aMCI) and 109 healthy controls. aMCI patients were followed for at least two years or until conversion to dementia. Images were evaluated by means of visual read by either moderately-skilled or expert readers, and by means of a summary metric of AD-like hypometabolism (PALZ score). Seventy-seven patients

Current address: Department of Nuclear Medicine, University of Cologne, Cologne, Germany.
Correspondence to: Flavio Nobile, MD, Clinical Neurology, Department of Neuroscience (DINOGMI), University of Genoa, IRCCS AOU San Martino-IST, Largo P. Dante, 3, 16132 Genoa, Italy. Tel: +390103537568; Fax: +390103535689; E-mail: flaviomariano.nobili@hsanmartino.it.

ISSN 1387-2677/15/$35.00 © 2015 – IOS Press and the authors. All rights reserved
converted to AD dementia after 28.6 ± 19.3 months of follow-up. Expert reading was the most accurate tool to detect these MCI converters from healthy controls (sensitivity 89.6%, specificity 89.0%, accuracy 89.2%) while two moderately-skilled readers were less (p < 0.05) specific (sensitivity 85.7%, specificity 79.8%, accuracy 82.3%) and PALZ score was less (p < 0.001) sensitive (sensitivity 62.3%, specificity 91.7%, accuracy 79.6%). Among the remaining 67 aMCI patients, 50 were confirmed as aMCI after an average of 42.3 months, 12 developed other dementia, and 3 reverted to normalcy. In 30/50 persistent MCI patients, the expert recognized the AD hypometabolic pattern. In 13/50 aMCI, both the expert and PALZ score were negative while in 7/50, only the PALZ score was positive due to sparse hypometabolic clusters mainly in frontal lobes. Visual FDG-PET reads by an expert is the most accurate method but an automated, validated system may be particularly helpful to moderately-skilled readers because of high specificity, and should be mandatory when even a moderately-skilled reader is unavailable.

Keywords: Alzheimer’s disease, amnestic mild cognitive impairment, FDG-PET, PALZ score, visual read

INTRODUCTION

18F-fluorodeoxyglucose positron emission tomography (FDG-PET) has long been recognized as an accurate biomarker for the diagnosis Alzheimer’s disease (AD) dementia [1, 2]. The typical ‘AD hypometabolic pattern’ is well known and mainly includes the posterior cingulate cortex, precuneus, posterior temporoparietal cortex, and medial temporal lobe hypometabolism. With different metrics, sensitivity and specificity above 80% have been consistently reported in AD dementia [3].

As FDG-PET correlates well with the severity of cognitive impairment [4], the AD hypometabolic pattern may be less evident at the stage of mild cognitive impairment (MCI) due to AD and consequently the sensitivity could be lower. Nowadays, the main question is to identify the AD hypometabolic phenotype at the MCI stage, or even earlier, in order to intervene promptly and to access clinical trials with disease-modifying drugs.

Recent FDG-PET data are emerging in MCI cohorts from large multicenter studies [5–6], with medium-term follow-up (i.e., 2–3 years), showing variable sensitivity and specificity, mainly depending on the metric used to analyze data [5]. In this recent meta-analysis in pooled 532 MCI patients, average FDG-PET prognostic sensitivity to the development of AD dementia, obtained with different metrics, was 76% and average specificity versus non-converter patients was 74%. Visual read was the most sensitive (94%) but the least specific (68%) tool, individual case analysis with statistical parametric mapping (SPM) being the most specific (92%) but the second least sensitive (72%).

However, while diagnostic sensitivity can be easily computed in MCI converters, specificity computation versus non-converters is seriously flawed by the limited time of follow-up in most studies, as it has been shown that a non-negligible part of MCI patients can take up to 7 years to convert to dementia [10]. Indeed, hypometabolism within the AD pattern can be identified also in the group of non-converter MCI patients [11–13], especially in those showing cognitive worsening during the follow-up [14].

As for all the other AD biomarkers, one of the main issues is how FDG-PET images are evaluated and/or measured. In the above mentioned meta-analysis, the authors concluded that diagnostic accuracy of imaging biomarkers is at least as dependent on how the biomarker is measured as on the choice of biomarker itself [3]. As a matter of fact, FDG-PET scans are still usually reported visually by nuclear medicine physicians which introduces another ‘variable in the equation’ of PET reporting because readers’ expertise may range widely. On the other side, the automatic or semi-automatic available tools have a lower inter-rater variability but some of them, such as the SPM maps, are not user-friendly and the statistical output maps still require to be evaluated, especially if few and sparse hypometabolic clusters are highlighted. Only some of them are freely available and they are not widely distributed, especially in nuclear medicine departments without expertise or specific interest in brain imaging.

Even more importantly, the direct comparison between visual read and statistically-assisted interpretation is poorly known in MCI patients. In a recent paper comparing MCI patients and controls, statistical surface projection (SSP) software improved specificity versus visual read while the sensitivity was similar versus both the beginner and the expert reader [15]. The FDG-PET guidelines of the European Association of Nuclear Medicine (EANM) [16] admit the use of both visual and software-aided approaches, recommending their combined use but not suggesting further procedural reading guidelines, and a similar position has been taken by the Society of Nuclear Medicine in its 2009 guidelines. Accordingly, especially in daily clinical practice, the best way of reporting FDG-PET scans...
is still a matter of debate and clear statements from international societies are lacking.

We compared the diagnostic performance of visual read with that of an automatic tool [17, 18] yielding a statistical classification of FDG-PET scans in 109 healthy subjects and 142 patients with amnestic MCI (aMCI) recruited by the European Alzheimer Disease Consortium (EADC) PET project, followed-up for at least 2 years or until conversion to dementia.

METHODS

Subjects

Clinical, neuropsychological, and FDG-PET data derive from the FDG-PET project of the EADC (http://www.eadc.info/sito/pagine/a07.php?nav=a0), aimed at joining together data from patients with MCI and normal controls. Six EADC centers (namely, Amsterdam, Brescia, Genoa, Marseille, Munich, Perugia) in four countries participated.

Patients with aMCI with or without impairment in other cognitive domains and healthy controls have been enrolled. To date, data from 109 healthy controls and 225 aMCI patients are available, but for the purpose of the present study, we considered only those aMCI patients with a clinical follow-up of either at least two years or until the development of dementia. These were 142 patients whose main demographic and clinical characteristics are shown in Table 1.

As for aMCI, the inclusion and exclusion criteria are explained in previous papers [19, 20]. In summary, they basically included outpatients newly referred for cognitive complaints to a center dedicated to the evaluation of cognitive disorders. Cognitive complaints mainly included memory complaints but they could also include difficulties in other cognitive domains, such as attention and orientation. Main exclusion criteria were dementia; any somatic, metabolic, psychiatric, or neurological disorder that may cause cognitive impairment. Dementia was excluded by the clinical interviews with patients and caregivers that ruled out significant impairment in activities of daily living, and by means of the Clinical Dementia Rating (CDR) scale, scoring 0.5 in all aMCI patients. General cognition was assessed by means of the Mini-Mental State Examination (MMSE) in all centers.

All subjects underwent rating scales for depression and neuropsychiatric symptoms according to the routine in use in each center. Patients with major depression, delusions, or hallucinations were excluded. Structural neuroimaging was mainly performed by means of magnetic resonance imaging (MRI) while computed tomography (CT) was performed when MRI was unfeasible because of contraindications or patient intolerance.

Neuropsychological tests were administered in the domains of memory, language, executive function, attention, and visuoconstruction, according to the routine of each center, in order to define the aMCI syndrome, according to the Petersen’s criteria [21]. Raw scores were converted to age-, education-, and gender-corrected Z-scores according to locally collected or published normative data in use in each center that vary among countries and languages. Impairment was defined as a Z-score of −1.5 or lower. Subjects with impairment in the memory domain only (single-domain aMCI) or with impairment in the memory domain plus impairment in non-memory domains (multidomain aMCI) were defined as aMCI.

The 142 aMCI patients were regularly followed-up to identify those developing dementia. During this

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Main demographic and clinical characteristics of 142 patients with aMCI followed-up for at least 2 years (or until the conversion to dementia) in the EADC-PET project. Values are expressed as mean ± standard deviation.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>142</td>
</tr>
<tr>
<td><strong>Age (y)</strong></td>
<td>Gender M/F</td>
</tr>
<tr>
<td><strong>Education (y)</strong></td>
<td>9.64 ± 4.20</td>
</tr>
<tr>
<td><strong>MMSE (mean ± SD)</strong></td>
<td>28.50 ± 1.98</td>
</tr>
<tr>
<td><strong>Follow-up (mos)</strong></td>
<td>33.44 ± 19.50</td>
</tr>
<tr>
<td><strong>Neuropsychological Z scores</strong></td>
<td>3.61 ± 1.12</td>
</tr>
<tr>
<td><strong>Verbal memory Immediate recall</strong></td>
<td>3.61 ± 1.12</td>
</tr>
<tr>
<td><strong>Verbal memory Delayed recall</strong></td>
<td>3.61 ± 1.12</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>3.61 ± 1.12</td>
</tr>
<tr>
<td><strong>Visuoconstruction</strong></td>
<td>3.61 ± 1.12</td>
</tr>
<tr>
<td><strong>Executive Function</strong></td>
<td>3.61 ± 1.12</td>
</tr>
</tbody>
</table>
period, patients were carefully treated for systemic co-morbidity. None of the patients was taking antide-
mentia medication at the time of either PET scan or 
during the follow-up period, until a formal diagnosis 
of AD-dementia was made.

Diagnosis of dementia was established according to 
the criteria in use at the beginning of the study. Demen-
tia of the AD type was diagnosed according to the 
National Institute of Neurological and Communicative 
Disorders, Stroke, Alzheimer’s Disease and Related 
Disorders Association (NINCDS-ADRDA) criteria 
[22] and to the Diagnostic and Statistical Manual of 
Mental Disorders (DSM-IV R) criteria. Diagno-
sis of frontotemporal lobar degeneration (FTLD) and 
dementia with Lewy bodies (DLB) was formulated 
according to the Neary et al. [23] and McKeith et al. 
[24] criteria, respectively.

Healthy controls recruited by the same EADC cen-
ters have been described in previous papers [19, 20]. 
Briefly, they were volunteers undergoing the same 
clinical and neuropsychological evaluation as patients 
and were not taking neuropsychotropic drugs or drugs 
known to interfere with cerebral metabolism. Only sub-
jects with a MMSE score ≥26 and a CDR = 0 were 
included. There were 109 subjects, 53 males and 56 
females, mean age 67.0 ± 6.6 years, mean MMSE 
score: 29.3 ± 1.0, mean education: 11.4 ± 4.0 years. 
Their healthy status was checked again with a clini-
cal interview about one year later (mean: 11.8 ± 4.8 
months).

Subject information and scans uploaded by each 
center were as follows: Amsterdam, 16 patients and 20 
controls; Brescia, 43 patients and 27 controls; Genoa, 
40 patients and 33 controls; Marseille, 9 patients and 10 
controls; Munich, 34 patients and 19 controls; at the 
time of this data analysis, Perugia had not uploaded 
cases yet.

The study was carried out in accordance with the 
content of the Helsinki declaration and was approved 
by the local Ethics Committees; all subjects provided 
written informed consent.

PET procedures

FDG-PET was performed within 2 months from 
the baseline clinical-neuropsychological examination 
according to the EANM guidelines in use when the 
study began [25]. Subjects fasted for at least 6h. 
Before radiopharmaceutical injection, blood glucose 
was checked and was <140 mg/dL in all subjects. Sub-
jects were injected with 185–250 MBq of 18F-FDG 
via venous cannula. Required minimum time interval 
between injection and scan start was 30 min. Required 
minimum scan duration was 10 min. Emission scans 
were acquired in 3-dimensional mode. Images were 
reconstructed using an ordered subset expectation 
maximization (OSEM) algorithm in all centers but 
Amsterdam (filtered backprojection). Attenuation cor-
rection was based on CT scan in Brescia, Genoa, and 
Marseille, and on transmission scan in Munich and 
Amsterdam. Scatter correction was performed using 
standard software as supplied by scanner manufac-
turers. Technical details of the scanners used in the 
different centers are reported in previous papers [19, 
20]. Anonymous scans and clinical information were 
uploaded in a dedicated safe file transfer protocol 
web site.

Image analysis

Our aim was to compare the ability of moderately-
skilled readers, an expert reader, and an automatic 
voxel-based analysis method. In order to reproduce 
‘the real world’ of FDG-PET reporting, we imagined 
two conditions, i.e., one expert reporting alone and 
two moderately-skilled readers reporting together and 
reaching a final common opinion. We choose not to 
test the performance of a beginner reader because we 
believe that recognizing subtle metabolic abnormali-
ties in such a clinical context cannot be faced by a 
beginner left alone. Also, we chose not to evaluate 
the incremental diagnostic value of PALZ score with 
respect to qualitative readings because we wanted to 
directly compare the two reading systems and obtain 
accuracy values for both, which is poorly known in 
MCI patients, prior to evaluate in another series the 
incremental value of the automatic system to visual 
reads.

Then, we first defined a ‘moderately-skilled’ reader 
as a nuclear medicine physician with about a 3-year 
experience in weekly reporting of FDG-PET scans and 
a total of about 500 scans commented and reported 
together with an expert. Secondly, we defined an 
‘expert’ reader one with at least 10-year experience 
in weekly reporting of FDG-PET scans and with sus-
tained scientific activity in the field. We then identified 
one expert and two moderately-skilled readers, the 
ltter reading the scans together and were forced to 
reach a consensus. All raters read the 251 scans (142 
patients and 109 controls) without any other infor-
mation and forced to a dichotomous reading, i.e., 
Alzheimer hypometabolic pattern present or absent. 
Only when the Alzheimer pattern was absent, then 
the readers were invited to express whether another
hypometabolic pattern could be identified, including the FTLD and the DLB pattern. The Alzheimer pattern was defined according to the large literature evidence and included hypometabolism in at least one of the following: posterior cingulate cortex, precuneus, posterior temporoparietal cortex, and medial temporal lobe. The FTLD pattern included at least one of the following: orbitofrontal cortex, dorsolateral frontal cortex, ventrolateral frontal cortex, anterior temporal lobe, insula, and anterior cingulate cortex. The DLB pattern included the Alzheimer pattern without necessarily the medial temporal lobe but with the lateral occipital cortex.

The PMOD software (PMOD technologies, http://www.pmod.com) was used for automatic, voxel-based evaluation of scans with the ‘Alzheimer’ option computing the ‘Probability of Alzheimer’ or PALZ score. The PALZ score is a voxel-based parametric mapping method yielding diagnostic information on brain regions that are typically affected in AD [17]. Individual FDG-PET images are compared to a fixed database of normal elderly scans through a voxel-wise t-test, including age as confounding variable. The PALZ score is computed as voxel-by-voxel sum of t-scores in a pre-defined AD-pattern mask. PALZ was developed to distinguish AD from healthy controls above 50 years of age. The threshold for abnormality was set at 11,090 (t-sum higher than 11,090 = abnormal FDG-PET). As some centers provided Dicom files and some others provided Analyze files, Dicom files were converted to Analyze format. In fact, the use of different formats gives rise to slight differences in PALZ scores, as already outlined in a previous paper [6]. With analyze format, images must be checked for correct display orientation and eventually reoriented to anatomic position, preliminary to PALZ score computation.

**Statistics**

Confidence intervals (CI) for sensitivity, specificity, accuracy, positive (PPV) and negative (NPV) predictive values were computed by using the Wald interval for binomial proportions [26]. CI for positive (+LR) and negative (−LR) likelihood ratios were computed according to Simel et al. [27], while CI for the odds ratio were computed by the conventional log-transform method [28]. Paired comparisons between classifiers for sensitivity, specificity, and accuracy were performed by McNemar binomial test. Paired comparison between classifiers for PPV and NPV was performed according to the generalized score statistic proposed by Leisenring et al. [29]. Paired comparisons between classifiers for +LR and −LR were performed by the evaluation of log-transformed LR according to Nofuentes & Del Castillo [30]. Paired comparisons between classifiers for odds ratio were performed by evaluating conditional relative odds ratio, based on McNemar’s odds ratio, according to the method described by Suzuki [31]. Receiver operating characteristic comparison was unfeasible due to the dichotomous output of visual reading.

**RESULTS**

**Follow-up diagnosis**

During the follow-up period (mean in the 142 patients: 33.4 ± 19.5 months, range: 6–98), 77 patients (54.2%) progressed to AD dementia (‘prodromal AD’ patients, pAD), 10 patients (7%) developed FTLD, 2 patients progressed to DLB, and 3 patients reverted to a normal condition. The remaining 50 patients were confirmed to have aMCI at the last available follow-up, ranging from 24 to 98 months (details in Table 1).

**Control group**

The AD-hypometabolic pattern was seen in 22 healthy subjects by the moderately-skilled readers (79.8% specificity), in 12 subjects by the expert (89% specificity) while the PALZ score yielded a statistically significant result in 9 controls (91.7% specificity). Specificity was significantly higher for either the expert or PALZ score than for the moderately-skilled readers (p < 0.05) but did not differ between expert and PALZ score. Concordance between the two visual reading systems was 78%, between the expert and PALZ score was 84.4%, and between the moderately-skilled readers and PALZ score was 75.2%. Figure 1 shows an example of a negative PALZ score in a control subject in whom the AD hypometabolic pattern was erroneously recognized by the moderately-skilled readers.

**MCI patients who later converted to AD dementia (prodromal AD, pAD)**

The conversion to AD dementia in these 77 patients was fixed as the ‘gold standard’ to evaluate sensitivity of different PET readings. The moderately-skilled readers, the expert, and the PALZ score identified the AD hypometabolic pattern in 85.7%, 89.6%, and 62.3% of patients, respectively. Sensitivity was higher for both expert (p < 0.001) and moderately-skilled
Fig. 1. Example of false positive reading with the moderately-skilled readers. The AD hypometabolic pattern was erroneously recognized by the moderately-skilled readers in a 62 year-old male control subject with a normal PALZ score (T-sum within AD regions: 8.044). On top, FDG-PET visual analysis discloses very mild and sparse asymmetries in frontal-insular and inferior parietal regions (left minus) and in superior parietal region (right minus) (higher to lower uptake in white, red, yellow, green, and blue colors in decreasing order). On bottom, statistical maps obtained by means of the PMOD software showing very small and sparse significant hypometabolic clusters (in red).

Fig. 2. Example of false negative reading with the T-sum score. The AD hypometabolic pattern was recognized by the expert while the PALZ score was not significantly increased in a 72 year-old man converted to AD-dementia after 40 months of follow-up. On top, FDG-PET visual analysis discloses moderately reduced uptake in left temporo-parietal cortex while the statistical PMOD maps (on bottom) shows significant hypometabolic clusters in bilateral temporal regions and left parietal cortex without reaching the statistical threshold (T-sum = 7.225). Other details as in Fig. 1.

readers ($p < 0.01$) than for PALZ score while did not significantly differ between the two visual readings. Concordance between the two visual readings was 88.3%, between the expert and PALZ score was 62.3%, and between the moderately-skilled readers and PALZ score was 63.6%. In three patients, neither the expert nor the PALZ score were consistent for the presence of an AD hypometabolic pattern. Fig. 2 shows an example of a negative PALZ score in a patient in whom the AD hypometabolic pattern was recognized by the expert. In summary, total accuracy of moderately-skilled readers, expert and T-sum was 82.30%, 89.25%, and...
Table 2

Summary of FDG-PET analysis in 77 patients with aMCI and 109 controls from the EADC-PET project

<table>
<thead>
<tr>
<th></th>
<th>Moderately-skilled readers</th>
<th>Expert reader</th>
<th>PALZ score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity (95% C.I.)</td>
<td>0.857 (0.755–0.923)**</td>
<td>0.900 (0.803–0.931)</td>
<td>0.902 (0.806–0.940)**</td>
</tr>
<tr>
<td>Specificity (95% C.I.)</td>
<td>0.798 (0.708–0.867)**</td>
<td>0.800 (0.712–0.919)</td>
<td>0.817 (0.725–0.944)**</td>
</tr>
<tr>
<td>Total accuracy (95% C.I.)</td>
<td>0.825 (0.760–0.874)*</td>
<td>0.812 (0.739–0.878)*</td>
<td>0.796 (0.711–0.878)*</td>
</tr>
<tr>
<td>Positive predictive value (95% C.I.)</td>
<td>0.750 (0.646–0.833)*</td>
<td>0.852 (0.752–0.938)</td>
<td>0.842 (0.756–0.921)</td>
</tr>
<tr>
<td>Negative predictive value (95% C.I.)</td>
<td>0.888 (0.804–0.940)**</td>
<td>0.924 (0.851–0.964)</td>
<td>0.775 (0.691–0.842)**</td>
</tr>
<tr>
<td>-Likelihood ratio (95% C.I.)</td>
<td>4.25 (2.89–6.24)**</td>
<td>8.14 (4.75–13.96)</td>
<td>7.55 (4.94–14.45)</td>
</tr>
<tr>
<td>Likelihood ratio (95% C.I.)</td>
<td>0.18 (0.10–0.31)**</td>
<td>0.12 (0.06–0.23)</td>
<td>0.41 (0.31–0.55)**</td>
</tr>
<tr>
<td>Odds ratio (95% C.I.)</td>
<td>23.73 (10.75–52.35)</td>
<td>69.7 (27.06–179.62)</td>
<td>18.39 (8.07–41.89)*</td>
</tr>
<tr>
<td>Z value</td>
<td>7.84</td>
<td>8.79</td>
<td>6.93</td>
</tr>
</tbody>
</table>

**p<0.001; ***p<0.01; *p<0.05 in the comparison between either moderately-skilled readers or PALZ score and the expert reader.**

and 79.57%, respectively. Table 2 summarizes the results and reports positive (PPV) and negative (NPV) predictive values, positive (+LR) and negative (−LR) likelihood ratio, and odds ratio for the three reading systems.

Patients with persistent MCI at follow-up

FDG-PET scan reading in persistent MCI group without a definite clinical outcome was faced by considering the best figures of sensitivity and specificity reached in the 77 pAD patients and in the 109 controls, i.e., the expert reading was the gold-standard for sensitivity and the concordance of both expert reading and PALZ score was the gold-standard as for specificity. In fact, specificity values reached by experts and PALZ score were very similar. We thus identified a group of aMCI patients likely expressing the AD hypometabolic pattern based on the expert’s score and a group of patients in whom the AD hypometabolic pattern was unlike based on the concurring negative evaluation of the expert and of PALZ score. The expert identified the AD-hypometabolic pattern in 30 patients (aMCIExp+) (Fig. 3 shows a patient with a positive PALZ score also) while both the expert and the PALZ score were negative in 13 cases (aMCIExp-/PALZ-). In the remaining 7 patients, the PALZ score was positive but the expert did not identify the AD-hypometabolic pattern (aMCIExp-/PALZ+) (example in Fig. 4).

Table 3 reports the main data for the 30 aMCI-Exp+ and 13 aMCIExp-/PALZ- subgroups. Briefly, the two subgroups did not show significant differences for...
Fig. 4. Example of absence of AD hypometabolic pattern according to the expert but with a positive PALZ score (T-sum = 15,180) in a 79 year-old woman with persistent aMCI after a 42-month follow-up. On top, FDG-PET visual analysis does not clearly show reduced uptake in the typical AD regions. On bottom, statistical PMOD maps shows significant hypometabolic clusters sparsely distributed in bilateral frontal and temporal lobes, the anterior cingulate cortex, right cuneus and posterior cingulate cortex, and left parietal lobe. Other details as in Fig. 1.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>aMCIExp+</th>
<th>aMCIExp-/PALZ-</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>30</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Age (y; mean ± SD)</td>
<td>69.7 ± 8.69</td>
<td>69.1 ± 7.70</td>
<td>n.s.</td>
</tr>
<tr>
<td>Gender M/F</td>
<td>22/8</td>
<td>6/7</td>
<td>n.s.</td>
</tr>
<tr>
<td>Education (y; mean ± SD)</td>
<td>10.0 ± 4.75</td>
<td>10.5 ± 3.26</td>
<td>n.s.</td>
</tr>
<tr>
<td>PALZ (T-sum score; mean ± SD)</td>
<td>20359.53 ± 15799.30</td>
<td>4993.47 ± 3248.69</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Basal MMSE (mean ± SD)</td>
<td>26.77 ± 1.74</td>
<td>27.78 ± 2.28</td>
<td>n.s.</td>
</tr>
<tr>
<td>Follow-up MMSE (mean ± SD)</td>
<td>24.43 ± 2.18</td>
<td>26.15 ± 3.56</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Follow-up length (mo; mean ± SD)</td>
<td>41.67 ± 16.67</td>
<td>44.23 ± 21.89</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

aMCIExp+, patients in whom the expert identified the AD hypometabolic pattern; aMCIExp-/PALZ-, patients in whom neither the expert nor the PALZ score identified an AD hypometabolic pattern. PALZ (T-sum score), number of voxels in the AD hypometabolic regions as identified by the automated analysis of PALZ.

DISCUSSION

This study shows that visual read of FDG-PET is quite sensitive in identifying the AD hypometabolic pattern in aMCI patients who later converted to AD dementia, if adequate training and expertise is provided. Sensitivity was slightly but not significantly
lower in the moderately-skilled readers than in expert reader while was significantly lower for PALZ score. Thus the trained reader can pick-up subtle abnor-
malities that do not reach the threshold of statistical significance set for the T-sum score. This better sensitivity can be due to several factors, including the ability of the skilled reader to highlight inter-hemispheric asymmetries in critical areas, a feature that is not available in the PALZ scoring system, or by other voxel-based statistical mapping system, such as SPM. Only very recently some automatic tools for the analysis of FDG-PET in dementia have embedded asymmetry computation [32]. Evaluation of asymme-
tries was already regarded as a very sensitive approach to detect abnormalities in brain function in dementia [33] and in cerebrovascular disease [34]. Indeed, eval-
uation of asymmetries is among the cornerstones of reader training in specialized centers and one of the main index on which the expert reader bases his/her evaluation. In milder cases, such as aMCI patients, a metabolic asymmetry can be highlighted by the visual reading but either the relatively hypometabolic region may be not large enough or the hypometabolism may be not severe enough to reach the statistical thresh-
old on PALZ score. On the other hand, the PALZ score sensitivity was set based on patients with AD dementia and even if it has been shown to reach 83% in very mild AD patients [17] and 85% in mild AD [18], it has not been specifically trained and validated in prodromal AD patients (i.e., aMCI patients later converted to AD dementia). Indeed, few data have dealt with sensitiv-
ity of PALZ score in MCI converter patients showing a sensitivity of 57% in 30 MCI converter patients coming from the ADNI database [18] and of 79% in 14 MCI convert-
er patients in a naturalistic population [35]. Thus, the present series of MCI converter patients is the largest ever assessed with the PALZ score. A similar phenomenon has been verified for an MRI automatic analysis method measuring temporal atrophy that was less accurate in distinguishing between MCI convert-
ers and non-converters when the tool was trained in AD-dementia patients versus controls rather than in MCI converters versus non-converters [36].

Sensitivity of FDG-PET between 80% and 90% in prodromal AD patients has been recently found in other cohorts with different metrics. Landau et al. [37] reported 82% sensitivity with region of interest analysis in critical regions in 28 patients with MCI due to AD from the ADNI database, followed for 12–36 months (mean 21). In the study by Shaffer et al. [7] in 43 patients with MCI due to AD, sensitivity by means of principal component analysis was 86%.

Using a ‘hypometabolic convergence index’, sensitivity of 81.8% was found in 36 MCI ADNI patients converted to AD dementia in 18 months [5]. A recent meta-analysis pooling 532 MCI patients described in 10 studies reported a mean sensitivity of 76%, obtained with different metrics, but in particular the sensitivity reached with visual read was 94%, thus close to the figure obtained in our study [3].

Instead, specificity was fairly good especially with the PALZ score, slightly lower with the expert but significantly lower with the moderately-skilled read-
ers. The 91.7% specificity is close to the high figure reported in previous studies with the PALZ score, comparing AD patients and controls, ranging between 78.4% and 94.4% [17, 18, 38]. The moderately-skilled reader tends therefore to overemphasize minimal changes that are part of the between-subject variability in brain function. Thus, only the expert could reason-
avably avoid the assistance of the automatic system while the moderately-skilled readers may take advan-
tage of it to improve specificity. We choose not to test the ‘beginner’ reader, but the results obtained in the moderately-skilled readers strongly suggest that the assistance of an automatic and validated reading sys-
tem is mandatory for beginners who cannot access an expert-guided reader training and have not completed a relevant number of scans reported with supervi-
sion. The significant support of an automatic system to visual reading for moderately-skilled readers is in sub-
stantial agreement with the few other papers assessing this issue in smaller groups. In 39 AD patients and 40 subjects without AD, 3D-SSP significantly improved accuracy of novice readers allowing them to score similar-
ly as expert readers, but 3D-SSP did not significantly improve the performance of experts which was already rather high [39]. In another paper in 30 AD patients and 30 controls, adopting volumes of interest analy-
thesis as derived by 3D-SSP, experts performed better than beginners and the automatic analysis significantly improved beginner performance in the 15 patients with late-onset AD, but not expert performance [40]. The only other available paper in MCI concerned a group of 18 patients and showed that 3D-SSP increased speci-
ficity, but not sensitivity, of both beginner and expert readers [15].

An intriguing part of this study concerns the 50 patients still in an aMCI stage after a mean time of 3 years and a half since the first evaluation. We know that either these patients have been assessed very early on the path of AD evolution (late converters) or they have MCI not related to a neurodegenerative disease. In thirty of them, the expert identified the typical AD
interpretation of scans [48]. This subgroup of patients
a software-generated map) is mandatory for the final
of dementia (obtained either visually or by means of
hypometabolic patterns associated with different types
This finding underlines that the adequate knowledge of
tia, they can develop other types of dementia [47].
is in keeping with the notion that, although aMCI
or DLB, and FDG-PET visual reading was consistent
with the diagnosis in all but one patient. This result
was not specifically addressed in this study. In seven
patients, the PALZ score was positive but the expert
did not recognize the AD hypometabolic pattern. The
interpretation of this unexpected result, given the lower
sensitivity and the paired specificity of PALZ score as
compared to the expert, seems to rely in the pecu-
liar construct of PALZ score that includes several
areas in the frontal lobes as pertaining to the AD-
hypometabolic pattern (Fig. 4). However, the expert
interprets frontal hypometabolism as potentially due
to AD only if temporoparietal hypometabolism can be
concomitantly visualized, an operation not performed
by the PALZ score computation. The peculiar results
obtained in these seven patients highlight another prob-
lem, the statistical maps obtained at individual level
through PMOD or by other automatic systems, such as
SPM, that need eventually to be interpreted.

It is worth noting that 12 patients converted to FTLD
or DLB, and FDG-PET visual reading was consistent
with the diagnosis in all but one patient. This result
is in keeping with the notion that, although aMCI
patients are more prone to convert to AD demen-
tia, they can develop other types of dementia [47].
This finding underlines that the adequate knowledge of
hypometabolic patterns associated with different types
of dementia (obtained either visually or by means of
a software-generated map) is mandatory for the final
interpretation of scans [48]. This subgroup of patients
further proves the unbroken role of FDG-PET in aMCI
patients even in the era of amyloid PET. By exploring
the topographical distribution of hypometabolism in
a large spectrum of diseases, FDG-PET can improve
diagnostic accuracy at a time when the clinician needs
to know whether his/her patient with aMCI is actually
affected by a neurodegenerative disease by identifying
a specific hypometabolic pattern [2].

Some limitations of this study should be acknowl-
edged. These aMCI patients were recruited at
specialized centers and therefore conclusions may not
be generalized in the whole MCI population. In fact, the
conversion rate to dementia was high (26% per year)
but it has been shown that patients with aMCI have a
higher conversion rate than the general population of
MCI, including patients with non-amnestic MCI [49].
We do not have histopathological verification of the
clinical diagnoses, or demonstration of brain amylo-
dosis by means of PET imaging or decreased Aβ1–42
in cerebrospinal fluid (available only in a small sub-
group). Although we considered only those patients
with at least two years of follow-up (unless they con-
verted to dementia before), the follow-up time was not
homogeneous. Finally, data were acquired by different
PET scanners although we tried to minimize this het-
rogeney by including balanced numbers of patients
and controls from each center.

In conclusion, this study has shown that FDG-PET
is an accurate biomarker of AD-type neurodegenera-
tion in a large series of aMCI patients later converting
to AD dementia and in healthy controls. Even more
important, it can pick up the presence or absence of
an AD phenotype in a subgroup of aMCI patients not
yet converted to dementia after a mean follow up of
more than three years. Visual reads by an expert is the
most accurate method but an automated, validated sys-
tem may be particularly helpful to moderately-skilled
readers especially because of its high specificity, and
should be mandatory when even a moderately-skilled
reader is unavailable.

ACKNOWLEDGMENTS

There has been no financial or material support for
this research.

Authors’ disclosures available online (http://www.j-

REFERENCES

(2011) Review and meta-analysis of biomarkers and diag-


S. Morelli et al. / Comparative Analysis of FDG-PET in AβcI