Autopsy as Gold Standard in FDG-PET Studies in Dementia

Pascali Durand-Martel, Dominic Tremblay, Catherine Brodeur, Nancy Paquet

ABSTRACT: Positron emission tomography (PET) imaging with F18-fluorodeoxyglucose (FDG) is increasingly used as an adjunct to clinical evaluation in the diagnosis of dementia. Considering that most FDG-PET studies in dementia use clinical diagnosis as gold standard and that clinical diagnosis is approximately 80% sensitive or accurate, we aim to review the evidence-based data on the diagnostic accuracy of brain FDG-PET in dementia when cerebral autopsy is used as gold standard. We searched the PubMed and Medline databases for dementia-related articles that correlate histopathological diagnosis at autopsy with FDG-PET imaging and found 47 articles among which there were only 5 studies of 20 patients or more. We were able to conclude that sensitivity and specificity of FDG-PET for Alzheimer's disease are good, but more studies using histopathological diagnosis at autopsy as gold standard are needed in order to evaluate what FDG-PET truly adds to premortem diagnostic accuracy in dementia.

Cognitive impairment is an area of great concern in the media and for clinicians around the world. In the past 20 years, many articles have been written about positron emission tomography (PET) imaging with F18-fluorodeoxyglucose (FDG), either to investigate its qualities as a diagnostic tool or, more recently, as inclusion criteria in clinical trials1. However, the routine use of FDG-PET in the investigation of dementia is not recommended in the guidelines published by the American Academy of Neurology2 or by the Third Canadian Consensus Conference on Dementia3. Several publications have studied the ability of FDG-PET to differentiate Alzheimer’s disease (AD) from other dementia4, most of them using clinical diagnosis as gold standard. However, to this day, clinical diagnosis, even when combined with extensive neuropsychological testing, is neither sensitive nor specific enough to be considered a gold standard in dementia. In studies of clinicopathological correlation, the sensitivity of clinical diagnosis for AD varies between 75 and 98%, with an average of 82% when adjusted for the number of patients per study5-9. In clinicopathological correlation studies of 170 or more patients with a clinical diagnosis of AD the accuracy varies between 78% and 88%, with an adjusted average of 81%.10-12 In 2006, Forman et al identified 21% of AD pathology in a series of 114 cases presenting clinically as frontotemporal dementia13. In other words, up to one out of five patients participating in FDG-PET studies could, in fact, present a pathology that is different from the one that was clinically...
diagnosed. We believe that FDG-PET is a useful adjunctive diagnostic tool, but in order to evaluate what this exam truly adds to the accuracy of premortem diagnosis of dementing illnesses, we need to review the studies that use brain autopsy as gold standard. In this article, we present a systematic review of the correlations between FDG-PET and autopsy in literature.

**METHODS**

Methods and results of our literature review are summarized in the Figure. On June 24th 2009 we searched the Pubmed database for relevant papers. First, we combined the entry “FDG” with “dementia”, and then, in turn, with “Lewy Body disease”, “frontotemporal lobar degeneration”, “frontotemporal dementia”, “Alzheimer”, “Pick’s disease”, “progressive supranuclear palsy”, “corticobasal degeneration”, “semantic dementia”, “primary progressive aphasia”, “Parkinson’s disease dementia”, “vascular dementia” and “Creutzfeldt-Jakob”. We repeated the same searches replacing “FDG” by “positron emission tomography”. Titles and abstracts for each result were scanned for our inclusion criteria: presence and correlation of FDG-PET imaging and cerebral autopsy for the same patient(s). We excluded works if they: were written in languages other than English or French, reported the results of animal researches, did not used FDG as a tracer for PET, studied only mild cognitive impairment (MCI) or used clinical diagnosis, genetic testing or histopathological data from a brain biopsy as gold standard. Brain biopsy was excluded as a gold standard because it can overlook mild or mixed pathologies. We also excluded genetic testing as gold standard because most mutations known to be associated with dementia have important phenotypical variability and are rare occurrences. We also searched the Medline database for articles published between 1950 and June 2009.

**Figure**: Literature review methodology and results. FDG = Fluorodeoxyglucose, PET = Positron Emission Tomography, AD = Alzheimer’s disease, CJD = Creutzfeldt-Jakob disease, DLBD = Diffuse Lewy Body disease, FTLD = frontotemporal lobar degeneration, FTD = frontotemporal dementia, PSP = progressive supranuclear palsy, CBD = corticobasal degeneration, SD = semantic dementia, PPA = primary progressive aphasia, VaD = vascular dementia, PDD = Parkinson’s disease related dementia.
24th 2009 corresponding to either “Positron Emission Tomography” or “Fluorodeoxyglucose F18” and either “Dementia”, “Pick Disease of the Brain”, “Creutzfeldt-Jakob Syndrome”, “Lewy Bodies”, “Lewy Body disease”, “Alzheimer Disease”, “Vascular Dementia”, “Multi-Infarct Dementia”, “Progressive Supranuclear Palsy”, “Primary Progressive Aphasia”, “corticobasal degeneration” (as keyword only (kwd)), “semantic dementia” (kwd), “Parkinson’s disease dementia” (kwd), “frontotemporal lobar degeneration” (kwd) or “frontotemporal dementia” (kwd). We restricted this search to articles referring to humans that were written in English or in French and we scanned the results for the inclusion and exclusion criteria described above. Several studies and case series evaluated FDG-PET as well as other imaging techniques or included several patients for which the final diagnosis had been decided over different criteria. These articles were included in our study but we only considered the patients for which there had been both a FDG-PET and a cerebral autopsy.

**Results**

A total of 3696 results were obtained from PubMed and their titles and abstracts were scanned. Among these results, we initially selected 151 FDG-PET-related publications. After a complete reading of these articles and their references, 43 articles meeting our criteria remained. Four additional papers were found through the references of the initial 151 articles selection. The Medline search provided us with 797 results that were also scanned for the inclusion and exclusion criteria listed in the previous section. No additional publication was found through this database. The final number of articles included in this review is 47.

Among these 47 articles, the overwhelming majority are case reports or series. Considering only the reported patients for which there has been both a FDG-PET and a cerebral autopsy, there are 26 single case reports, 16 case series that include between 4 and 11 patients, and only 5 studies of 20 patients or more. The Table summarizes these five studies and their results. The pathological diagnoses of the case reports are diverse. There are eight cases of prion diseases14-21, five isolated AD22-26, four frontotemporal lobar degeneration (FTLD)27-30, four pure diffuse Lewy body disease (DLBD)31-34, two mixed AD and DLBD35-36, two progressive supranuclear palsy37-38 and one Parkinson’s disease39. In contrast, most of the case series refer to patients with a pathological diagnosis of AD40-47 or to a group of patients presenting either AD or another diagnosis such as DLBD48 or FTLD49. Among the remaining series, three present Creutzfeldt-Jakob disease cases50-52, two present progressive supranuclear palsy patients53-54 and one presents patients with frontotemporal lobar degeneration linked to chromosome 17 mutations55. One of the case series that reports Creutzfeldt-Jakob disease cases also presents a case of metachromatic leukodystrophy52. Two neuropathologically proven cases of Pick’s disease and five of AD are also included in the series but we were unable to know whether a biopsy or an autopsy had occurred in all but one case of AD for which there had been a biopsy. Fortunately, one of the larger autopsy-only studies includes some patients from this series56.

Four out of the five larger studies included between 20 and 45 patients. Hoffman et al evaluated the accuracy, sensitivity and specificity of bilaterally decreased glucose metabolism in the temporal and parietal lobes for Alzheimer’s disease in a group of 20 patients recruited on the basis of a challenging or difficult clinical diagnosis57. In this paper, the analysis also included two patients whose diagnosis had been made with cerebral biopsy. The premortem clinical diagnoses are included in the publication. While, in this study, PET is more sensitive for AD than clinical diagnosis of probable and possible AD with the criteria of both the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer’s Disease and Related Disorders Association (NINDS-ADRDA) and the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) (92.9% vs. 78.6%), its accuracy is equal (81.8% vs. 81.8%). In 2001, Minoshima et al studied the value of occipital hypometabolism (Z score ≤-2.4) in differentiating DLBD from AD pathology. The sensitivity and specificity were 90% and 80%, respectively. They did a similar analysis with a group of clinically diagnosed AD among which some patients had developed clinical criteria for DLBD over time. F18-fluorodeoxyglucose PET had been obtained before the patients presented symptoms suggestive of DLBD. The specificity was greater and the sensitivity was lower in this clinically diagnosed group, 95% and 38%, respectively, with a Z score value of ≤-2.5.

The only two studies we found that evaluated the effect on diagnostic accuracy of adding an FDG-PET to the clinical evaluation are those of Jagust et al and Foster et al, both published in 200759,60. The former classified patients as having AD or non-AD pathology on both PET and cerebral autopsy. In this study, the clinical evaluation made an average of 4.9 years before death was 76% sensitive and 58% specific for AD pathology, while the FDG-PET, obtained on average 3.6 years before death, was 84% sensitive and 74% specific. The likelihood of detecting AD pathology on the basis of the initial evaluation increased from 70% to 84% with the addition of a positive FDG-PET scan, and decreased to 31% with a negative scan. Foster et al evaluated six neurologists’ interpretations of FDG-PET scans, as well as their clinical diagnoses made from clinical summaries and checklists created retrospectively from the files of 45 patients60. Thirty-one patients had a pathological diagnosis of AD while the remaining 14 had FTLD. The accuracy of FDG-PET was significantly better than that of clinical diagnosis (p<0.02). There was an improvement of overall diagnostic accuracy from 79% to 90% (p=0.03) when raters added the PET results to their evaluation of clinical scenarios. In subgroup analysis however, this improvement was only statistically significant for FTLD patients (p=0.01), not for AD patients (p=0.3). The addition of PET was significantly more likely to have a beneficial rather than adverse effect on diagnostic accuracy (p=0.0001). It also increased significantly the diagnostic confidence in cases where the raters were uncertain about their initial diagnosis (p=0.003).

The largest PET-autopsy correlation study is that of Silverman et al, which includes 138 patients who underwent both FDG-PET and cerebral autopsy at eight different centers in the United States of America, Belgium and Germany. The interpreters visually rated and classified the images according to several patterns of cerebral metabolism suggestive of either progressive or non progressive clinical course, and more specifically of AD. The premortem clinical diagnoses of these
patients are unfortunately not discussed in the article, but information about the severity of cognitive impairment at the moment of FDG-PET is available. This allowed the authors to separately analyze a subgroup of 55 patients with mild or questionable dementia at the time of the PET scan, giving 89% accuracy, 95% sensitivity and 71% specificity.

The five main studies used different pathological criteria for Alzheimer’s disease but all the criteria used are well recognised and validated. None of the five main studies mentioned whether the pathologists were blind to the results of the FDG-PET studies. As for PET reading, the techniques were also varied. In Minoshima et al, mean regional reduction in metabolism is measured in percentages for each group (AD, DLBD and Lewy Body variant AD), and Z scores were chosen in order to optimize sensibility and sensitivity for DLBD in the four other papers, exams were either interpreted by a blinded nuclear medicine specialist who agreed with the initial diagnostic classification in all but one case. In the two remaining papers, exams were either interpreted by a single physician or by two investigators who had to come to a consensus. We did not find any reported case or series with pathological confirmation of diagnosis where automated analysis of FDG-PET, as with the method developed by Heiss, in Herholz et al, was used. The use of those automated techniques would limit inter-observer variability although it is presently unclear whether it would improve the overall sensitivity and specificity for AD.

**DISCUSSION**

The purpose of our review was to verify whether studies using autopsy as gold standard have proven that FDG-PET improves the premortem diagnostic accuracy of dementia. The five larger studies we found fulfilling our criteria provide interesting information on the accuracy of FDG-PET diagnosis from a total of 268 patients. The results for accuracy, sensitivity and specificity for AD in these papers are in fact similar to those obtained in FDG-PET studies that use clinical diagnosis as gold standard. This finding is reassuring because Alzheimer’s disease

<table>
<thead>
<tr>
<th>Reference</th>
<th>Patients</th>
<th>Methods</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoffman, 2000</td>
<td>9 AD only</td>
<td>Prospective, monocentric</td>
<td>Ac for AD only : 81.8%</td>
</tr>
<tr>
<td></td>
<td>1 AD + DLBD</td>
<td>PET to autopsy : 2.1 yrs average</td>
<td>Se for AD only : 92.9%</td>
</tr>
<tr>
<td></td>
<td>1 AD + PSP</td>
<td>PET : transaxial</td>
<td>Sp for AD only : 62.5%</td>
</tr>
<tr>
<td></td>
<td>9 others, not AD</td>
<td>Pathology : CERAD</td>
<td>Ac for AD + other pathology : 81.8%</td>
</tr>
<tr>
<td></td>
<td>MMSE : unspecified</td>
<td></td>
<td>Se for AD + other pathology : 87.5%</td>
</tr>
<tr>
<td>Minoshima, 2001</td>
<td>10 AD</td>
<td>Retrospective, monocentric</td>
<td>Ac for DLBD : N/A</td>
</tr>
<tr>
<td></td>
<td>4 DLBD</td>
<td>PET to autopsy : 3.2 yrs average</td>
<td>Se for DLBD : 90%</td>
</tr>
<tr>
<td></td>
<td>7 AD + DLBD</td>
<td>PET : 3D-SSP. Z score for visual cortex</td>
<td>Sp for DLBD : 80%</td>
</tr>
<tr>
<td></td>
<td>MMSE : 14*</td>
<td>Pathology : Khachaturian, Lewy bodies†</td>
<td></td>
</tr>
<tr>
<td>Jagust, 2007</td>
<td>25 AD</td>
<td>Retrospective, monocentric</td>
<td>Ac : N/A</td>
</tr>
<tr>
<td></td>
<td>19 non AD</td>
<td>PET to autopsy : 3.6 yrs average</td>
<td>Se for AD : 84%</td>
</tr>
<tr>
<td></td>
<td>MMSE : 23 +/- 5.7</td>
<td>PET : axial, sagittal and coronal</td>
<td>Sp for AD : 74%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pathology : CERAD and NIA-Reagan</td>
<td>PPV for AD : 81%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NPV for AD : 78%</td>
</tr>
<tr>
<td>Foster, 2007</td>
<td>31 AD</td>
<td>Retrospective, monocentric</td>
<td>Ac of SSP : 89.2%</td>
</tr>
<tr>
<td></td>
<td>14 FTLD</td>
<td>PET to autopsy : N/A</td>
<td>Se of SSP for AD : 97.6%</td>
</tr>
<tr>
<td></td>
<td>MMSE : 14 +/- 8.8</td>
<td>PET : transaxial and SSP</td>
<td>Sp of SSP for AD : 73.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pathology : NIA-Reagan, FTLD</td>
<td>PPV of SSP for AD : 89%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NPV of SSP for AD : 93%</td>
</tr>
<tr>
<td>Silverman, 2001</td>
<td>97 AD +/- other pathology</td>
<td>Multicentric</td>
<td>Ac for AD : 88%</td>
</tr>
<tr>
<td></td>
<td>23 other pathologies</td>
<td>PET to autopsy : 2.9 yrs average</td>
<td>Se for AD : 94%</td>
</tr>
<tr>
<td></td>
<td>18 no pathology</td>
<td>PET : axial and coronal</td>
<td>Sp for AD : 73%</td>
</tr>
<tr>
<td></td>
<td>MMSE : 24 +/- 6 (N = 85)</td>
<td>Pathology : Varied, unspecified</td>
<td>Se for any degenerative disease : 94%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sp for any degenerative disease : 78%</td>
</tr>
</tbody>
</table>

AD = Alzheimer’s disease, DLBD = Diffuse Lewy Body disease, PSP = progressive supranuclear palsy, MMSE = Minimal Status Examination, FTLD = frontotemporal lobar degeneration, Ac = accuracy, Se = sensitivity, Sp = specificity, CERAD = Consortium to Establish a Registry for Alzheimer’s disease, NIA = National Institute on Aging, N/A = Not available, PPV = positive predictive value, NPV = negative predictive value, 3D = three dimensional, SSP = stereotactic surface projection, *Adjusted mean calculated by the authors of this review, †3 Lewy bodies per x20 field in 4 fields in 3 of transentorhinal cortex, anterior, cingulate cortex, amygdala, and insular cortex.
is the most common cause of dementia in developed countries. Even though there currently is no pharmacological treatment that can reverse or stop biological progression of AD, acetylcholinesterase inhibitors as well as N-methyl D-aspartate (NMDA) antagonists provide symptomatic relief, and hence improve quality of life while decreasing caregiver burden. Diagnosis of DLBD and FTLD by PET appears accurate and sensitive, but since we found only 11 and 14 cases, respectively, in large studies we cannot make valid conclusions. Through this review, we also realised that there is little information available on the usefulness of FDG-PET for diseases other than AD, even in articles using clinical diagnosis as gold standard.

Unfortunately, only two out of the five studies that we found analysed the effect of combining FDG-PET with clinical evaluation. Hoffman et al and Silverman et al studied the diagnostic accuracy for AD in subjects with different pathological diagnosis but, since the PET results were dichotomized as AD or non-AD, or progressive vs non progressive, calculation of the overall accuracy is not possible. We have not found any study that evaluated prospectively the accuracy of FDG-PET for differentiating the various dementia subtypes. While Alzheimer’s disease is part of the differential diagnosis in most cases, AD vs. non-AD is not the only dilemma that clinicians interested in cognitive impairment face in their practice. The clinical and pathological diagnoses were also dichotomized in the evaluations of the adjunctive effect of PET on diagnostic accuracy by Jagust et al and by Foster et al. This limits the application of their results to our clinical practice. Additionally, in Foster et al, the FDG-PET studies were interpreted by neurologists instead of nuclear medicine specialists or radiologists and the examiners could only interpret these studies as suggestive of either AD or FTLD. We believe that these features, as well as the use of clinical vignettes that include even late presenting symptoms do not represent clinicians’ reality.

In Jagust et al, initial clinical diagnosis was proven to be inferior to the latest diagnosis in life after a long-term follow-up. Conversely, Silverman et al noted a slight superiority of FDG-PET diagnosis in a subgroup of only mildly affected individuals. This would argue in favour of adding FDG-PET exams to the initial evaluation of patients presenting with dementia in order to achieve better accuracy earlier in the course of the disease. In a case series by Mosconi et al, four normal elderly were followed clinically and with FDG-PET to the onset of clinically probable AD in two cases and amnesic MCI in the two others. Alzheimer’s disease changes were present at autopsy for these four patients and preclinical as well as early changes on the FDG-PET had been detected in all of them. Several papers report cohorts of MCI patients in which FDG-PET is a good indicator of progression towards clinically probable AD. However, in order to truly identify and measure the benefits of FDG-PET early in the clinical course, there needs to be a multi-center and, ideally, prospective autopsy correlation study of a large population with varied pathological diagnoses in which the collected data include both initial and follow-up standardised clinical diagnoses, pharmacological treatment information, as well as FDG-PET results that are categorized into dementia subtypes instead of dichotomized. If such a study brought positive results, clinicians would have to take into consideration the cost of the exam, which varies from one center to another. As an example, the estimated cost of an FDG-PET in Quebec was below 1000 $ in 2001.

Our review has some limitations. First, we did not consider works that were either published in languages other than English or French, later than June 24th 2009, or in sources not accounted for in PubMed and Medline. Second, as in most literature review, we present studies with varied samples and methods. In these studies, both FDG-PET and autopsy have been made and interpreted by many different techniques and investigators, limiting our ability to compare their results. One might argue that in reviewing only FDG-PET studies, we overlooked the contribution of single photon emission computed tomography (SPECT) studies in dementia. However, the accuracy of PET has been proven to be superior to that of SPECT. We also overlooked amyloid, dopaminergic and other PET tracers, but these are not currently available to most clinicians in Canada. It is also possible to use visual or automated rating of hippocampal atrophy on magnetic resonance imaging (MRI) in the evaluation of patients with possible Alzheimer’s disease. Advantages of this technique are greater availability of MRI compared to PET and the possibility to evaluate combined vascular lesions. The principal disadvantage is that hippocampal volume decrease occurs after the diminution of glucose metabolism which might make a single exam less useful than FDG-PET in MCI and early dementia. In one extensive literature review, the sensitivity of this technique varied between 35 and 95% (with most values between 76 and 95%) and the specificity between 76 and 94%, compared to clinical diagnosis. These results seem comparable to FDG-PET. However, while there is fair evidence to support selective use of structural MRI or computed tomography in the evaluation of dementia patients, the actual usefulness of morphometric MRI implies serial measurements in order to track the progression of AD in clinical trials (Grade B, Level 2). Finally, we are confident that our search was thorough and that our inclusion and exclusion criteria allowed us to select only those FDG-PET and dementia studies that were clinically significant and used the appropriate gold standard.

Among the excluded articles, two are worth mentioning. We excluded an article by Tedeschi et al in which all the histological diagnoses were obtained through frontal biopsy. Even though we believe that a single diagnosis made from a biopsy is reliable, it is likely that mixed pathologies would be missed by these procedures, and grading is unreliable. Finally, we excluded a study of medical statistics published in 2004 that was based on data from a study with a principal investigator, population, methods and results entirely identical to Foster et al. This article provided us with no additional information that was related to our objective and we believed it was most likely from the same sample as Foster et al, even though it was published earlier.

**Conclusion**

In conclusion, there are few studies that compare FDG-PET results with autopsy diagnoses in dementia. The studies we found show good accuracy and sensitivity for Alzheimer’s disease pathology, including in the early clinical stages when it is the most clinically valuable. This argues in favour of adding FDG-PET exams to the initial evaluation of patients presenting with atypical dementia in order to achieve better accuracy earlier.
in the course of the disease, and hence provide them the most appropriate treatment. The guidelines published by the Third Canadian Consensus Conference on Dementia in 2007 state that “there is fair evidence that functional imaging with PET or SPECT scanning might assist specialists in the differential diagnosis of dementia, particularly those with questionable early stage dementia or those with FTD (Grade B, Level 2)”. However, only one study investigates the accuracy and sensitivity for each DLBD and FTLD. No PET-autopsy study includes more than two major clinical subtypes of dementia in order to evaluate the overall accuracy of FDG-PET in a group of patients with many different clinical diagnoses. Even if it is reassuring that the results for AD in these papers are similar to those obtained from studies using clinical diagnosis as gold standard, we believe further studies with autopsy confirmation of diagnosis are needed in the field of PET and dementia.

REFERENCES