Seizure 30 (2015) 132–135

Contents lists available at ScienceDirect

Seizure

journal homepage: www.elsevier.com/locate/yseiz

Short communication

Variability in clinical assessment of neuroimaging in temporal lobe epilepsy

Aaron F. Struck *, Michael B. Westover

Massachusetts General Hospital, Harvard Medical School, Division of Epilepsy, Wang 735, 55 Fruit St. Boston, MA 02114, United States

ARTICLE INFO

Article history:
Received 22 April 2015
Received in revised form 21 June 2015
Accepted 23 June 2015

Keywords:
FDG-PET
MRI
Temporal lobe epilepsy
Inter-rater reliability

ABSTRACT

Purpose: Neuroimaging is critical in deciding candidacy for epilepsy surgery. Currently imaging is primarily assessed qualitatively, which may affect patient selection and outcomes.
Method: The epilepsy surgery database at MGH was reviewed for temporal lobectomy patients from the last 10 years. Radiology reports for MRI and FDG-PET were compared to the epilepsy conference consensus. First, specific findings of ipsi/contra hippocampal atrophy and T2 signal changes were directly compared. Next the overall impression of presence of hippocampal sclerosis (HS) for MRI and temporal hypometabolism for PET was used for sensitivity/specificity analysis. To assess predictive power of imaging findings logistic regression was used.
Results: 104 subjects were identified. 70% of subjects were ILAE class I at 1-year. Radiology reports and the conference consensus differed in 31% of FDG-PET studies and 41% of MRIs. For PET most disagreement (50%) stemmed for discrepancy regarding contralateral temporal hypometabolism. For MRI discrepancy in ipsilateral hippocampal atrophy/T2 signal accounted for 59% of disagreements. When overall impression of the image was used the overall reliability between groups was high with only MRI sensitivity to detect HS (0.75 radiology, 0.91 conference, $p = 0.02$) was significantly different between groups. On logistic regression MRI was a significant predictor of HS, but still 36% of patients with normal MRI as read by both groups had HS on pathology.
Conclusion: Despite some difference in specific radiologic findings, overall accuracy for MRI and PET is similar in clinical practice between radiology and conference; nonetheless there are still cases of hippocampal pathology not detected by standard imaging methods.

1. Introduction

Temporal lobectomy is an effective treatment for medication refractory temporal lobe epilepsy [9], but not all patients benefit equally. 10–40% of patients still have seizures post-operatively [7,9]. While there are many reasons for surgical failures one potential etiology with minimal investigation is the variability in the interpretation of pre-operative imaging. Pre-surgical neuroimaging typically consists of MRI and ^{18}F-Fluorodeoxyglucose (FDG)-positron emission tomography (PET). The primary findings of interests are the presence of T2 hyper-intensity and atrophy of the hippocampus for MRI and temporal lobe hypometabolism for FDG-PET [4]. Typically images are interpreted by neuro-radiologists and then again by clinicians at an epilepsy surgical conference. Findings can be readily apparent, but in many cases they are subtle and controversial. Most centers rely on expert opinion to interpret images. This methodology is by its nature subjective and may be susceptible to issues of intra/inter-rater reliability. It is unclear how this subjectivity may affect patient selection and thereby surgical outcomes. This study assesses this variability in a real-world clinical setting by comparing radiology reports to epilepsy surgery conference notes. Data is then compared to endpoints including seizure-free rates and presence of hippocampal sclerosis on pathology.

2. Methods

The epilepsy surgery database at Massachusetts General Hospital (MGH) was queried for patients that underwent temporal lobectomy from 2003 to 2013. The study was conducted with approval from the IRB. Inclusion criteria were, age >18, epilepsy surgery conference presentation, and MRI and PET imaging at MGH. Patients were excluded if they had a structural lesion of the temporal lobe other than hippocampal sclerosis (HS), based on the

* Corresponding author. Tel.: +1 608 628 4630; fax: +1 617 726 9250.
E-mail address: astruck@mgh.harvard.edu (A.F. Struck).

http://dx.doi.org/10.1016/j.seizure.2015.06.011
1059-1311/

pre-operative MRI, prior epilepsy surgery, or surgery beyond an anterior temporal lobectomy. Imaging reports and conference notes were assessed for ipsilateral and contralateral findings including MRI findings of hippocampal atrophy and increased hippocampal T2 signal, and FDG-PET findings of temporal lobe hypometabolism. The epilepsy surgery conference was attended by a neuroradiologist that aided in the interpretation of the study. Only infrequently was this the same neuroradiologist that read the clinical study. Imaging reports were categorized on a scale of 0 for no findings suggestive of HS or temporal lobe hypometabolism, 1 for mild, possible, or subtle findings, and 2 for definite or probable findings. For each subject the interpretations were compared for change in the degree of certainty of findings and for new findings. The following findings were assessed: change in ipsilateral and contralateral T2 signal in the hippocampus, ipsilateral and contralateral hippocampal atrophy, and in the case of PET ipsilateral contralateral hypometabolism. For example if a radiology report did not find hippocampal atrophy and the conference felt that is might be present that would be considered a new finding and the interpretation was different. If the conference felt there was definite increased hippocampal T2 signal and the radiology found possible increased T2 signal this would be considered a difference in the strength of the finding. For purposes of sensitivity/specificity analysis only ipsilateral findings were considered and any finding, independent of the strength of that finding, suggestive of ipsilateral hippocampal sclerosis or in the case of PET ipsilateral hypometabolism was considered positive. These binominal data were used to develop 2X2 contingency tables to assess sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy. Proportions were compared using Z-test comparisons of proportions. Logistic regression was used to assess the relationship between interpretation and the categorical outcomes variables of HS on pathology and seizure freedom at 1-year post-operation (ILAE class 1) [1]. Univariate and multivariate logistic regression using a Chi-square statistic with a $p < 0.05$ was used to determine significance of predictor variables and a McFadden pseudo-R^2 was used to assess goodness of fit [5]. Statistical analysis was done with the *Matlab Statistics Toolbox Release 2013b* (MathWorks, Natick, MA).

The MR images are acquired with the MGH epilepsy protocol primarily acquired on a 3 T or 1.5 T Siemens MRI with sequences including high-resolution MPRAGE, coronal/axial FLAIR, T2 TSE through hippocampi, SPACE T2 FLAIR, SPACE T2, and axial SWI. PET imaging was performed approximately 45 min following administration of 5.0-mCi of [18F] fluorodeoxyglucose. Patients were screened with finger stick glucose prior to FDG administration. Imaging was performed on an ECAT HR scanner (CTI-Seimens, Knoxville, Tennessee) and acquired in 3D mode with attenuation correction from a transmission scan or with CT. A maximum likelihood reconstruction method was used.

3. Results

104 subjects were eligible for inclusion in the study. There was incomplete data from 40 subjects. These subjects lacked specific mention of conference interpretation of either PET or MR data. Mean age at surgery was 40 years. Mean follow-up duration was 36 months. 70% of the patients were ILAE class I at 1-year. Of the 104 cases, 71 (68%) of them had HS on pathology, while the others had non-specific gliosis or normal pathology.

76 patients had FDG-PET reports from radiology and documentation of discussion at conference. In 31% (24 patients), 95% confidence interval (CI): 21–41%, there was a difference in findings. In 12 patients (50% of cases with difference) the reason for difference was the conference disagreed with the finding of contralateral PET hypometabolism found in radiology reports. In 6 patients (25% of cases with difference) the conference found ipsilateral hypometabolism not found by radiology. In 1 patient (4%) radiology felt there was ipsilateral hypometabolism and the conference did not agree. In 6 patients (25%) radiology felt there was possible ipsilateral hypometabolism, but conference felt this finding was definite. Note that one patient had two disagreements with both a change in the strength of the ipsilateral hypometabolism and a disagreement regarding presence of contralateral hypometabolism.

With MRI 82 patients had both a report from radiology and documentation from conference. 34 patients (41%, 95%-CI 31–53%) had a difference in interpretation of MRI. 24 patients had new findings described by one group and not the other. 20 patients were felt to have ipsilateral hippocampal changes by conference and not by radiology with 8 having increased T2 signal, 5 having hippocampal atrophy, and 7 having both signal change and atrophy. Two patients were felt to have hippocampal atrophy by radiology and not by conference. One patient was felt to have contralateral hippocampal atrophy by radiology, but not by conference. One patient was felt to have abnormal ipsilateral hippocampal T2 signal by radiology but not by conference and hippocampal atrophy by conference but not by radiology. The remaining 10 patients had a disagreement regarding the strength of the findings. Conference found definite findings where radiology found only possible or questionable findings. All of these were in regard to ipsilateral hippocampal changes with 4 related to T2 signal change, 3 to hippocampal atrophy, and 3 involving both signal change and atrophy.

For the next analysis the interpretations were compared to surgical outcomes and surgical pathology. For MRI the presence of HS on surgical pathology was used as a reference standard. MRIs read as normal by radiology had HS on pathology 50% of the time, while that number was 36% for conference. For MRIs read as having possibly abnormal mesial temporal structures, HS was found on pathology 62% for radiology and 40% for epilepsy conference. For MRIs read as probably or definitely abnormal the percentages were 83% for radiology and 88% for epilepsy.

For FDG-PET, complete seizure-freedom at one year (ILAE class 1) was used as the reference standard. 17% of patients with a PET read as normal by radiology were a seizure-free, while 0% of patients with a normal PET as read by conference were a seizure free. For possibly abnormal PET imaging the numbers were 75% for radiology and 57% of conference. For patients whose PETs were read as definitely abnormal 72% were seizure-free for both the radiology and conference.

While there were many apparent differences between interpretations regarding which abnormalities (atrophy versus signal change) were present and to what extent the abnormalities were present, ultimately the critical factor is whether the medial temporal structures were felt to be abnormal or not. So for the sensitivity/specificity analysis any possible or probable abnormality was considered a positive test. Of the patients with a normal MR study as read by radiology 17/34 (50%) had HS, and 53/69 (77.7%) read as abnormal had HS. For conference interpretation of MR 5/14 (35.7%) read as normal had HS while 51/69 (74.0%) read as abnormal had HS. For the radiology interpretation of PET, 1/6 (16.7%) read as normal had ILAE class 1 at 1-year, while 53/73 (72.6%) read as abnormal were ILAE class 1 at 1-year. For conference interpretation of PET 0/3 (0%) read as normal were ILAE class 1 at 1-year and 43/61 (70.5%) of abnormal studies were ILAE class 1 at 1-year. Using this methodology the sensitivity, specificity, NPV, PPV, and overall accuracy were calculated (Fig. 1). The sensitivity of MRI to detect hippocampal sclerosis was the only measure that differed significantly between epilepsy conference and radiology interpretations (0.91 0.76 $p = 0.02$).