Global temporal lobe asymmetry as a semi-quantitative imaging biomarker for temporal lobe epilepsy lateralization: A machine learning classification study

Abstract

**Objectives:** The purpose of this study was to evaluate the utility of global semi-quantitative analysis via fluorine-18-fluorodeoxyglucose positron emission tomography (18F-FDG PET) at lateralizing seizure foci and diagnosing patients with unilateral temporal lobe epilepsy (TLE). **Methods:** Seventeen patients with unilateral TLE (11 right TLE and 6 left TLE) were retrospectively selected for semi-quantitative 18F-FDG PET analysis. Twenty-three control subjects with a Mini Mental State Examination (MMSE) score of 29 or greater were selected for comparison. Globally averaged standardized uptake value (gSUVmean) was computed for each temporal lobe. Lateralization indices (LI) and the absolute value of lateralization indices (|LI|) were calculated to assess the degree of asymmetry in each subject. Logistic regression analyses were performed at a probability cutoff of 0.5 to classify TLE patients as left or right TLE and to discriminate patients from control subjects. Receiver operating characteristic (ROC) curves were generated to evaluate the utility of LI and |LI| as classification predictors. The Bland Altman test was used to evaluate the reproducibility of the measurements.

**Results:** There was a statistically significant difference in gSUVmean computed LI between left and right TLE patients (P<0.01). There was no statistically significant difference in |LI| between the patient and control groups (P=0.22). Logistic regression revealed that 82% of TLE patients were lateralized correctly using LI as the sole predictor. The area under the ROC curve (AUC) was 0.80. Logistic regression using |LI| on the combined patient/control population showed a diagnostic accuracy of 65% and an AUC of 0.44.

**Conclusion:** We conclude that gSUVmean computed LI is a reliable and reproducible measure for predicting seizure lateralization in unilateral TLE patients. However, gSUVmean computed |LI| does not appear to be particularly effective at diagnosing TLE patients from control subjects. Further studies with more patients should investigate other machine learning techniques that combine gSUVmean with other diagnostic predictors.

Introduction

Drug-resistant temporal lobe epilepsy (TLE) often requires temporal lobectomy in order to mitigate seizure onset. Although life-threatening complication rates are low, postoperative cognitive decline is common [1]. In particular, studies have shown that memory deficits occur in as many as 63% of patients after temporal lobectomy [2]. Moreover, in 20%-40% of medically intractable cases, surgical resection is unable to improve symptomatology [3]. The success of surgery at relieving epileptic symptoms while minimizing postoperative complications relies on successful lateralization of the epilepto genic zone. Precise localization of seizure onset sites allows for the affected tissue to be removed, while sparing neighboring structures. Thus, diagnostic presurgical assessments are critical to the long-term prognosis of patients with medically intractable TLE.

Determination of seizure lateralization in patients with unilateral TLE typically involves invasive depth electrode stereoelectroencephalography (SEEG) [4]. Although widely used, SEEG is associated with a number of postoperative risk factors including hemorrhaging and infection, with complication rates between 3%-17% [5, 6]. Furthermore, administration of SEEG tests are often accompanied by extended stay in an epilepsy monitoring unit (EMU) where patients are weaned off of antiepileptic drugs in an attempt to elicit seizure activity. Such diagnostic techniques are incredibly stressful for the patients, and in some cases can lead to permanent neurological deficits [5].

Noninvasive neuroimaging modalities such as fluorine-18-fluorodeoxyglucose positron emission tomography (18F-FDG PET) and magnetic resonance imaging (MRI) are
Methods

Population information

A large population of unilateral TLE patients was obtained for retrospective analysis. Because TLE is associated with a number of severe comorbidities, strict exclusion criteria were applied to eliminate patients who were deemed likely to exhibit altered brain function based on other disorders. As a result, patients presenting with a history of traumatic brain injury, auto-vehicular accidents, major brain tissue loss, structural abnormalities, brain tumors, chemotherapy, extratemporal epilepsy, stroke, prior brain tissue resections, genetic disorders affecting the brain, or infections of the central nervous system were excluded. A total of 17 patients (11 right TLE and 6 left TLE) were identified and used for subsequent analysis.

All patients were lateralized at the Hospital of the University of Pennsylvania and given a clinical diagnosis of either left or right TLE. Seizure lateralization was determined based on a combination of diagnostic tests administered as necessary including depth electrode placement, surface-EEG, visual 18F-FDG PET, visual MRI and extended stay in the EMU. A total of 23 control subjects were selected for statistical comparison. All subjects were free of notable central nervous system abnormalities and scored 29 or higher on Mini Mental State Examinations (MMSE). Control subjects underwent 18F-FDG PET at Jefferson Hospital.

Global semi-quantitative analysis

Global semi-quantitative analysis of temporal lobe 18F-FDG PET activity was performed on each patient and control subject by the same observer according to our previously published methodology [8]. Manually drawn regions of interest (ROI) encapsulating the entire temporal lobe were drawn on individual slices using the medical imaging application, OsiriX. Notable subcortical structures such as the thalamus and basal ganglia were excluded. Slice ROIs were interpolated to produce a three-dimensional temporal lobe reconstruction (Figure 1). Globally averaged standardized uptake value (gSUVmean) was calculated for each temporal lobe.

Figure 1. An image of a 3-D temporal lobe reconstruction based on interpolation of manually drawn ROI encompassing the entire temporal lobe.

Lateralization indices (LI) were computed based on these global measurements to assess temporal lobe asymmetry according to the following formula, where RTL and LTL refer to the gSUVmean of the right and left temporal lobes respectively:

$$\text{RTL} - \text{LTL}$$

$$\text{RTL} + \text{LTL}$$

Thus, higher LI magnitude is associated with greater metabolic asymmetry in the temporal lobes. Moreover, a negative LI indicates hypometabolism on the right side. Differences in mean LI and LI magnitude (|LI|) were reported using the Welch’s unequal variances t-test at a 0.01 alpha level.

Classification analysis

Binary classification was applied first to the TLE population alone to assess the lateralization ability of gSUVmean computed LI at discriminating between left and right TLE. A machine learning logistic regression analysis was performed to predict seizure lateralization using a probability of 0.5 as the classification cut-off. Leave-one-out cross validation (LO-OCV) was performed to assess the accuracy of the classifier and to determine the uncertainty in the cut-off parameter. The LI score was used as the sole predictor for this analysis with the left-sided lateralization taken as the test positive case.

A similar regression analysis was applied on the combined patient and control subject populations to evaluate the accuracy of temporal lobe asymmetry at diagnosing patients from controls. Absolute value of LI (|LI|) was used as the sole predictor for binary classification. A machine learning logistic regression analysis was performed to predict patient or control status, with greater |LI| (and thus, greater temporal lobe asymmetry) associated with a positive patient classification. A probability of 0.5 was used as the classification cutoff. Leave-one-out cross validation was again performed to assess the accuracy of the classifier and deter-
mine the uncertainty in the cut-off parameter. Confusion matrices and receiver operating characteristic (ROC) curves were calculated to visualize the sensitivity and specificity of LI and |LI| as a function of varying cut-off values.

Reproducibility analysis
The Bland-Altman method was used to assess the inter-reader and intra-reader reproducibility of our segmentation method. Temporal lobe global measurements were performed twice by one reader and repeated by another independent reader for each lobe using OsiriX software.

The Bland-Altman graph displays a scatter plot of the differences in measurements plotted against the averages of the two measurements. The limits of agreement (LOA) are defined as the mean difference ±1.96 standard deviations of differences.

Results

Asymmetry analysis
Comparison between the left (n=6) and right (n=11) TLE populations using the Welch’s unequal variances t-test revealed a statistically significant difference in LI between the means of the two groups (P<0.01). Boxplots showing data for the patient populations are shown in Figure 2. There was a trend towards greater magnitude of asymmetry (|LI|) in left TLE patients, though this finding was not statistically significant (P=0.26). Summary statistics are presented in Table 1.

![Figure 2. LI in left and right TLE patients. The LI statistics for the left (n=6) and right (n=11) TLE populations. The centerlines of the boxplots denote the median LI values. The edges of the boxplots represent the interquartile ranges, while the whiskers cover the entire range of the data. Welch's unequal variances t-test revealed a significant difference between the means of the two groups (P<0.01).](image)

The degree of asymmetry between control subjects (n=23) and TLE patients (n=17) was evaluated by comparing the absolute value of the LI (|LI|) between the two populations (Figure 3). Temporal lobe epilepsy patients trended towards greater asymmetry between temporal lobes, though this difference was not statistically significant (P=0.26). Summary statistics are presented in Table 2.

![Figure 3. |LI| in TLE Patients vs. Controls. The |LI| statistics for the control subject (n=23) and TLE (n=17) populations. The centerlines of the boxplots denote the median |LI| values. The edges of the boxplots represent the interquartile ranges, while the whiskers cover the entire range of the data. Welch's unequal variances t-test failed to reveal a significant difference between the means of the two groups (P<0.22).](image)

Lateralization classification
Table 3 shows the confusion matrix for the left/right classification of the TLE patient population using LOOCV. Left TLE was taken as the condition positive case for this analysis. 82% of TLE patients (14/17) were correctly lateralized using gSUVmean computed LI as the sole predictor for logistic regression. Specifically, 4/6 left TLE patients and 10/11 right TLE patients were lateralized correctly. The probability >0.5 classification cut-off of the logistic function corresponded to an LI of 0.009±0.001 (mean±standard deviation) across the LOOCV.

The ROC curve for the lateralization classification is shown in Figure 4. The AUC across all discrimination boundaries was 0.80, indicating that gSUVmean computed LI is an informative predictor for TLE lateralization.

<table>
<thead>
<tr>
<th>Table 1. Lateralization Summary statistics. The summary statistics of lateralization index (LI) and magnitude of lateralization index (</th>
<th>LI</th>
<th>) for the left (n=6) and right (n=11) TLE patient populations (mean±standard deviation). P values were computed using the Welch’s unequal variances t-test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>Right</td>
<td>P value</td>
</tr>
<tr>
<td>LI</td>
<td>0.046±0.038</td>
<td>-0.019±0.028</td>
</tr>
<tr>
<td></td>
<td>LI</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Diagnosis Summary Statistics. The magnitude of lateralization index (</th>
<th>LI</th>
<th>) for the control (n=23) and patient (n=17) populations (mean±standard deviation). P values were computed using the Welch’s unequal variances t-test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>TLE</td>
<td>P value</td>
</tr>
<tr>
<td></td>
<td>LI</td>
<td></td>
</tr>
</tbody>
</table>
The diagnosis ROC curve using gSUVmean computed \(|LI|\) is plotted against an uninformative predictor in Figure 5. The AUC across all diagnostic thresholds was 0.44, indicating poor performance of gSUVmean computed \(|LI|\) as a diagnostic predictor.

Diagnostic classification

The confusion matrix for the patient/control diagnostic classification is shown in Table 4. A total of 26/40 cases were classified correctly (65%). Five out of seventeen (29%) patients and 21/23 (91%) control subjects were correctly diagnosed, corresponding to 12 false negatives and 2 false positives. The LOOCV revealed the probability >0.5 logistic function boundary to be \(|LI|=0.040\pm0.003\) (mean±standard deviation).

Reproducibility

We evaluated the reproducibility of our global analysis technique using the Bland-Altman test. We report a high level of intra-reader reproducibility, with paired measurements performed by the same observer revealing that 96% of differences lie between LOA (Table 5). We also report fairly high levels of inter-reader reproducibility as well. Paired measurements between the two readers showed that 96% and 91% of differences lie between LOA on successive trials.

Discussion

In this study we have shown that temporal lobe asymmetry determined by global semi-quantitative analysis via \(^{18}\)F-FDG PET is a reliable predictor of seizure lateralization in unilateral TLE patients. Global LI scores were significantly different between patients clinically diagnosed as either right or left TLE (P<0.01), showing ipsilateral hypometabolism in the affected lobe. At an asymmetry cut-off averaged across LOOCV iterations of LI >0.009±0.001 (mean±standard deviation) classified as right TLE, we achieved a lateralization accuracy of 82%. As expected, the derived boundary between left and right TLE was measured close to LI=0, which represents the metabolically symmetric case.

The AUC analysis can be interpreted semantically in a number of ways. Two of the most useful ways of interpreting AUC are as follows: as the average sensitivity across all possible values of specificity, or equivalently, the probability of correctly classifying both a randomly chosen true positive and a randomly chosen true negative [9, 10]. As a result, the
**Table 5.** The Bland-Altman inter-reader and intra-reader reproducibility test results of temporal lobe measurements. Temporal lobe global measurements performed twice by one reader and repeated by another independent reader using OsiriX software to assess the inter-reader and intra-reader reproducibility. The limits of agreement (LOA) are defined as the mean difference ±1.96 standard deviations of differences. Abbreviations: R: Reader, M: Measurement, LOA: Limits of Agreement

<table>
<thead>
<tr>
<th>Reproducibility test</th>
<th>Side of measurement</th>
<th>Mean difference (Bias)</th>
<th>Standard deviation (SD)</th>
<th>Bland - Altman 95% LOA</th>
<th>% Within LOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-reader R1M1 - R1M2</td>
<td>Right</td>
<td>0.0769</td>
<td>0.0765</td>
<td>-0.0730</td>
<td>0.2268</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>0.0716</td>
<td>0.0768</td>
<td>-0.0789</td>
<td>0.2221</td>
</tr>
<tr>
<td>Inter-reader R1M1 - R2</td>
<td>Right</td>
<td>0.0391</td>
<td>0.1435</td>
<td>-0.2422</td>
<td>0.3204</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>-0.1698</td>
<td>0.1861</td>
<td>-0.5345</td>
<td>0.1949</td>
</tr>
<tr>
<td>Inter-reader R2M2 - R2</td>
<td>Right</td>
<td>-0.0378</td>
<td>0.1628</td>
<td>-0.3568</td>
<td>0.2812</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>-0.2414</td>
<td>0.1623</td>
<td>-0.5596</td>
<td>0.0768</td>
</tr>
</tbody>
</table>

**Figure 6.** The Bland-Altman plots to assess the inter-reader and intra-reader reproducibility and repeatability of our segmentation method. The Y axis shows the difference between the two paired gSUVmean measurements, and the X axis represents the average of those measurements. The limits of agreement (LOA) are defined as the mean difference ±1.96 standard deviations of differences.
AUC describes the predictive power of the specified measurement relative to the reference value of 0.5, which represents classification by chance. Our finding of an AUC = 0.80 for lateralization suggests that globally averaged LI is a strong semi-quantitative biomarker for localizing seizure foci. Our Bland-Altman analysis revealed that our technique for computing gSUVmean across entire temporal lobes is highly reproducible. Ninety six percent of differences between paired measurements made by the same observer were between LOA, indicating a high level of intra-reader reproducibility. Moreover, 96% and 91% of differences between paired measurements made by different observers were between LOA. We suspect that for a greater sample size, inter-reader reproducibility would increase even further. This provides a convincing argument for the superiority of global semi-quantitative analyses over simple visual analyses which are subject to interpretation bias.

Our results are consistent with other studies indicating that semi-quantitative $^{18}$F-FDG PET is an accurate predictor of seizure lateralization. Several studies employing regional ROI analyses have also demonstrated this high predictive ability by examining a number of metabolic regions throughout the brain. Pustina et al. (2015) used a similar logistic regression technique and found that asymmetry across regional masks drawn on mesial and lateral temporal lobes accurately predicted lateralization in 97% of patients [11]. Another $^{18}$F-FDG PET study conducted by Kerr et al. (2013) reported an accuracy of 89% for both manual and computer aided segmentation techniques [12]. Muzik et al. (1998) calculated an AUC of 0.94 using localized $^{18}$F-FDG PET abnormalities less than 1cm² in area [13]. Such studies have generally utilized the placement of small ROI involving multidimensional analyses of many predictors combined with complex mask smoothing procedures and image segmentation. Although our reported lateralization accuracy is somewhat lower (82%), our comparatively simpler analysis procedure lends itself well to clinical adaptation. Our results indicate that global metabolic asymmetry is a reliable metric that can be applied clinically with a minimal amount of training. We therefore present these results within the context of sacrificing some degree of accuracy for the added benefit of widespread clinical use.

A number of MRI based machine learning studies have been able to predict seizure lateralization with a high degree of accuracy as well. Yang et al. (2015) applied a support vector machine classification using a variety of resting state fMRI functional connectivity parameters and reported a LOO-CV lateralization accuracy of 83% [14]. Other studies using different combinations of structural parameters quantified with MRI reported lateralization accuracies of 67%-100% [11]. Another study examining the role of hippocampal atrophy in seizure lateralization was able to achieve accuracies of 100% and 94% for patients with and without atrophy respectively using a variety of MRI predictors. However, these studies rely on the presence of MRI-detectable lesions for accurate classification. Indeed, it has been reported that even unilateral TLE may result in bilateral structural lesions that make lateralization via MRI difficult to perform [15]. It is well known that metabolic changes which can be detected with $^{18}$F-FDG PET predate these structural changes as well. Moreover, studies have shown $^{18}$F-FDG PET to be a more informative predictor of seizure lateralization than structural imaging, and confirmed $^{18}$F-FDG PET accuracy at predicting postoperative outcome [11, 16]. We therefore advocate for the use of semi-quantitative $^{18}$F-FDG PET in clinical diagnoses, particularly when patients are MRI-negative.

Although globally averaged LI seems to be an accurate predictor of seizure lateralization, the absolute magnitude of temporal lobe asymmetry ([LI]) does not appear to be a reliable biomarker for diagnosing TLE. Our regression analysis revealed a diagnostic cut-off of [LI] = 0.040±0.003, which yielded an accuracy of only 65% on the LOO-CV set. Moreover, the AUC analysis showed that temporal lobe asymmetry is not a particularly useful predictor across all thresholds. Consistent with our results, Kerr et al. (2013) reported a lower accuracy at diagnosing TLE patients from healthy controls compared to determining unilateral lateralization amongst the TLE group alone [12]. Still, there have been several PET studies reporting greater diagnostic accuracies (82%-90%) than we present here [12, 13, 17]. Muzik et al. (1998) reported a diagnostic accuracy of 90% using a threshold equivalent to an [LI] = 0.075 [13]. However, when using a threshold more similar to ours ([LI] = 0.05), the authors reported a much lower accuracy, with an extremely high false positive rate (0.89) [13]. We therefore conclude that semi-quantitative PET may still be an accurate diagnostic tool, but only in patient populations with high degrees of metabolic asymmetry. Consistent with this notion, our error mostly took the form of false negatives-suggesting that our patient population was much less asymmetric than populations used in other $^{18}$F-FDG PET studies. As a result, our low diagnostic accuracy may be an artifact of our relatively small patient population. We suggest further studies be conducted that account for the severity of TLE symptoms so that the diagnostic ability of PET can be evaluated on populations with different levels of asymmetry.

We also found that on average, our control subjects exhibited mild left temporal lobe hypometabolism (LI = 0.019±0.015). Whether or not this deviation from pure symmetry is an artifact of our subject population or represents a notable metabolic finding should be explored in future studies. Indeed, Kerr et al. (2013) reported a higher accuracy in diagnosing right TLE patients from controls than left TLE patients [12]. It is therefore feasible that left temporal lobe asymmetry in healthy controls could contribute to lower diagnostic accuracies. We also report that left TLE patients tended to have greater asymmetry than right TLE patients, although this finding was not statistically significant (P = 0.26). Nevertheless, this trend is consistent with several studies reporting greater degrees of atrophy in left TLE patients [11, 12]. We suspect that a greater sample size would result in statistical significance for global LI on $^{18}$F-FDG PET as well.

Although we provide convincing evidence for the utility of global semi-quantitative analysis of temporal lobes on $^{18}$F-FDG PET, our study is not without limitations. The relatively small sample sizes for the left and right TLE groups may not
be indicative of the larger TLE population. We attempted to apply relatively strict exclusion criteria to limit effects from comorbid diseases, but this exclusion process was not as robust as is common for prospective studies. We also were not able to age-match our controls to our patient population, leaving open the possibility for age-related interactions. We also report that the clinical evaluations used to represent the gold standard diagnoses were based on a consensus of clinical data, and not necessarily made under consistent conditions for each patient. However, this is of limited concern given that each patient was eventually diagnosed to the best ability of the observing physician.

In summary, we found that global semi-quantitative analysis of temporal lobe asymmetry is a good predictor of seizure lateralization, but not TLE diagnosis relative to control subjects. Future studies should assess whether or not different statistical methods reveal the same results. Specifically, a regression threshold of 0.5 may not be the optimal cutoff for diagnosing TLE. Further work should also be conducted to evaluate the synergistic effects of combining global $^{18}$F-FDG PET asymmetry scores with those of focal abnormalities in a multi-dimensional machine learning context. Subsequent investigations will advise on how global asymmetry can be incorporated with other clinical measures to provide optimal diagnostic accuracy.

The authors of this study declare no conflicts of interest.

Bibliography