Diagnostic performance of an automated analysis software for the diagnosis of Alzheimer’s dementia with $^{18}$F FDG PET

Sasan Partovi$^1$, Roger Yuh$^1$, Sara Pirozzi$^2$, Ziang Lu$^1$, Spencer Couturier$^1$, Ulrich Grosse$^1$, Mark D Schluchter$^3$, Aaron Nelson$^2$, Robert Jones$^1$, James K O’Donnell$^1$, Peter Faulhaber$^1$

$^1$Department of Radiology, University Hospitals Cleveland Medical Center, Case Western Reserve University, Cleveland, Ohio; $^2$MIM Software Inc., Cleveland, Ohio; $^3$Department of Biostatistics and Epidemiology, Case Western Reserve University, Cleveland, Ohio

Received September 16, 2016; Accepted November 27, 2016; Epub January 15, 2017; Published January 30, 2017

Abstract: The objective of this study was to assess the ability of a quantitative software-aided approach to improve the diagnostic accuracy of $^{18}$F FDG PET for Alzheimer’s dementia over visual analysis alone. Twenty normal subjects (M:F-12:8; mean age 80.6 years) and twenty mild AD subjects (M:F-12:8; mean age 70.6 years) with $^{18}$F FDG PET scans were obtained from the ADNI database. Three blinded readers interpreted these PET images first using a visual qualitative approach and then using a quantitative software-aided approach. Images were classified on two five-point scales based on normal/abnormal (1-definitely normal; 5-definitely abnormal) and presence of AD (1-definitely not AD; 5-definitely AD). Diagnostic sensitivity, specificity, and accuracy for both approaches were compared based on the aforementioned scales. The sensitivity, specificity, and accuracy for the normal vs. abnormal readings of all readers combined were higher when comparing the software-aided vs. visual approach (sensitivity 0.93 vs. 0.83 $P = 0.0466$; specificity 0.85 vs. 0.60 $P = 0.0005$; accuracy 0.89 vs. 0.72 $P<0.0001$). The specificity and accuracy for absence vs. presence of AD of all readers combined were higher when comparing the software-aided vs. visual approach (specificity 0.90 vs. 0.70 $P = 0.0008$; accuracy 0.81 vs. 0.72 $P = 0.0356$). Sensitivities of the software-aided and visual approaches did not differ significantly (0.72 vs. 0.73 $P = 0.74$). The quantitative software-aided approach appears to improve the performance of $^{18}$F FDG PET for the diagnosis of mild AD. It may be helpful for experienced $^{18}$F FDG PET readers analyzing challenging cases.

Keywords: PET, Alzheimer’s dementia, quantitative, software aided

Introduction

The prevalence of Alzheimer’s disease (AD) is continuously increasing in the United States with an expected doubling of the annual number of incident cases in 2050 [1]. Symptoms of the disease include cognitive impairment including memory dysfunction [2]. The diagnosis is most frequently made clinically based on neuropsychological testing, in particular with the Mini Mental Status Examination (MMSE) [3, 4]. However, the MMSE can be of limited value in early forms of dementia and mild cognitive impairment (MCI) [5, 6]. Imaging in the workup of AD is evolving with several neuroimaging modalities applied in clinical settings, including positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetic resonance imaging (MRI) and computed tomography (CT). PET with attenuation correction by low dose CT has been found to be a highly useful imaging modality for the diagnosis of neurodegenerative disorders [7]. According to the Center for Medicare and Medicaid Services, PET imaging with 2-deoxy-2-[$^{18}$F]fluoro-D-glucose (FDG) is considered an important step in the workup of a patient diagnosed with dementia who is suffering from at least 6 months of documented cognitive decline, has previously been evaluated for alternative degenerative diseases, and met the criteria for both AD and/or frontotemporal dementia (FTD) [8]. FDG as a radiotracer is reflecting glucose metabolism [9] and thus enables depiction of reduced activity of glucose metabolism in affected brain areas of AD [10, 11]. Pattern of abnormalities of cerebral glucose...
metabolism in AD can involve the parieto-temporal, posterior cingulate, precuneus and frontal cortex [2, 12-18]. Sparing of the sensorimotor cortex, subcortical gray matter, visual cortex, basal ganglia, cerebellum, and thalami is a distinguishing feature of AD as opposed to other forms of dementia [8]. FDG-PET revealed sensitivity above 90% and specificity above 70% for the diagnosis of AD while correctly identifying the presence of AD in 88% of patients when compared with histopathology [17]. A recently published meta-analysis pooled data from 27 studies and determined a sensitivity of 90% and a specificity of 89% for the diagnosis of AD against non-demented healthy controls with FDG-PET [19]. Interpretation of FDG-PET in the setting AD is usually in a manual approach using visual qualitative reading. A manual approach can be prone to interpretation errors and strongly depends on the experience and training of the reading physician [20].

FDG-PET with a quantitative component using software-aided analysis may increase diagnostic performance for AD detection. Previous studies for quantitative AD diagnosis with FDG-PET applied voxel-based procedures with age-adjusted t-statistics, statistical parametric mapping approaches and standardized stereotactic surface projections (SSPs) [21-23]. The purpose of this study was to determine whether a commercial quantitative software tool, MIMneuro™ (MIM Software Inc. Cleveland, OH) could improve the diagnosis of AD compared to qualitative visual analysis alone for three experienced board certified radiology and nuclear medicine physicians (25 years, 12 years and 6 years of experience).

**Subjects and methods**

**ADNI (Alzheimer’s disease neuroimaging initiative)**

Forty subjects with FDG PET brain scans were obtained from the ADNI database (adni.loni.usc.edu). All subjects included within the ADNI database gave written informed consent to participate in the study. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). For up-to-date information, see www.adni-info.org. In addition to FDG PET scans, all ADNI subjects received baseline comprehensive neuropsychological evaluation, laboratory tests and blood samples for genetic analysis, and structural MR exams [24].

**Subject selection**

Twenty normal controls and twenty AD subjects were chosen from the ADNI database. The normal control patients (12 male, 8 female) ranged in age from 71 to 94, with a mean age ± SD of 80.6±7.2 y, and fulfilled all eligibility criteria for enrollment as an ADNI normal control including a Clinical Dementia Rating (CDR) score of 0, a MMSE score between 24-30, and no diagnosis of depression, MCI, or dementia. Normal controls included in this study also had a negative PET amyloid scan and did not progress to MCI or AD within 24 month follow up. The mild AD subjects (12 male, 8 female) used in this study were chosen from the ADNI database. The normal control patients (12 male, 8 female) ranged in age from 71 to 94, with a mean age ± SD of 80.6±7.2 y, and fulfilled all eligibility criteria for enrollment as an ADNI normal control including a Clinical Dementia Rating (CDR) score of 0, a MMSE score between 24-30, and no diagnosis of depression, MCI, or dementia. Normal controls included in this study also had a negative PET amyloid scan and did not progress to MCI or AD within 24 month follow up. The mild AD subjects (12 male, 8 female) used in this study were chosen from the ADNI database. The normal control patients (12 male, 8 female) ranged in age from 71 to 94, with a mean age ± SD of 80.6±7.2 y, and fulfilled all eligibility criteria for enrollment as an ADNI normal control including a Clinical Dementia Rating (CDR) score of 0, a MMSE score between 24-30, and no diagnosis of depression, MCI, or dementia. Normal controls included in this study also had a negative PET amyloid scan and did not progress to MCI or AD within 24 month follow up.

<table>
<thead>
<tr>
<th>Number of subjects</th>
<th>Healthy controls</th>
<th>Alzheimer disease patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (%)</td>
<td>12 (60)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Mean age ± standard deviation</td>
<td>80.6±7.2*</td>
<td>70.6±10.4*</td>
</tr>
<tr>
<td>Mini-Mental State Examination (MMSE) ± standard deviation</td>
<td>29±1.4</td>
<td>22.8±1.9</td>
</tr>
<tr>
<td>Clinical Dementia Rating (CDR)</td>
<td>0</td>
<td>0.5/1</td>
</tr>
</tbody>
</table>

*Mean age differed significantly between healthy controls and Alzheimer disease patients (P = 0.005).
shown in Table 1. All subjects were clinically followed and imaged at 6, 12, and 24 months.

\textit{\textsuperscript{18}F-FDG PET imaging data}

All FDG PET scans were acquired between 2007 and 2011 under common standardized FDG PET acquisition protocols consistent with ADNI standards. Data were corrected for scatter and radiation-attenuation using transmission scans for systems having rod sources, or by CT scan for sites with a PET/CT scanner. Data were reconstructed using measured-attenuation correction and image reconstruction algorithms specified for each scanner (http://adni.loni.usc.edu). Raw PET data in DICOM format were first uploaded to the University of Michigan for pre-processing to correct for differences across PET scanners used at various ADNI sites. During preproces-
ing, first each of the five-minute emission frames acquired in every FDG scan were co-registered and then averaged to the first frame. The co-registered, averaged image was then reoriented to a common spatial orientation, such that the anterior-posterior axis of the subject ran parallel to the anterior commissure-posterior commissure (AC-PC) line, and interpolated onto a uniform 60×160×96 voxel image grid, with 1.5 mm cubic voxels (http://adni.loni.usc.edu/methods/pet-analysis/pre-processing). Finally a subject-specific mask was applied for intensity normalization (where average in the mask was one). Further details regarding ADNI image acquisition and processing are described in Jagust et al [24].

Visual image interpretation

Pre-processed ADNI images (co-registered, averaged, standardized image and voxel size) were downloaded from the ADNI LONI database in DICOM format (https://ida.loni.usc.edu). Three observers who were blinded to the diagnoses were selected as participants in this study. All three readers were highly experienced US board certified nuclear medicine and radiology physicians (25 years, 12 years and 6 years of experience). Before image interpretation, a 30-minute training session was provided to review the criteria for defining typical metabolic abnormalities for AD on FDG PET brain scans. The following criteria were used for a visually positive diagnosis of AD: hypometabolism in the temporal and parietal lobes, the posterior cingulate gyrus and the precuneus, with relative sparing of the sensorimotor and primary visual cortices and cerebellum. The striatum and thalamus are also spared. Readers were reminded that in early stages the hypometabolism might appear asymmetric, however, as the disease progresses it will often become bilateral. Bilateral involvement is more common, however, there may still be some of the earlier asymmetry present. With more advanced disease frontal lobe deficits can be seen as well. Occasionally there may be frontal dominant disease where frontal deficits are more prominent than the parietotemporal deficits [12, 25-30].

Prior to actual interpretation sessions, FDG PET images of a healthy subject and an AD subject were shown in a standard three plane display as sample cases to highlight key differences between normal and AD brain scans (Figure 1). After the training session, PET images were presented to readers for visual interpretation in a conventional three plane display (axial, sagittal, and coronal) with a standard gray scale over the course of several weeks. Order of normal and AD images were randomized and presented by a non-observer who transcribed patient rankings into an excel spreadsheet. FDG PET images were visually classified using

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.90</td>
<td>0.75</td>
<td></td>
<td>0.85</td>
<td>0.45</td>
<td></td>
<td>0.875</td>
<td>0.600</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.80</td>
<td></td>
<td>0.95</td>
<td>0.75</td>
<td></td>
<td>0.975</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.90</td>
<td>0.95</td>
<td></td>
<td>0.75</td>
<td>0.60</td>
<td></td>
<td>0.825</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>0.933</td>
<td>0.833</td>
<td>0.047</td>
<td>0.85</td>
<td>0.60</td>
<td>0.0005</td>
<td>0.892</td>
<td>0.717</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.45</td>
<td>0.65</td>
<td></td>
<td>0.95</td>
<td>0.50</td>
<td></td>
<td>0.70</td>
<td>0.575</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.85</td>
<td>0.70</td>
<td></td>
<td>0.95</td>
<td>0.85</td>
<td></td>
<td>0.90</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.85</td>
<td>0.85</td>
<td></td>
<td>0.80</td>
<td>0.75</td>
<td></td>
<td>0.825</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>0.717</td>
<td>0.733</td>
<td>0.74</td>
<td>0.90</td>
<td>0.70</td>
<td>0.0008</td>
<td>0.808</td>
<td>0.717</td>
<td>0.0356</td>
</tr>
</tbody>
</table>
two rating scales: Normal/Abnormal and AD/No AD. Each reader rated each case on a scale from 1 = definitely normal to 5 = definitely abnormal, and on a scale from 1 = definitely not AD to 5 = definitely AD. For the normal/abnormal scale, a 1 was considered “definitely normal”, and a 2 was considered “probably normal”. A score of 4 or 5 corresponded to “probably abnormal” and “definitely abnormal”, respectively. For the AD/No AD scale rating, a 1 was considered “definitely not AD”, and a 2 was considered “probably not AD”. A score of 4 or 5 corresponded to “probably AD” and “definitely AD”, respectively. For both rating systems a score of 3 was considered equivocal, and counted as positive for both scales. The number of patients that were assigned to each score for Normal/Abnormal AD/No AD and can be observed in Table 6. Visual analysis was completed first for all subjects prior to beginning the quantitative analysis portion of the study.

Aided image interpretation

Following visual assessment, quantitative analysis was performed on the same set of 40 FDG PET scans. A non-observer presented images to the three readers who were again asked to rate each case with the additional aid of quantification results generated by MIMneuro. Readers rated the images using the same two scales used for visual assessment (Normal/Abnormal and AD/No AD). All FDG PET brains were first spatially normalized to a standard FDG PET brain template using a 9-parameter affine registration followed by a nonlinear landmark-based deformation. This registration allowed for comparisons to a normal database which had also been registered to the template. The normal database was comprised of 43 healthy controls between the ages of 41-80, with a mean age ± SD of 63.8±10 y. Both the normal comparison set and subject’s brains were normalized to the mean activity of the whole brain, pons, and cerebellum before comparison. Statistical comparisons were made between the subject and the normal database for each voxel in the brain. Z scores were calculated from the normalized values for voxel-based analysis using the following formula: 

\[ Z = \frac{\text{mean}_{\text{subject}} - \text{mean}_{\text{normals}}}{\text{SD}_{\text{normals}}} \]

Stereotactic surface projections (SSPs) were also utilized, providing an overview of cortical FDG uptake. SSPs are obtained by looking for the highest activity voxel along a vector perpendicular to the surface of the brain and projecting that voxel on to the surface [25]. Statistically significant differences were highlighted for all of the voxels in the brain on both the multi-planar reconstructed images and the SSPs. The differences were represented as z-scores or the number of standard deviations away from the mean. Voxels with z-scores < -1.65 were highlighted in a cool color scale on the subject’s registered brain (Figure 2). A z-score cutoff of -1.65 using a one-tailed t-test corresponds to statistically significant difference from normal at the 0.05 significance level. Processed quantitative results displaying voxel Z scores and SSPs normalized to each reference region were saved into a session, which allowed the software to be restored to a given state (including display layout and quantitative voxel and SSP results) for a consistent presentation to each reader. Cases were re-randomized for the second time to minimize recall bias and presented to readers in the saved session format with the processed quantitative results.

<table>
<thead>
<tr>
<th>Reader</th>
<th>Auto</th>
<th>Manual</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.896</td>
<td>0.735</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.994</td>
<td>0.821</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.895</td>
<td>0.833</td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>0.978</td>
<td>0.841</td>
<td>0.0154</td>
</tr>
</tbody>
</table>

Table 4. Area under the ROC curves for normal vs. abnormal readings when comparing software-aided vs. manual readings in all three readers and averaged over all readers

<table>
<thead>
<tr>
<th>Reader</th>
<th>Auto</th>
<th>Manual</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.800</td>
<td>0.626</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.956</td>
<td>0.821</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.890</td>
<td>0.884</td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>0.943</td>
<td>0.814</td>
<td>0.0264</td>
</tr>
</tbody>
</table>

Table 5. Area under the ROC curves for absence vs. presence of Alzheimer’s disease when comparing software-aided vs. manual readings in all three readers and averaged over all readers
Software-aided FDG-PET in Alzheimer’s dementia

Statistical analysis

Accuracy of visual and aided readings were compared for the Normal/Abnormal scale, and also for the AD/No AD scale. Receiver operating characteristic (ROC) analysis was performed to determine diagnostic accuracy. ROC curves were estimated by method (visual, aided) for each scale (Normal/Abnormal and AD/No AD) for each reader, and using the average rating across the three readers. Nonparametric estimates of ROC curve areas were estimated and compared between methods using a non-parametric approach, implemented in the Proc Logistic procedure of SAS version 9.4 (SAS Institute, Inc., Cary, NC). A generalized estimating equations logistic regression model was used to compare sensitivity, specificity, and accuracy of aided vs. visual methods using data from all three readers where readings of 3 or higher on the 5-point scale were considered positive/abnormal, and readings of 1-2 were considered negative/normal.

Results

Representative example FDG-PET scans acquired from the ADNI database are shown for an AD patient and healthy control in Figure 1. SSPs with overlaid z-scores for a healthy control and AD patient using MIMneuro is demonstrated in Figure 2. The distribution of scores for both the visual and software-aided approaches is listed in Table 6.

Sensitivity, specificity and accuracy for software-aided vs. manual readings

When comparing normal vs. abnormal readings, the specificity and accuracy observed for each of the three readers when using the software-aided approach were consistently higher for each of the three readers when using the software-aided approach. Observed sensitivity was higher in reader 2, equal in reader 3, and was lower in reader 1 when using the aided vs. the visual approach. When combining the results of all readers the specificity and accuracy were significantly increased (P = 0.0008 and P = 0.0356) when using the software-aided approach vs. visual inspection alone. Sensitivities of the software-aided and visual approaches did not differ significantly (0.72 vs. 0.73 P = 0.74). The results are listed in Table 2.

ROC curves for software-aided vs. manual readings

For the ROC analyses using combined data from all three readers, the area under the ROC curve was higher when using software-aided vs. visual readings, both for normal vs. abnormal readings (P = 0.0154), and for absence vs. presence of AD readings (P = 0.0264). The results of each reader and combined are listed in Table 4 (normal vs. abnormal readings) and Table 5 (absence vs. presence readings of Alzheimer’s disease). The ROC curve using the average of all three readers along with the ROC associated statistics comparing software-aided vs. visual readings is shown in Figure 3 for normal vs. abnormal readings and in Figure 4 for absence vs. presence readings of AD.

Discussion

In this study, we aimed to show the effect of a quantitative software-aided approach for the diagnosis of early stage AD with $^{18}$F FDG PET. The MIMneuro software-aided approach demonstrated an improvement of AD diagnostic specificity and accuracy while maintaining similar sensitivity compared to the qualitative visual...
Software-aided FDG-PET in Alzheimer’s dementia

Furthermore, the AUC ROC increased with statistical significance comparing the software-aided approach versus the manual assessment for the detection of both, abnormality and AD.

The subject selection in our study focused on patients with mild or questionable AD in the early disease stage. We wanted to focus on this population subtype for two reasons: 1) early diagnosis of AD is important as treatment is most effective during early stages of the disease [31] and 2) early stage AD is more challenging to diagnose compared to moderate to late stage AD [25].

Due to the increasing number of therapeutic options, the early and accurate diagnosis of AD is playing an increasingly important role [32,
Software-aided FDG-PET in Alzheimer’s dementia

33] as emphasized by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) [34, 35]. The current clinical climate stresses the importance of having an effective approach to diagnose AD non-invasively at early stages.

The clinical diagnosis of AD without the aid of imaging may be challenging. Depression, aphasia, and apraxia can mimic clinical criteria of AD [36, 37]. This coupled with the difficulty of distinguishing normal age related memory loss and rapid memory deterioration in early stage AD increases the complexity of AD diagnosis [38]. \(^{18}\)F FDG PET imaging has made advances in enabling the diagnosis of AD in patients with confounding mental conditions [39] and has shown good accuracy in the discrimination of probable AD patients from normal subjects (95% sensitivity and 100% specificity using voxel-based z-scores and stereotactic surface projections) [21].

The addition of a software aided approach alongside qualitative visual interpretation could significantly improve the current clinical diagnostic performance of early stage AD.

The use of quantitation to improve the diagnosis of AD using FDG PET has been studied previously. One study demonstrated that 3D-SSP could improve reader diagnostic accuracy in subjects with AD compared to standard transaxial section presentation of the FDG PET scan [25]. The benefit of 3D-SSP was attributed to readers only being presented with the surface of the brain, allowing them to interpret less information. This study included subjects with moderate to severe dementia (CDR 2.0 and 3.0) in addition to those with milder dementia (CDR of 0.5 or 1.0) which were used in our study. In another study it was shown that using regional cerebral blood flow reduction in posterior cingulate gyri and precunei on voxel-based 3D-SSP as diagnostic criteria [21] produced a diagnostic accuracy of greater than 85% for patients with mild AD (CDR score of 0.5) and that 3D-SSP was a feasible tool for aiding to the visual diagnosis of AD [26]. Contrary to our study, the authors used SPECT rather than \(^{18}\)F FDG PET. In a study looking at the effect of a voxel-based 3D-SSP approach for both beginners and experts over visual analysis alone it was found that that 3D-SSP significantly improved diagnostic specificity in both beginners (0.26 to 0.63) and experienced readers (0.56 to 0.87) without a significant decrease in sensitivity (0.83 to 0.82) [40]. This minimal decrease in sensitivity from the voxel based 3D-SSP approach corresponded to the change in sensitivity from the software-aided approach in our study (0.72 to 0.73 P = 0.74). In addition, the use of voxel-based 3D-SSP significantly improved the interpretative confidence for all readers compared to the visual approach.

Fully automated quantitative software approaches without reader input have shown improvements in terms of AD diagnostic accuracy compared to beginner readers [41, 42]. However, a previous study has shown that an automatic diagnostic system relying solely on voxel based parametric mapping without reader input produced significantly lower AD diagnostic sensitivity compared to experienced readers [43]. Nevertheless, one benefit of a software-aided approach is its ability to improve reader specificity, especially for readers who may overemphasize physiological minimal changes that are part of inter-subject variability [43]. Our study has shown that voxel based quantitative adjunct software improves early AD diagnostic specificity and accuracy while maintaining similar sensitivity for experienced readers, emphasizing the importance of a software aided approach with final assessment by reading physicians as opposed to a fully automatic approach.

Our study has several limitations including the confounding effect of senile brain atrophy, which mimics glucose hypometabolism seen in AD [44]. The minimal amount of metabolic and perfusion changes seen in both mild stages of late-onset AD and senile dementia may influence interpretation of PET images, especially for elderly patients [45-47]. The training session provided to readers before PET interpretation was aimed to reduce the confounding effect of brain atrophy, a challenge encountered during study readouts. In addition, our study included a limited sample of subjects, as only 20 subjects with mild AD were included along with 20 normal control subjects. Hence, the direct scope of our results and conclusion was limited to patients that exhibit mild, borderline AD. Nevertheless, the changes in diagnostic specificity and accuracy were statistically
significant despite the small sample size. Despite this limitation, the statistically significant results suggest that our software aided approach would be successful in aiding diagnosis of both moderate and severe AD. As AD progresses in severity the physiologic manifestations and hallmark metabolic changes become more pronounced. We hypothesize that software aided evaluation would continue to demonstrate statistically significant increases in diagnostic capability, though as the physiologic derangements become more pronounced the magnitude of effect may decrease compared to the values we achieved with mild AD. Furthermore, the diagnosis of AD for the study subjects was determined by clinical criteria rather than the gold standard histopathology. This may impact the diagnostic accuracy of both approaches in the study, as the true presence of AD in each subject was not determined. However, the long clinical follow-up and imaging at 6, 12, and 24 months strengthened our assessment of the clinical diagnosis as well as positive amyloid scans for the AD subjects and negative amyloid scans for the controls. In addition, reader 1 (our most experienced reader) had a twenty percent increase in sensitivity of presence/absence of AD when switching from software-aided to manual approach. The accuracy, sensitivity, and specificity of reader 1 with quantitative software guidance for normal/abnormal reading were very high (0.875, 0.90, and 0.85, respectively). Hence, it may be that this increase in sensitivity of AD diagnosis was due to the reader attributing atrophy related changes seen on quantitative analysis to causes other than AD. Furthermore, our study utilized two parameters, normal/abnormal and presence of AD/absence of AD, to evaluate each $^{18}$F FDG PET scan. This was done to address that the ADNI database has classified subjects based on clinical evaluation—which has been shown to potentially include subjects with glucose metabolism patterns more consistent with FTD or Lewy Body Dementia (LBD) as AD [48]. Although our subject selection criteria require a positive amyloid scan for mild AD subjects, LBD subjects may also present with a positive amyloid scan. Therefore, we used an additional parameter of normal vs. abnormal pattern of metabolism to include scans that indicate an abnormal pattern of metabolism without expressing a pattern consistent with AD. Finally, our study considered a score 3 on a 5 point scale to be indicative of an abnormal finding/presence of AD. This is an equivocal and subjective decision as it may also be interpreted as normal/absence of AD or an indeterminate outcome. However, scores of 3 out of 5 represented less than 10\% of total reads and thus did not play a role in determining the results of our study.

For future studies, additional quantitative metrics could be used including region-based analysis to assess characteristic regional hypometabolism, which could further aid diagnostic performance. Furthermore, a comparison of diagnostic effectiveness of voxel SSP and voxel transaxial section quantitative approaches could further highlight the effect of 3D projections of the brain surface, as previously mentioned [25]. Finally, our study indicated that quantitative software-aided improved diagnostic accuracy of abnormality and AD in comparison to the manual approach. At the same time, it would be of interest to further investigate the potential diagnostic benefit of a dual-read approach that incorporates a preliminary manual read followed by a software-aided diagnosis.

Conclusion

The quantitative software-aided approach appears to improve the performance of $^{18}$F FDG PET for the diagnosis of mild AD. The software-aided approach improves the diagnostic accuracy and specificity of mild AD while maintaining a similar level of sensitivity. Quantitative voxel based software may be helpful for experienced $^{18}$F FDG PET readers analyzing early onset AD.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.;
Disclosure of conflict of interest

This study was investigator initiated and supported by MIM Software Inc. Peter Faulhaber is consultant for MIM Software Inc. Peter Faulhaber is speaker for Philips Healthcare and has research grants from Philips Healthcare. Aaron Nelson has part ownership and is an employee of MIM Software Inc. Sara Pirozzi is an employee of MIM Software Inc.

Authors’ contribution

*Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpc-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Address correspondence to: Peter Faulhaber, Department of Radiology, University Hospitals Cleveland Medical Center, Case Western Reserve University, 11100 Euclid Avenue, 44106, Cleveland, OH, USA. Tel: 216-844-8140; Fax: 216-844-3106; E-mail: peter.faulhaber@uhhospitals.org

References

Software-aided FDG-PET in Alzheimer’s dementia


