Impact of expectation-maximization reconstruction iterations on the diagnosis of temporal lobe epilepsy with PET

John M Floberg¹, Aaron F Struck², Brooke K Peters³, Christine J Jaskowiak³, Scott B Perlman³, Lance T Hall³

¹Department of Medical Physics; ²Department of Neurology; ³Department of Radiology, University of Wisconsin-Madison, Madison, WI, USA

Received April 26, 2012; accepted May 17, 2012; Epub July 10, 2012; Published July 30, 2012

Abstract: There is a well known tradeoff between image noise and image sharpness that is dependent on the number of iterations performed in ordered subset expectation maximization (OSEM) reconstruction of PET data. We aim to evaluate the impact of this tradeoff on the sensitivity and specificity of ¹⁸F-FDG PET for the diagnosis of temporal lobe epilepsy. A retrospective blinded reader study was performed on two OSEM reconstructions, using either 2 or 5 iterations, of 32 ¹⁸F-FDG PET studies acquired at our institution for the diagnosis of temporal lobe epilepsy. The sensitivity and specificity of each reconstruction for identifying patients who were ultimately determined to be surgical candidates was assessed using an ROC analysis. The sensitivity of each reconstruction for identifying patients who showed clinical improvement following surgery was also assessed. Our results showed no significant difference between the two reconstructions studied for either the sensitivity and specificity of ¹⁸F-FDG PET for predicting surgical candidacy, or its sensitivity for predicting positive surgical outcomes. This implies that the number of iterations performed during OSEM reconstruction will have little impact on a reader based interpretation of ¹⁸F-FDG PET scans acquired for the diagnosis of temporal lobe epilepsy, and can be determined by physician and institutional preference.

Keywords: ¹⁸F-FDG PET, temporal lobe epilepsy, OSEM reconstruction, ROC analysis

Introduction

Positron emission tomography with 2-deoxy-2-(¹⁸F)fluoro-D-glucose (¹⁸F-FDG PET) has become an integral part in the diagnosis and presurgical evaluation of temporal lobe epilepsy (TLE), and its diagnostic efficacy and ability to predict surgical outcomes have been demonstrated in numerous studies [1-4]. TLE is diagnosed on ¹⁸F-FDG PET by identifying hypometabolic regions with decreased ¹⁸F-FDG uptake in a diseased temporal lobe relative to a healthy contralateral one. The diagnosis may be difficult to make if the degree of hypometabolism is subtle, and ¹⁸F-FDG uptake in small regions may be obscured if resolution is poor or noise variance is high. A number of approaches have been developed to improve diagnosis by reducing partial volume effects, for example by using anatomical information from MRI scans, to make small hypometabolic regions more detectable [5-7]. However, such correction methods are not readily available at most clinical centers and their use is only beginning to be validated.

A more clinically accessible approach to reduce partial volume effects is simply increasing the number of iterations performed during expectation-maximization (EM) reconstruction. EM reconstructions, particularly with ordered subsets (OSEM) [8, 9], are now widely used for diagnostic PET. There is a well-known tradeoff between increased image sharpness and increased noise variance as the number of EM iterations is increased during reconstruction [10, 11]. In the context of ¹⁸F-FDG PET for TLE, a greater number of iterations may increase image sharpness and thus improve the detectability of small regions of hypometabolism, but this may be offset by the associated increase in image noise. The impact of the number of iterations performed during reconstruction on diagnostic accuracy has been studied in the context of several imaging tasks [12-16], and unsurprisingly the opti-
mal implementation of EM reconstruction is dependent on the task. However, to our knowledge such a study has not been undertaken for $^{18}$F-FDG PET acquired to diagnose TLE, and the specific reconstruction algorithm used is typically determined by physician preference. Our aim here is to evaluate the impact of the number of EM reconstruction iterations, and hence the tradeoff between image sharpness and noise, on the diagnostic accuracy of clinical $^{18}$F-FDG PET scans acquired for TLE diagnosis. We do this for two OSEM reconstructions with a blinded reader study using a receiver operating characteristic (ROC) analysis of the predictive power of $^{18}$F-FDG PET for identifying surgical candidates, and a comparison of the sensitivity of the two different reconstructions for identifying patients who improved with surgery.

**Materials and methods**

**Subject selection**

This study was carried out as a retrospective analysis of clinical $^{18}$F-FDG PET studies previously acquired at our institution for the diagnosis of medically intractable TLE. The medical and imaging records of all patients (n=184) who received $^{18}$F-FDG PET scans for the diagnosis of TLE between 2000-2010 were investigated for inclusion and exclusion criteria as follows. Inclusion criteria included diagnosis or suspicion of TLE; consideration for surgical treatment of epilepsy; age $\geq$ 18 years; clinical documentation of age, gender, seizure onset, and seizure frequency, antiepileptic drug (AED) trials, EEG report, PET report, MRI report, and in patients ultimately receiving surgery: postoperative follow-up, seizure frequency, and seizure character. Exclusion criteria included a history of cerebral vascular accident (CVA), brain tumor, head trauma, tuberous sclerosis, prior cranial surgery, and hemispheric congenital malformations (e.g., porencephaly, lissencephaly, perisylvian polymicrogyria, hemimegalencephaly).

Medical records were then reviewed to determine the documented findings of $^{18}$F-FDG PET, MRI, EEG exams, and postoperative outcomes. $^{18}$F-FDG PET was deemed positive if unilateral temporal lobe hypometabolism was noted in the medical record. MRI was deemed positive if mesiotemporal sclerosis, hippocampal atrophy, unilateral temporal atrophy, or temporal gliosis were noted. EEG was deemed localizing if reports indicated that seizures originated from one temporal lobe. Surgical outcome data were evaluated in patients who had at least one assessment of their postoperative seizure course in the electronic medical record. Surgical outcomes were graded according to the International League Against Epilepsy (ILAE) scale [17]. For the purposes of this study, surgical outcomes were further categorized as positive for ILAE scores of 1-4 (1 = absence of seizures, 4 = 4 seizures per year up to a 50% seizure reduction from baseline), and negative for scores of 5 or 6 (less than a 50% reduction in seizures from baseline). Of the patients receiving surgery, only one had a negative outcome (ILAE score of 5). Following the compilation of these data, subjects were anonymized. The general characteristics of this population of subjects and an analysis of their outcomes has previously been published [18].

A subset of 32 subjects were then selected to be used in the blinded reader study. This subset of subjects was selected to maximize the number of patients whose scans were initially read as negative with positive surgical outcomes (i.e. false negatives), as the interpretation of these scans is plausibly the most likely to change. The other scans included in the reader study were selected such that the population of patients used in the reader study maintained the general diagnostic characteristics of the overall population of patients who received $^{18}$F-FDG PET for TLE. In particular, the percentage of scans initially read as positive and negative was kept approximately the same between the subset of subjects used in the reader study and the overall number of $^{18}$F-FDG PET studies for TLE, as was the proportion of scans with findings concordant and discordant with MRI and/or EEG.

**PET image acquisition and reconstruction protocols**

Patients fasted for 6 hours prior to injection of $^{18}$F-FDG. Diabetic patients were instructed to withhold diabetic medications for 6 hours and blood glucose measurements were required to be < 200 mg/dL at the time of tracer injection. Patients were injected intravenously with 0.14 mCi/kg (minimum of 10 mCi) $^{18}$F-FDG, and were then instructed to relax quietly for 45 minutes in a dimly lit room. Patients were imaged at 60 minutes after injection with one of two scanners: the Advance and the Discovery VCT (GE
OSEM iterations for epilepsy diagnosis with PET

Table 1. Reconstruction parameters for each scanner used in the study. The number of iterations was the only parameter varied between the two OSEM reconstructions for each patient exam.

<table>
<thead>
<tr>
<th>Scanner</th>
<th>GE Advance</th>
<th>GE Discovery VCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction</td>
<td>2D OSEM</td>
<td>3D OSEM</td>
</tr>
<tr>
<td>Matrix Size</td>
<td>128x128x35</td>
<td>128x128x47</td>
</tr>
<tr>
<td>Voxel Size</td>
<td>2.3x2.3x4.25 mm³</td>
<td>2.3x2.3x3.27 mm³</td>
</tr>
<tr>
<td>Number of Subsets</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>Number of Iterations: Smooth</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Number of Iterations: Sharp</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Post-Smoothing Filter (Gaussian FWHM)</td>
<td>3.27 mm</td>
<td>2.57 mm</td>
</tr>
</tbody>
</table>

Figure 1. Example images of the reconstructions studied. Example slices are shown of the 2 different reconstructions that were compared for both the GE Advance and GE Discovery VCT scanners used in the study. Examples from each scanner for patients with ¹⁸F-FDG PET scans initially read as positive and with positive surgical outcomes, and for patients with ¹⁸F-FDG PET scans initially read as negative who did not receive surgery are both shown. Hypometabolic regions are indicated with arrows.

Healthcare).

Two OSEM reconstructions were performed on each ¹⁸F-FDG PET exam. The first used reconstruction parameters typical of our institution for the type of scanner used. These reconstructions were considered the smooth, relatively low-resolution, and low-noise standards. The second reconstruction increased the number of iterations used during reconstruction, while keeping all other parameters constant. The specific number of iterations to be used for the second set of reconstructions was determined by a nuclear medicine physician using a test set of subjects by qualitatively determining the number of iterations at which possible increased confidence in the diagnosis would be offset by increased noise. These reconstructions served as the sharper, higher noise comparisons. Pertinent reconstruction parameters are summarized in Table 1, and example reconstructions from each scanner for ¹⁸F-FDG PET scans initially read as both positive and negative are shown in Figure 1. Note that 3D OSEM reconstructions of shorter acquisitions with more subsets (35 versus 28) were used for the Discovery
OSEM iterations for epilepsy diagnosis with PET

VCT, which is standard practice at our institution. Corrections for normalization, deadtime, and scatter radiation were applied using system software. Attenuation correction was applied to the scans acquired on the GE Advance using a transmission scan acquired with two Ge-68 rod sources, and to the scans acquired on the GE Discovery VCT using a co-registered CT scan.

Reader study design and analysis

Each individual reconstruction was assigned a random number and all associated patient, exam, and reconstruction information were removed. All reconstructions were then presented to the blinded readers interspersed randomly. Two readers, reader 1 and reader 2, assigned a diagnostic score of 1-5 to each reconstructed image (1 = unequivocal hypometabolic focus, 2 = strong confidence of hypometabolic focus, 3 = moderate confidence of hypometabolic focus, 4 = equivocal for hypometabolic focus, 5 = no hypometabolic focus). Readers were also blinded to MRI, EEG, and other clinical findings.

ROC curves for surgical candidacy were then generated for each reconstruction, and analyzed for each reader separately and with their results combined. The area under the curve (AUC) was calculated for each ROC curve, and curves for smooth reconstructions (2 EM iterations) were compared with curves for sharp reconstructions (5 EM iterations) using the non-parametric comparison approach of DeLong et al. [19].

As only one patient out of all the patients meeting the inclusion and exclusion criteria for the study had a negative surgical outcome as we have defined (ILAE score of 5 or 6), the specificity of 18F-FDG-PET for predicting surgical outcomes cannot be assessed. We therefore compared only the sensitivities of the two OSEM reconstructions for identifying patients who improved with surgery at each level of reader confidence. Ninety-five percent confidence intervals (CIs) were found for the sensitivities at each level of diagnostic confidence using the Clopper-Pearson interval, and McNemar’s test was used to test for significance.

Access to imaging and medical records of all patients for the purpose of this study, and permission to reprocess and reinterpret imaging studies, was approved by the local institutional review board.

Results

Patient selection

A total of 120 patients met the initial inclusion and exclusion criteria and 32 were included in the reader study. As outlined above, the subset of 32 scans used for the reader study were selected to maximize the number of scans whose initial interpretations were false negative while keeping the percentage of scans initially read as 18F-FDG PET positive and negative for hypometabolism, and with concordant and discordant MRI and/or EEG findings, approximately the same as in the 120 patients meeting the inclusion and exclusion criteria. Of the 32 patients included in the reader study, 26 were scanned on the GE advance and 6 were scanned on the GE Discovery VCT. The diagnostic characteristics of the scans selected for the reconstruction study and for the overall population of patients undergoing 18F-FDG PET for TLE are summarized in Figure 2. Of the patients receiving surgery who were included in the reader study (n=18), 9 had an ILAE outcome of 1 (seizure free), 4 had an ILAE outcome of 2 (auras, but no seizures), 4 had ILAE outcomes of 3 (1-3 seizures per year), and 1 had an outcome of 5 (<50% reduction from baseline). The representation of these outcomes in the reader study population is likewise similar to their representation in the overall population of patients receiving surgery. Two patients were identified as surgical candidates but had not received surgery at the time of data collection. One patient was awaiting surgery at the time the study was conducted, and one did not proceed with surgical treatment. Both of these patients were scanned on the GE Discovery VCT.

Reader study results

The two readers were very consistent in their interpretations of the 18F-FDG scans, regardless of the number of iterations used. When the results of the two readers are combined, 48/64 (75%) of the studies were given an identical rating between the two reconstructions, and 13/64 (20.3%) were given ratings that differed by one degree of reader confidence.

The ROC curves for surgical candidacy are shown in Figure 3. The ROC curves of the individual readers and their combined results are both included. The area under each curve and the results of the nonparametric statistical com-
OSEM iterations for epilepsy diagnosis with PET

Comparison between them are summarized in Table 2. There was no statistically significant difference between the AUCs for the reconstructions with 2 iterations and 5 iterations for either of the outcomes. This was true for the results of both readers individually and with their results combined. As the acquisitions and reconstructions from the GE Discovery VCT were 3-dimensional and those from those from the GE Advance were 2-dimensional, the areas under ROC curves excluding the scans acquired on the GE DVCT (n= 6) were also examined (Table 2). Excluding the GE Discovery VCT scans also excludes the two patients identified as surgical candidates but who had not received surgery at the time of data collection. Excluding these scans did not alter the results.

The sensitivities of the 2 iteration and 5 iteration OSEM reconstructions for predicting surgi-
cal outcome at a moderate level of diagnostic
confidence (images rated 3 or higher counted
as positive reads) are shown in Figure 4. There
was no significant difference between the two
reconstructions for each reader separately or
with their results combined. This was true at all
other levels of diagnostic confidence as well
(data not shown). As with the ROC curves for
surgical candidacy, excluding the scans ac-
quired on the Discovery VCT did not alter the
results.

Discussion

The aim of this study was to retrospectively in-
vestigate the impact of the number of iterations
performed during OSEM reconstruction on the
sensitivity and specificity of 18F-FDG PET for pre-
dicting surgical candidacy and surgical out-
comes in TLE. This was done with a blinded
reader study comparing two OSEM reconstruc-
tions, differing in the number of iterations per-
formed. The ability of 18F-FDG PET to predict
surgical candidacy was evaluated with a ROC
analysis of the two reconstructions, and the
sensitivities of the two reconstructions for pre-
dicting surgical outcomes were compared. In
the cases studied here, the number of iterations
performed during OSEM reconstruction had no
statistically significant impact on the sensitivity
and specificity of 18F-FDG PET for predicting
surgical candidacy, or its sensitivity for predict-
ing surgical outcome. The nuclear medicine phy-
sicians’ interpretations of the PET studies were
essentially unchanged by the different recon-
structions, illustrated by the consistency with
which they interpreted the images between the
two reconstructions studied. Therefore using
physician preference to determine reconstruc-
tion parameters seems justified and acceptable
in this case. These results should be tempered
by the limited statistical power of our data.

The hypometabolic regions identified by nuclear
medicine physicians tended to be more global
decreases in FDG uptake in the temporal lobe
(Figure 1). Smaller lesions may not have been
identified in either reconstruction, as they may
have remained too blurred in the original recon-
struc-

Table 2. The areas under the ROC curves from each reconstruction for the individual readers and with
their results combined. The P-values for the difference between the AUCs of each reconstruction are
shown. None reached significance. The AUCs and the P-values comparing them after excluding the
scans acquired of the GE Discovery VCT are also shown. The exclusion of these scans did not alter the
results.

<table>
<thead>
<tr>
<th>Surgical Candidacy</th>
<th>All Scans</th>
<th>Discovery VCT Excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSEM Iterations</td>
<td>Reader 1</td>
<td>Reader 2</td>
</tr>
<tr>
<td>2</td>
<td>0.839</td>
<td>0.812</td>
</tr>
<tr>
<td>5</td>
<td>0.806</td>
<td>0.831</td>
</tr>
<tr>
<td>P-value</td>
<td>0.563</td>
<td>0.280</td>
</tr>
</tbody>
</table>

Figure 4. Surgical outcome sensitivity. The sensitivity of 18F-FDG PET for predicting a positive surgical outcome
for the two OSEM reconstructions studied. There was no statistically significant difference in the sensitivi-
ties of the reconstructions for either
of the readers, or with their results combined. Excluding the scans ac-
quired on the GE Discovery VCT did not alter the results.
OSEM iterations for epilepsy diagnosis with PET

structions with more iterations. It is possible that reconstructing with a number of iterations between 2 and 5 would be more optimal and would result in a better trade-off between image sharpness and noise variance.

We did not consider partial volume correction methods that aim to increase image sharpness while suppressing noise to improve the identification of small hypometabolic regions on 18F-FDG-PET or on SPECT studies [5-7]. Such methods are not yet available as clinical tools, whereas changing the number of iterations used in OSEM reconstructions can readily be performed and therefore may have a more immediate impact. Given that our results do not indicate any benefit to using the sharper but noisier images obtained with more iterations, more advanced partial volume correction methods might be needed to identify smaller regions of hypometabolism. However, if physicians rely on identifying a pattern of globally reduced 18F-FDG uptake, such methods may make little difference in subjective interpretation. If this is the case, one of the more objective methods of detecting small regions of hypometabolism might be required [20, 21].

We have focused on the number of iterations performed during OSEM reconstruction in order to study the tradeoff between bias and variance (resolution and noise in this case), but there are a number of other factors that will influence resolution and noise. The most notable extraneous factors in this study are the two PET cameras that were used, the GE Advance and the Discovery VCT, and the different acquisitions and reconstruction used for each scanner, 2D acquisition with 2D OSEM reconstruction for the Advance and a shorter 3D acquisition with 3D OSEM reconstruction using more subsets (35 versus 28) for the Discovery VCT (Table 1). The overall impact of these differences on the images from the two scanners is difficult to determine. The 3D acquisition and reconstruction of the Discovery VCT images should result in better noise properties in the images, but this may be offset by the shorter acquisition time. Likewise, the greater number of subsets used during reconstruction should give sharper images for the same number of iterations, though image sharpness is also likely influenced by the 3D acquisition. We do not attempt to address each of these issues, but instead try and focus only on the impact of iteration number by demonstrat-

ing that excluding the Discovery VCT scans does not significantly alter the results. In addition, the smooth, clinically standard reconstructions from the two scanners are qualitatively closer to each other in their resolution and noise properties than to the corresponding sharper reconstructions (Figure 1).

All reconstructions used for PET imaging have parameters that will affect the tradeoff between bias and variance. While these parameters are frequently determined by physician and institutional preference, it is possible that they could have an impact on diagnostic outcomes. Our investigation here indicates that this may not be the case with OSEM reconstructions of 18F-FDG PET images acquired for the diagnosis of TLE, as we found little difference between the two reconstructions studied. Our results are limited by their statistical power, by extraneous variables such as the two PET cameras used, and in that we only examined two possible numbers of iterations. However, the consistency with which the readers interpreted the images indicates that a substantial number of scans would have to be read to identify any difference between reconstructions. Different reconstruction parameters, such as an intermediate number of iterations between 2 and 5, may have made a greater impact on the interpretation of scans, but the consistency of the interpretations makes this unlikely as well. If changing the number of iterations performed in reconstruction had potential to change interpretations, less consistency would be expected in the data presented here, even if the area under the ROC curves and the sensitivities are nearly equivalent. As such it appears perfectly reasonable to use images that nuclear medicine physicians are most familiar and comfortable with. Nevertheless, such studies could be helpful in validating and optimizing the reconstruction and image processing methods used in different clinical imaging tasks. And in the case of 18F-FDG PET for the diagnosis of TLE a more rigorous study could be performed with more patients and varying more reconstruction parameters to validate the results presented here.

Conclusions

In this retrospective blinded analysis, we have investigated the impact of the number of iterations performed during OSEM reconstruction on the interpretation of 18F-FDG PET scans ac-
quired for the diagnosis of TLE. We found no difference between the reconstructions studied. This implies that the reconstructions used for the subjective clinical interpretation of \(^{18}\text{F}\)-FDG PET scans acquired for TLE can be determined by physician preference. More sophisticated means of partial volume correction may have a more significant impact on the diagnostic interpretation of such scans.

Acknowledgements

The authors would like to thank Dr. Alejandro Munoz del Rio of the Department of Radiology at the University of Wisconsin School of Medicine and Public Health for the helpful discussions and assistance with statistical analysis, and Mark McNall from the Department of Radiology at the University of Wisconsin School of Medicine and Public Health for his assistance.

The authors would also like to acknowledge financial support from the University of Wisconsin Medical Scientist Training Program, the University of Wisconsin Department of Radiology, and the NIH Radiological Sciences Training grant #T32 CA009206.

Address correspondence to: John M Floberg, Department of Medical Physics, University of Wisconsin-Madison, Madison, WI, USA E-mail: jfloberg@wisc.edu

References


[18] Struck AF, Hall LT, Floberg JM,Perlman SB, Dulli DA. Surgical decision making in temporal lobe epilepsy: A comparison of \(^{18}\text{F}\)FDG-PET,
