

Prediction of Naming Outcome With fMRI Language Lateralization in Left Temporal Epilepsy Surgery

William Louis Gross, MD, PhD, Alexander I. Helfand, PhD, Sara J. Swanson, PhD, ABPP, Lisa L. Conant, PhD, Colin J. Humphries, PhD, Manoj Raghavan, MD, PhD, Wade M. Mueller, MD, Robyn M. Busch, PhD, Linda Allen, BSN, RN, Christopher Todd Anderson, MD, Chad E. Carlson, MD, Mark J. Lowe, PhD, John T. Langfitt, PhD, Madalina E. Tivarus, PhD, Daniel L. Drane, PhD, David W. Loring, PhD, Monica Jacobs, PsyD, Victoria L. Morgan, PhD, Jane B. Allendorfer, PhD, Jerzy P. Szaflarski, MD, PhD, Leonardo Bonilha, MD, PhD, Susan Bookheimer, PhD, Thomas Grabowski, MD, Jennifer Vannest, PhD, and Jeffrey R. Binder, MD, on behalf of the fMRI in Anterior Temporal Epilepsy Surgery (FATES) Study

Correspondence
Dr. Gross
bgross@mcw.edu

Neurology® 2022;98:e2337-e2346. doi:10.1212/WNL.000000000200552

Abstract

Background and Objectives

Naming decline after left temporal lobe epilepsy (TLE) surgery is common and difficult to predict. Preoperative language fMRI may predict naming decline, but this application is still lacking evidence. We performed a large multicenter cohort study of the effectiveness of fMRI in predicting naming deficits after left TLE surgery.

Methods

At 10 US epilepsy centers, 81 patients with left TLE were prospectively recruited and given the Boston Naming Test (BNT) before and ≈ 7 months after anterior temporal lobectomy. An fMRI language laterality index (LI) was measured with an auditory semantic decision–tone decision task contrast. Correlations and a multiple regression model were built with a priori chosen predictors.

Results

Naming decline occurred in 56% of patients and correlated with fMRI LI ($r = -0.41, p < 0.001$), age at epilepsy onset ($r = -0.30, p = 0.006$), age at surgery ($r = -0.23, p = 0.039$), and years of education ($r = 0.24, p = 0.032$). Preoperative BNT score and duration of epilepsy were not correlated with naming decline. The regression model explained 31% of the variance, with fMRI contributing 14%, with a 96% sensitivity and 44% specificity for predicting meaningful naming decline. Cross-validation resulted in an average prediction error of 6 points.

Discussion

An fMRI-based regression model predicted naming outcome after left TLE surgery in a large, prospective multicenter sample, with fMRI as the strongest predictor. These results provide evidence supporting the use of preoperative language fMRI to predict language outcome in patients undergoing left TLE surgery.


Classification of Evidence

This study provides Class I evidence that fMRI language lateralization can help in predicting naming decline after left TLE surgery.

RELATED ARTICLE

 **Editorial**
fMRI Has Added Value in Predicting Naming After Epilepsy Surgery
Page 959

MORE ONLINE

 **Class of Evidence**
Criteria for rating therapeutic and diagnostic studies
[NPub.org/coe](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9000000/)

From the Departments of Neurology (W.L.G., A.H., S.J.S., L.L.C., C.H., M.R., L.A., C.T.A., C.E.C., J.R.B.), Anesthesiology (W.L.G.), and Neurosurgery (W.M.M.), Medical College of Wisconsin, Milwaukee; Departments of Neurology (R.M.B.) and Radiology (M.J.L.), Cleveland Clinic Foundation, OH; Departments of Neurology (J.T.L.) and Imaging Sciences (M.E.T.), University of Rochester, NY; Departments of Neurology (D.L.D., D.W.L.) and Pediatrics (D.L.D.), Emory University, Atlanta, GA; Department of Neurology (D.L.D., T.G.), University of Washington, Seattle; Departments of Psychology (M.J.) and Radiology (V.L.M.), Vanderbilt University Medical Center, Nashville, TN; Department of Neurology (J.B.A., J.P.S.), University of Alabama at Birmingham; Department of Neurology (L.B.), Medical University of South Carolina, Charleston; Department of Neurology (S.B.), University of California, Los Angeles; and Department of Neurology (J.V.), University of Cincinnati, OH.

Go to [Neurology.org/N](https://www.neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

fMRI in Anterior Temporal Epilepsy Surgery (FATES) study coinvestigators are listed in the Appendix 2 at the end of the article.

Glossary

BNT = Boston Naming Test; **CCF** = Cleveland Clinic Foundation; **FWHM** = full width at half-maximum; **IAP** = intracarotid anesthesia procedure; **LI** = laterality index; **MCW** = Medical College of Wisconsin; **RCI** = reliable change index; **ROI** = region of interest; **SD-TD** = semantic decision–tone decision; **TLE** = temporal lobe epilepsy; **UR** = University of Rochester.

Surgical removal of networks responsible for seizure generation is effective at reducing or eliminating seizures in many people with drug-resistant temporal lobe epilepsy (TLE).^{1,2} Roughly 30% to 50% of people with TLE who undergo surgery in the left temporal lobe show decline in their ability to name visual objects,^{3,4} and these deficits are correlated with worse quality of life.⁵ Predictors of this decline include older age at onset of epilepsy,^{3,6,7} older age at surgery,³ higher preoperative naming scores,³ and absence of medial temporal sclerosis.⁸ Models built from these factors, however, account for only a fraction of the wide variance in naming outcome. Naming decline is much more frequent in patients undergoing left-sided surgery, consistent with the long-held assumption that resection in the language-dominant hemisphere entails a greater risk. This suggests that preoperative language lateralization may provide additional predictive value within this left-sided surgery group.

Assessment of language dominance is a clinical staple in the preoperative evaluation for epilepsy surgery.⁹ Although left hemisphere surgery clearly confers greater naming risk,⁴ the predictive value of language dominance within this population remains unclear. Current evidence is based on 3 small fMRI studies showing that stronger preoperative lateralization of language to the left hemisphere was associated with greater naming decline.¹⁰⁻¹² A fourth small study, however, showed no correlation between these variables,¹³ and a fifth small study actually reported less naming decline in patients with stronger left dominance compared to patients with mixed or right dominance.¹⁴ This paucity of fMRI data and inconsistency between studies led the authors of a recent American Academy of Neurology practice guideline to conclude that evidence supporting the use of fMRI for predicting naming outcome in TLE surgery reached only Level C in quality.¹⁵ Similarly, the European Union's E-PILEPSY consortium concluded that meta-analysis of fMRI naming outcome prediction studies could not be performed given the small number of published studies.¹⁶

We addressed this knowledge gap via a multicenter, prospective study of picture naming outcomes in a large cohort of patients undergoing left anterior temporal lobe surgery for drug-resistant epilepsy. Patients underwent a standardized fMRI language lateralization protocol before surgery and standardized picture naming assessment before and \approx 7 months after surgery. The predictive value of fMRI laterality indices (LIs) was assessed relative to other demographic and clinical variables known to account for naming outcome variance. Our primary research question was whether fMRI language lateralization would improve the prediction of naming decline after left TLE surgery.

Methods

Participants

Patients undergoing surgery for left TLE were prospectively recruited from 10 epilepsy centers in the United States (eTable 1, links.lww.com/WNL/B945) from 2011 through 2017. Patients were included in the present analysis only if they underwent resection that included the anterior temporal lobe neocortex; patients with selective hippocampal or amygdala resections or ablations were excluded because such patients have not experienced significant naming declines in prior studies.^{17,18} From all centers, 84 patients were initially recruited and met surgical criteria, and 81 had complete datasets. The 3 missing patients were excluded for not completing follow-up ($n = 2$) and having corrupted MRI data ($n = 1$). Tailoring of resections based on language mapping with direct electric stimulation was performed in 50 patients (no mapping was performed in 26 patients, 5 had missing data). The decision to use or not use electric stimulation mapping was made by the treating clinical team. None of the patients in the current sample were included in previous fMRI naming prediction studies.

To differentiate decline associated with left-sided surgery from nonspecific effects of surgical intervention, we identified 58 prospectively recruited patients with right TLE with left-lateralized language from the Medical College of Wisconsin (MCW) who underwent right-sided temporal lobectomies. This control population was selected because patients with right TLE share a similar surgical experience with patients with left TLE but show little or no decline in naming.⁴ Deriving cutoffs for decline from this population will focus our analysis on decline due to cortical resection of language networks rather than other surgical factors. The control sample was selected with the use of the same inclusion criteria and underwent similar neuropsychological testing.

Standard Protocol Approvals, Registrations, and Patient Consents

Written informed consent was obtained from all patients before participating, and the study was monitored under each institution's local ethics standards review board. Results were prepared according to Strengthening the Reporting of Observational Studies in Epidemiology guidelines.

fMRI Methods

Imaging was performed at 3T with GE (Chicago, IL) Discovery MR750 (MCW), Siemens (Munich, Germany) Tim Trio (Cleveland Clinic Foundation [CCF], Emory University,

Medical University of South Carolina, University of Rochester [UR], University of California, Los Angeles), Siemens Prisma (CCF), Siemens Allegra (University of Alabama at Birmingham), or Philips (Best, the Netherlands) Achieva (University of Cincinnati, University of Washington, Vanderbilt University) scanners and head array receive coils with 8, 12, or 32 channels. Functional scans used a gradient-echo echoplanar imaging sequence. Gapless axial 2.5-mm slices were used in all cases. The number of slices ranged from 45 to 46. Complete coverage of the temporal lobes was required, and coverage included nearly the entire cerebrum, excluding only the most superior frontoparietal region in some participants. In-plane resolution was 2.5 × 2.5 mm (resulting in isometric cubic voxels). A short echo time (23 milliseconds) and relatively small voxel size were used to minimize signal dropout from intravoxel dephasing. Other parameters included repetition time 3 seconds, matrix 96 × 96, bandwidth ≈250 kHz, and flip angle 84°. A high-resolution T1-weighted image (spoiled gradient recalled or magnetization-prepared rapid gradient echo sequence) was acquired in each participant using ≈1-mm cubic voxels.

Analysis of MRIs was done at MCW using a single pipeline with the AFNI software package.¹⁹ Images were slice-time aligned, followed by rigid-body registration across repetitions. Spatial smoothing was then applied to the raw echo planar imaging data. Given that different system configurations add different amounts of smoothing to the raw MRIs,^{20,21} which can substantially affect signal-to-noise levels, smoothing was performed to a uniform target value across all patients with the AFNI program 3dBlurToFWHM, which iteratively applies small amounts of gaussian kernel smoothing until the target value is reached. Raw image smoothness across sites averaged 3.1-mm full width at half-maximum (FWHM) (SD 0.3 mm). Images were blurred to a smoothness of 6-mm FWHM, which resulted in an average additional blur of 2.9-mm FWHM. Blood oxygen–level dependent responses were extracted using multiple linear regression with stimuli coded as alternating blocks and motion parameters included as nuisance regressors.

Language lateralization was determined with a semantic decision–tone decision (SD-TD) contrast.^{22,23} In the semantic decision task, patients were presented recordings of spoken names of animals and asked to respond if the animal was both found in the United States and used by people. In the tone decision task, patients listened to sequences of 3 to 7 high (750 Hz) and low (500 Hz) pure tones and were instructed to respond if the sequence contained exactly 2 high tones. LIs derived from this contrast have been extensively validated by comparison with Wada language testing²⁴ and verbal memory outcome predictions.²⁵

LIs were calculated with the bootstrapped, multiple threshold method.²⁶ This method does not rely on a single threshold, so it is able to calculate a stable LI across a wider data quality range relative to classic fixed-threshold methods. LI values can in theory range from –100 (all activation in the right

hemisphere) to +100 (all activation in the left hemisphere). LIs were calculated using voxels within a region of interest (ROI) created from activations in a healthy cohort of 80 participants and the right hemisphere mirror image of these activations,²⁵ illustrated in eFigure 1, links.lww.com/WNL/B945. Most of this ROI is restricted to lateral cortical regions, distant from midline, making lateralization of activated voxels within it unambiguous. The only exception to this was a small area in the parietal lobe where the most medial portion of the ROI was within 3 mm of midline. Because this ROI covers a large area of the lateral cortex, supplemental analyses were repeated using its frontal, temporal, and parietal components to define the individual contributions of each.

We note that the LI metric is a continuous variable, reflecting evidence that language lateralization is primarily a graded phenomenon^{27–29} and our goal of developing a regression model for predicting naming outcome that is based on a combination of continuous variables. In general, division of continuous variables into categories serves no statistical purpose and results in loss of power; therefore, we did not divide patients into dominance subgroups (such as left, right, and bilateral) for the purposes of our analyses.

Naming Assessment

Picture naming was assessed using the Boston Naming Test (BNT) before and ≈7 months after surgery. In this test, patients are presented 60 line drawings of common objects and asked to verbally name them. The raw number of correctly named items is used as the final score. Each clinical site was responsible for scheduling the follow-up visit locally and performing neuropsychological testing.

To define postoperative decline, a reliable change index (RCI) was computed from the sample of patients with right TLE with the method of Jacobson and Truax.^{30,31} Briefly, an expected variability between preoperative and postoperative scores (SE_{diff}) is derived from a combination of their individual variability and correlation with each other. Change can then be assessed with the analogous *t* test formula (difference of means/ SE_{diff}). To simplify the calculation, an RCI cutoff value is chosen at a selected significance level by multiplying the SE_{diff} by the corresponding gaussian cumulative distribution function value (e.g., 1.64 for the 90th percentile).

Statistical Analyses

Preoperative predictors of interest, chosen a priori, included preoperative BNT score, age at surgery, age at epilepsy onset, duration of epilepsy, years of education, and the fMRI LI. We first assessed univariate relationships between each of these predictors of interest and BNT score change using the Pearson correlation. Predictor variables that had significant simple correlations with BNT score change were then entered into a 2-step hierarchical linear multiple regression model. In this model, postoperative change in BNT was entered as a continuous, linear variable. An initial model was built with all significant predictors aside from fMRI LI. Then, this model

Table 1 Descriptive Statistics for Preoperative Naming Outcome Predictors of Interest and Their Simple Correlations With BNT Change

	Mean (SD)	Range	r Value	t Value	df	p Value
Age at surgery, y	36.9 (12.2)	18.0–68.2	–0.23	–2.10	79	0.039
Age at epilepsy onset, y	21.4 (12.4)	0.0–57.0	–0.30	–2.81	79	0.006
Epilepsy duration, y	15.5 (11.6)	2.2–50.7	0.08	0.70	79	0.486
Education, y	13.2 (2.3)	7.0–19.0	0.24	2.18	79	0.032
Preoperative BNT score	47.5 (8.1)	25–60	–0.08	–0.72	79	0.471
fMRI LI	39.8 (46.4)	–82.3 – 88.9	–0.41	–3.95	79	<0.001

Abbreviations: BNT = Boston Naming Test; LI = laterality index.

was compared with a model that included the fMRI LI. This sequence was chosen to specifically highlight the additional information contributed by fMRI after all other available predictors have been taken into account.

Finally, Monte Carlo cross-validation analysis was used to estimate prediction error. The final regression model was fit using 90% of the randomly sampled patients, and the average prediction error was calculated from the mean absolute difference between predicted and actual scores in the remaining 10% of patients. This procedure was repeated with random resampling for 10,000 iterations to develop an empirical bootstrap distribution.

Data Availability

Additional data will be made available to any qualified investigators by request to the corresponding author (W.L.G.).

Results

Descriptive statistics for the preoperative predictors are listed in Table 1, and demographic data are listed in eTable 2, links.lww.com/WNL/B945. Because ≈40% of the patients came from MCW, analyses performed with the entire sample were qualitatively compared to an analysis obtained after removal of the MCW cohort to ensure that this cohort did not drive the effects. The majority of patients (70%) underwent a standard anterior temporal lobectomy, with the remainder undergoing tailored resections for clinical indications (details of each resection type are listed in eTable 3, links.lww.com/WNL/B945). Engel class was available for all but 1 patient; 60 patients (75%) had class 1 outcomes.

Functional Activation

Frequency of activation across brain regions is shown in Figure 1A, which illustrates the percentage of patients who had significant activation within voxels after application of a voxel-wise $p < 0.001$ threshold to each patient. As expected, activation was roughly homologous across hemispheres, with activation in the left hemisphere being more common and spatially more extensive. Areas of frequent activation included

regions known to be involved in linguistic and semantic processes, including the inferior frontal lobe, lateral and ventral temporal cortex, inferior lateral and medial parietal lobe, and superior and medial frontal lobe. Three examples of individual patient activation maps taken from the prediction analysis are shown in Figure 1B: the most right- and left-lateralized patients in our sample and a patient with representative symmetric activation. Of note, patients with atypical lateralization tend to activate homologous contralateral regions.

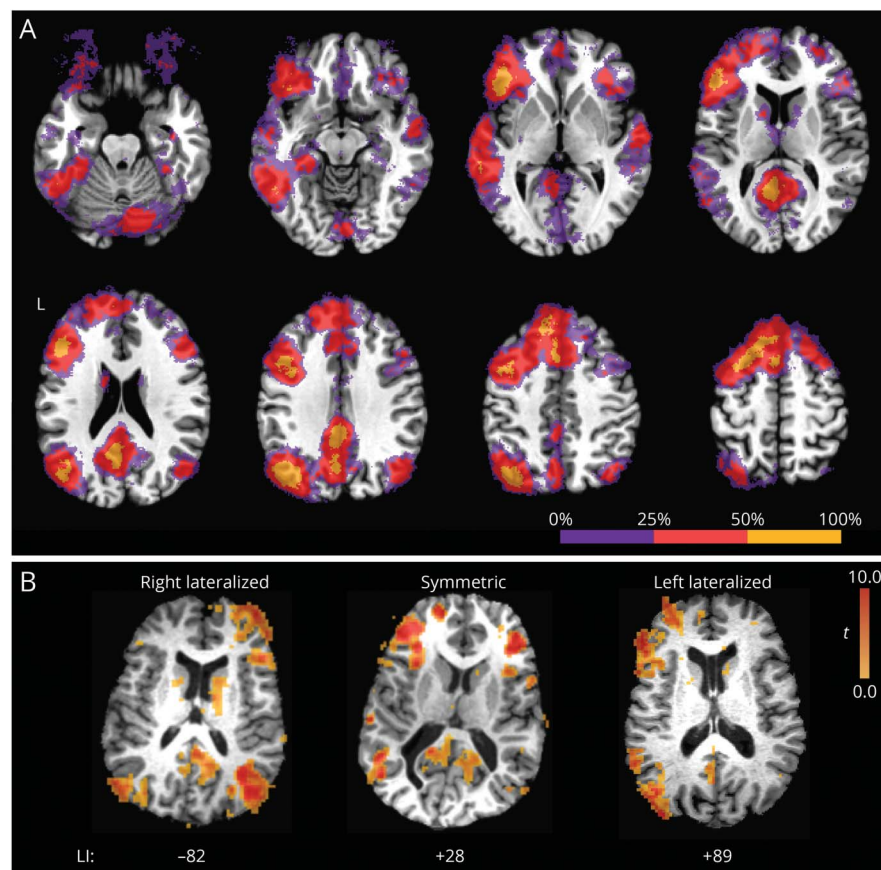
BNT Score Change and Simple Correlations

Average BNT score for the left TLE group preoperatively was 47.5 (SD 8.1) and postoperatively was 40.2 (SD 11.2). On average, patients with left TLE scored 7.3 points lower on the BNT after surgery ($t [80] = -7.75, p < 0.001$). This postoperative change was not different between male (7.8 points) and female (6.9 points) patients ($t [79] = -0.46, p = 0.647$). The distribution of BNT score change is shown in eFigure 2, links.lww.com/WNL/B945. Data from the right TLE surgery sample are shown in eTable 4, with the derivation of the RCI. Average BNT score change in this control sample was an increase of 1.9 points, and the 90th percentile RCI cutoff was 4.4, in line with estimates reported in nonsurgical patients.³² With the use of this cutoff, 45 of 81 (56%) of patients with left TLE had a clinically meaningful decline in naming after surgery, also consistent with previous reports.³

Although the small sample size (<5 patients) at some sites precluded direct comparisons, BNT score change was compared across 4 sites: MCW ($n = 32$), CCF ($n = 17$), UR ($n = 11$), and all other sites ($n = 21$). Using a nonparametric Kruskal-Wallis test, we found no difference in BNT score change among these groups ($H [3] = 1.07, p = 0.784$).

Simple correlations with BNT score change and the predictor variables of interest are shown in Table 1. Age at surgery, age at epilepsy onset, years of education, and fMRI-derived LI were all significantly correlated with change in BNT score. Preoperative BNT score and duration of epilepsy were not correlated with BNT score change. The relationship between

Figure 1 Typical Extent of fMRI Activation to the SD-TD Task Contrast



(A) Frequency of activations across individual patient maps. Voxels that contained significant activation in <25% of patients are shown in purple; voxels active in 25% to 50% of patients, in red; and voxels active in >50% of patients, in yellow. Activation tended to be left-lateralized in most patients and included areas known to be involved in language and semantic processing. (B) Example activation patterns in 3 individual patients taken from the prediction analysis: the most right and left-lateralized patients in our sample and a patient with representative symmetric activation. Of note, patients with atypical lateralization tend to activate homologous contralateral regions. LI = laterality index; SD-TD = semantic decision-tone decision.

BNT score change and fMRI LI is illustrated in Figure 2. Qualitatively similar results were found when we analyzed the sample without the MCW patients (eTable 5, links.lww.com/WNL/B945), confirming that these effects generalized across the sample. There was no difference in BNT score change between the 50 patients who underwent surgery tailored with electric stimulation language mapping (mean change -8.1) and the 26 patients who did not (mean change -5.8 ; $t [74] = -1.08$, $p = 0.285$). BNT score change did not differ across Engel outcome groups ($F_{3,76} = 2.075$, $p = 0.111$).

Multiple Regression Analysis

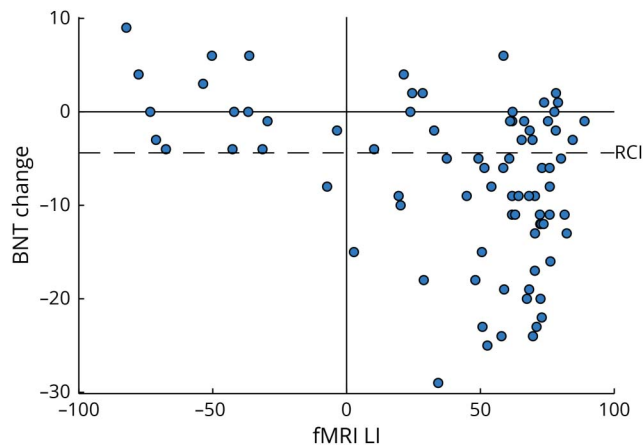
The variance explained and significance of both multiple regression models are shown in Table 2. The 3 clinical variables—age at onset, age at surgery, and education—together accounted for 13% of the variance in outcome as quantified by the adjusted R^2 value ($p = 0.004$). The fMRI LI provided an additional 14% of explained variance independently of all other predictors ($p < 0.001$). The final model accounted for 31% of the total variance in the sample. The fit parameters for the final models are shown in eTable 6, links.lww.com/WNL/B945. While both age at surgery and age at epilepsy onset were significantly correlated with outcome using simple correlations, they were also highly correlated with each other ($r = 0.553$, $p < 0.001$), resulting in

minimal additional information when either was added to the regression model after the other.

Predicted scores based on the full regression model vs actual change scores are plotted in Figure 3, illustrating the model accuracy. With the use of the RCI cutoffs for decline, the model had a sensitivity of 0.96 and specificity of 0.44 for correctly predicting binary naming outcomes (the full contingency table is shown in eTable 7, links.lww.com/WNL/B945).

The regression analysis was repeated excluding the MCW cohort to ensure that these results were not driven by that sample with qualitatively similar results (eTable 8, links.lww.com/WNL/B945). The full model R^2 values of the MCW and non-MCW cohorts were not significantly different when compared with the Fisher Z test ($z = -0.05$, $p = 0.962$). To explore the effects of the differential contributions of the frontal, temporal, and parietal lobes, the regression was also repeated using sub-ROIs from each of these lobes (as shown in eFigure 1). The results from these regressions are shown in eTable 9. In summary, the fMRI LI derived from each sub-ROI was a significant predictor of BNT score change. However, no single sub-ROI performed better than our original ROI.

Figure 2 Scatterplot of Individual Patient BNT Score Change by fMRI LI



Positive fMRI laterality indices (LIs) correspond to left-lateralized patients. Reliable change index (RCI) calculated from the right temporal lobe epilepsy cohort is shown as a dashed line, denoting a reliable decline. All but 1 of the declining patients had left-lateralized fMRI LIs. BNT = Boston Naming Test.

Because false-negative predictions result in distressing scenarios for patients (undergoing the surgery and experiencing an unexpected decline), the 2 patients who received false-negative predictions were examined further. Their individual data are listed in eTable 10, links.lww.com/WNL/B945. Neither patient had right-lateralized language by fMRI. One patient was predicted to have a decline of 4 points but actually declined by 5 points, which meets technical criteria for an unexpected decline, although the prediction was otherwise relatively accurate. The other patient had a strongly left-lateralized fMRI LI (58.8); however, the regression model resulted in a prediction of less decline because their age, age at onset, and education were all favorable.

Cross-Validation

The results of the cross-validation analysis are shown as a histogram in Figure 4. Average error in the held-back sample was 5.98, with a 10% to 90% interdecile of 4.1 to 8.0, suggesting that in a sample of novel patients, the average of BNT score change scores predicted with this regression model will likely be within 4 to 8 points of the actual average change.

Classification of Evidence

This study provides Class I evidence that fMRI language lateralization can help in predicting naming decline after left TLE surgery.

Discussion

We have shown, using a large sample of patients recruited prospectively from multiple epilepsy centers, that a multivariable regression model including an fMRI language LI is predictive of naming outcome after left anterior TLE surgery. We also confirmed the significance of age at surgery, age at

Table 2 Significance of Hierarchical Regression Models Created by First Adding Significant Non-fMRI Predictors and Then Adding fMRI

Model	R ² value	aR ² value	ΔaR ² value	p Value
Base model	0.16	0.13	—	0.004
fMRI model	0.31	0.27	0.14	<0.001

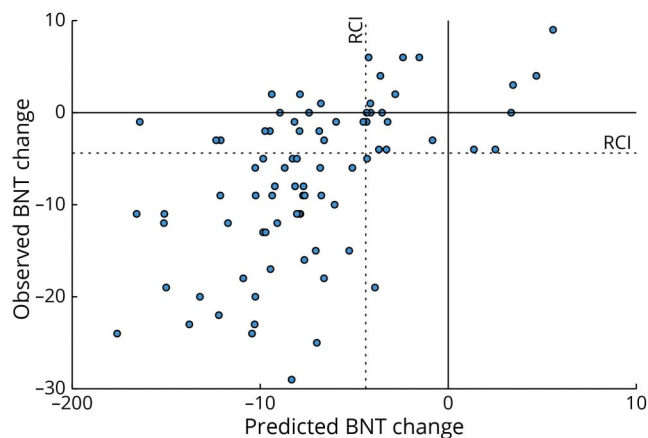
Abbreviation: aR² = adjusted R².

epilepsy onset, and education as predictive variables. It is important to note that a hierarchical multivariable regression model showed that fMRI added significant independent predictive value beyond the other predictors. These findings provide compelling evidence to support the clinical use of fMRI to assist with prediction of postsurgical naming outcome in patients with left TLE.

Prior evidence on this topic has involved small samples with somewhat inconsistent results. Sabsevitz et al.¹² studied 24 patients with left TLE using an SD-TD fMRI protocol similar to the one used here. All patients also underwent Wada language testing. The fMRI language LI was strongly correlated with outcome ($r = -0.64$, $p < 0.001$) and showed 100% sensitivity, 73% specificity, and a positive predictive value of 81% for predicting significant decline. In comparison, the Wada language LI showed a somewhat weaker correlation with decline ($r = -0.50$, $p < 0.05$), 92% sensitivity, 43% specificity, and positive predictive value of 67%. Similar fMRI results were reported by Bonelli et al.¹⁰ using a verbal fluency task in 24 patients with left TLE, although the correlation with naming outcome was more modest ($r = -0.46$, $p = 0.03$). Rosazza et al.¹¹ reported similar results ($r = -0.50$, $p = 0.04$) in a sample of 17 patients with left TLE. In contrast, Audrain et al.¹³ found no correlation between their fMRI LI, based on a combination of word production activation tasks, and BNT score change in 20 patients with left TLE. Last, You et al.¹⁴ used an auditory sentence comprehension task to measure lateralization in a temporoparietal ROI in patients with either left ($n = 20$) or right ($n = 15$) TLE. Surprisingly, stronger left lateralization was associated with less naming decline in their multivariable regression model. The explanation for these wide disparities is unclear but seems most likely to be a random outcome of the small sample sizes used in all these studies.

In contrast to previous studies, we did not find the preoperative BNT score or duration of epilepsy to be predictive of postoperative change. The lack of correlation with duration of epilepsy could be the result of changing clinical practice, with a shift toward earlier consideration of surgery. This pattern results in the time between diagnosis of epilepsy and surgical treatment becoming more homogeneous and less related to epilepsy severity. Supporting this, age at onset and age at surgery in our sample were highly correlated ($r = 0.56$)

Figure 3 Scatterplot of Predicted vs Actual Change in BNT Scores

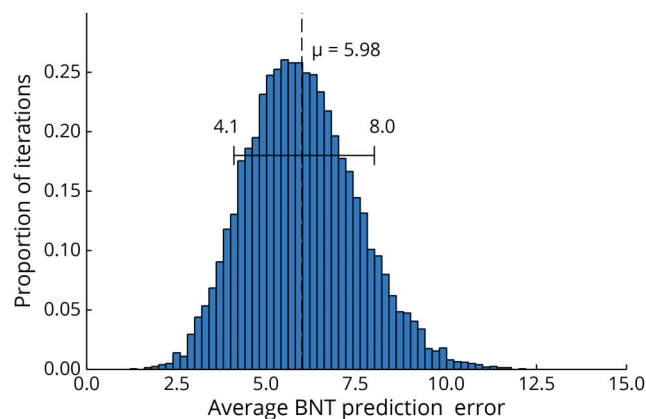


Red dotted lines indicate the reliable change index (RCI). Patients in the lower right quadrant were not predicted to have decline based on the regression model but experienced a postoperative decline, as defined by the RCI. BNT = Boston Naming Test.

and supplied mostly overlapping information to the regression model. The lack of correlation between preoperative BNT score and change score is more difficult to integrate with the previous literature. This relationship has been attributed to patients with higher preoperative performance having more functional tissue at risk and therefore a proportionally larger deficit postoperatively. While this effect has been reported in some previous studies,^{3,11} several studies have failed to find a robust correlation,^{12,14} suggesting that the effect may not be present in all samples.

Among our patients who experienced a naming decline postoperatively, 2 (4%) were unexpected according to the final prediction model. Avoiding false negatives (i.e., predicting a patient is not at risk who subsequently experiences a postoperative decline) is critical when counseling patients for surgery. While false positives (i.e., predicting a patient is at risk who subsequently has no deficit) are inaccurate, the final patient outcome if they choose to undergo surgery is better than expected, whereas with a false-negative prediction, the patient undergoes the procedure with confidence in a good outcome and subsequently experiences an unexpected deficit. Because the model presented here is based on linear regression, it has overall strength in predicting the approximate degree of deficit, but it is not optimized to make the categorical distinction of whether the patient will experience a deficit. As shown in the cross-validation analysis, a patient could experience a deficit 4 to 8 points worse than predicted, which may lead to a categorical change based on RCI for patients with smaller deficits. Notably, however, neither of the false-negative patients presented here had right-lateralized fMRI LIs. This suggests that a right-lateralized LI (although this is in the minority of cases) may allow the clinician to have strong confidence in predicting a positive outcome when the surgery is on the left.

Figure 4 Bootstrap Distribution of Average Prediction Error of the Complete Regression Model



Each model was fitted with 90% of the patients and error measured with the held back 10%. Bootstrap distribution was created by repeating this using random resampling for 10,000 iterations. Mean error is shown as the dashed line, with the range denoting the 10% and 90% quantiles. BNT = Boston Naming Test.

Our model taken as a whole, although predictive, accounts for only about one-third of the variance in naming change after surgery. This suggests that there may be other important factors that determine naming outcome. This sample consisted of patients with a relatively homogeneous surgical resection; however, residual variability in the location and extent of resection probably influences naming outcome. Our group recently published a voxel-based lesion symptom mapping study specifically associating naming decline with damage in the ventral temporal lobe.¹⁷ The critical focus was at the posterior aspect of the region typically removed in a standard anterior temporal lobectomy. Thus, variation in the posterior extent of the resection in our sample could account for some of the variance in outcome not explained by the prediction model. Overall functional status of the left temporal lobe may also be a factor. Language may be lateralized to the left side in some patients despite a severely dysfunctional left temporal lobe, with intrahemispheric reorganization compensating for the temporal lobe dysfunction. In such patients, even a large left temporal lobe resection might not produce naming decline. Supporting this hypothesis, You et al.¹⁴ provided evidence that the amount of language-related tissue removed during surgery accounts for significant variance in naming outcome. Thus, naming outcome after temporal lobe surgery is likely to reflect a complex interaction of language lateralization, functional status of the to-be-resected temporal lobe tissue, and extent and location of the resection.

The ability to predict naming outcome with preoperative fMRI could depend on several technical factors, particularly the validity of the fMRI task contrast and the algorithms used to calculate LI. Prior studies highlight the importance of incorporating sensorimotor and attention controls in the fMRI task contrast so that these systems do not dominate the activation map (i.e., enhancing specificity) and so that language systems normally active during

passive and resting states can be detected (i.e., enhancing sensitivity).^{22,33,34} The protocol used here incorporates such controls and produces activation maps that are strongly left-lateralized and sensitive to frontal, temporal, and parietal language networks in healthy controls. External validation against the intracarotid anesthesia procedure (IAP) is also an important step to ensure validity of an fMRI language lateralization protocol. The protocol used here has been validated in the largest fMRI-IAP comparison study yet published²⁴ and has been validated by correlation with verbal memory outcomes in a large series of patients undergoing left temporal lobe surgery.²⁵ In this complex, multicenter study, we elected to focus on this well-validated fMRI protocol rather than attempt to compare different protocols. We emphasize, however, the possibility that the current results may not apply to other fMRI task protocols.

Other factors that could affect validity of the fMRI LI include the mathematical method for calculating LI and the ROI chosen for the calculation. We chose a multithreshold, bootstrap sampling algorithm for LI calculation that is widely used and thought to provide more stable results over a wider activation range compared to fixed-threshold methods.²⁶ We also compared naming outcome models using different ROIs covering frontal, temporal, inferior parietal, and lateral fronto-temporo-parietal cortices, which showed little difference between models but a slight advantage for the lateral multilobar ROI used in the main analysis. This ROI may be slightly superior both because it covers a larger territory than the other ROIs and because it omits cortical regions near the midline, which can add noise to the LI calculation due to blurring of signals across the midline.

This study did not address the possible predictive value of language laterality derived from the IAP.⁹ Use of the IAP has declined in recent decades,^{35,36} and the IAP was not obtained in most of the patients in this sample, nor was it possible to enforce a standard IAP protocol across those centers that routinely perform an IAP. Although the IAP has been in clinical use for many decades, its ability to predict naming outcomes has seldom been formally assessed. One of the few studies to do so actually reported a lower risk of naming decline in patients with left language dominance undergoing left-sided surgery,³⁷ a finding that contradicts long-held assumptions about the IAP. Only 2 small studies have directly compared fMRI and IAP language tests on naming outcome prediction; in both cases, the fMRI LI, based on the same SD-TD protocol used in the present study, was more accurate than the language IAP.^{12,38} Although these studies were based on specific fMRI and IAP language lateralization protocols and may not generalize to other protocols, the generally high level of concordance between these tests (averaging ≈85% across many studies and centers²⁴) suggests that language IAP in general may not provide much independent information beyond language fMRI.

Our model incorporating an fMRI LI derived with an SD-TD task contrast was significantly predictive of naming change in patients undergoing left temporal lobe surgery, and fMRI

added independent predictive value to the model. These results provide further evidence for the role of fMRI in pre-operative prediction of naming outcomes before TLE surgery.

Acknowledgment

The authors thank the participating patients for their generous commitment of time and effort, along with the support of the many staff who helped coordinate this study, including Jeanette Albertson, Nancy Cohen, Noreen Connolly, Alexander Dagley, Sheri Davis, Lisa Ferguson, Mirjana Ivanisevic, Colleen Kenney, Megan LeDoux, Lucy Mendoza, Katherine Murphy, Gloria Novak, Brian Renner, Lisa Schwartz, and Diane Smith.

Study Funding

Collection of the data from this work was supported by NIH National Institute of Neurological Disorders and Stroke, R01 NS35929 (J.R.B.), and data analysis and manuscript preparation by NIH National Center for Advancing Translational Sciences, KL2 TR001438 (W.L.G.).

Disclosure

The authors report no disclosures relevant to the manuscript. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* March 18, 2021. Accepted in final form March 2, 2022.

Appendix 1 Authors

Name	Location	Contribution
William Gross, MD, PhD	Medical College of Wisconsin, Madison	Manuscript preparation, study design, and analysis
Alexander I. Helfand, PhD	Medical College of Wisconsin, Madison	Analysis or interpretation of data
Sara Swanson, PhD	Medical College of Wisconsin, Madison	Manuscript preparation and analysis
Lisa Conant, PhD	Medical College of Wisconsin, Madison	Manuscript preparation and analysis
Colin J. Humphries, PhD	Medical College of Wisconsin, Madison	Study design, acquisition of data, and analysis
Manoj Raghavan, MD, PhD	Medical College of Wisconsin, Madison	Manuscript preparation, study design, and acquisition of data
Wade Mueller, MD	Medical College of Wisconsin, Madison	Study design and acquisition of data
Robyn Busch, PhD	Cleveland Clinic, OH	Manuscript preparation and acquisition of data
Linda Allen, BSN, RN	Medical College of Wisconsin, Madison	Acquisition of data
Christopher Anderson, MD	Medical College of Wisconsin, Madison	Acquisition of data

Appendix 1 (continued)

Name	Location	Contribution
Chad Carlson, MD	Medical College of Wisconsin, Madison	Acquisition of data
Mark Lowe, PhD	Cleveland Clinic, OH	Acquisition of data
John Langfitt, PhD	University of Rochester, NY	Acquisition of data
Madalina Tivarus, PhD	University of Rochester, NY	Acquisition of data
Daniel Drane, PhD	University of Washington, Seattle	Manuscript preparation and acquisition of data
David Loring, PhD	Emory University, Atlanta, GA	Acquisition of data
Monica Jacobs, PsyD	Vanderbilt University, Nashville, TN	Acquisition of data
Victoria Morgan, PhD	Vanderbilt University, Nashville, TN	Manuscript preparation and acquisition of data
Jane Allendorfer, PhD	University of Alabama at Birmingham	Acquisition of data
Jerzy Szafarski, MD, PhD	University of Alabama at Birmingham	Manuscript preparation and acquisition of data
Leonardo Bonilha, MD, PhD	Medical University of South Carolina, Charleston	Acquisition of data
Susan Bookheimer, PhD	University of California, Los Angeles	Acquisition of data
Thomas Grabowski, MD	University of Washington, Seattle	Acquisition of data
Jennifer Vannest, PhD	Cincinnati Children's Hospital, OH	Acquisition of data
Jeffrey Binder, MD	Medical College of Wisconsin, Madison	Manuscript preparation, acquisition of data, study concept, and analysis

Appendix 2 Coinvestigators

Name	Location	Role	Contribution
Bassel Abou-Khalil, MD	Vanderbilt University	Site investigator	Study group member
Ellen Air, MD, PhD	University of Cincinnati	Site investigator	Study group member
Erik Beall, PhD	Cleveland Clinic	Site investigator	Study group member
Michael Berg, MD	University of Rochester	Site investigator	Study group member
William Bingaman, MD	Cleveland Clinic	Site investigator	Study group member
Truman Brown, PhD	Medical University of South Carolina	Site investigator	Study group member

Appendix 2 (continued)

Name	Location	Role	Contribution
Jonathan Edwards, MD	Medical University of South Carolina	Site investigator	Study group member
Steven Glazier, MD	Medical University of South Carolina	Site investigator	Study group member
Jorge Gonzalez-Martinez, MD	Cleveland Clinic	Site investigator	Study group member
Xiaoping Hu, PhD	Emory University	Site investigator	Study group member
Peter Konrad, MD, PhD	Vanderbilt University	Site investigator	Study group member
Imad Najm, MD	Cleveland Clinic	Site investigator	Study group member
Andrew Nencka, PhD	Medical College of Wisconsin	Site investigator	Study group member
Jeffrey Ojemann, MD	University of Washington	Site investigator	Study group member
Vaishali Phatak, PhD	University of Washington	Site investigator	Study group member
Webster Pilcher, MD	University of Rochester	Site investigator	Study group member
Michael Privitera, MD	University of Cincinnati	Site investigator	Study group member
Kristen Riley, MD	University of Alabama at Birmingham	Site investigator	Study group member
Fred Sabb, PhD	University of California, Los Angeles	Site investigator	Study group member
William Vandergrift III, MD	Medical University of South Carolina	Site investigator	Study group member
Mark Wagner, PhD	Medical University of South Carolina	Site investigator	Study group member
Hwa Shaun Yeh, MD	University of Cincinnati	Site investigator	Study group member

References

- Wiebe S, Blume WT, Girvin JP, Eliasziw M. A randomized, controlled trial of surgery for temporal-lobe epilepsy. *N Engl J Med*. 2001;345(5):311-318.
- Télez-Zenteno JF, Dhar R, Wiebe S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. *Brain*. 2005;128:1188-1198.
- Busch RM, Floden DP, Prayson B, et al. Estimating risk of word-finding problems in adults undergoing epilepsy surgery. *Neurology*. 2016;87(22):2363-2369.
- Sherman EM, Wiebe S, Fay-McClymont TB, et al. Neuropsychological outcomes after epilepsy surgery: systematic review and pooled estimates. *Epilepsia*. 2011;52(5):857-869.
- Perrine K, Hermann BP, Meador KJ, et al. The relationship of neuropsychological functioning to quality of life in epilepsy. *Arch Neurol*. 1995;52(10):997-1003.
- Davies KG, Risse GL, Gates JR. Naming ability after tailored left temporal resection with extraoperative language mapping: increased risk of decline with later epilepsy onset age. *Epilepsy Behav*. 2005;7(2):273-278.
- Ruff IM, Swanson SJ, Hammeke TA, Sabsevitz D, Mueller WM, Morris GL. Predictors of naming decline after dominant temporal lobectomy: age at onset of epilepsy and age of word acquisition. *Epilepsy Behav*. 2007;10(2):272-277.
- Davies KG, Bell BD, Bush AJ, Hermann BP, Dohan FC, Jaap AS. Naming decline after left anterior temporal lobectomy correlates with pathological status of resected hippocampus. *Epilepsia*. 1998;39(4):407-419.
- Wada J, Rasmussen T. Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observations. *J Neurosurg*. 1960;17:266-282.
- Bonelli SB, Thompson PJ, Yogarajah M, et al. Imaging language networks before and after anterior temporal lobe resection: results of a longitudinal fMRI study. *Epilepsia*. 2012;53(4):639-650.

11. Rosazza C, Ghielmetti F, Minati L, et al. Preoperative language lateralization in temporal lobe epilepsy (TLE) predicts peri-ictal, pre- and post-operative language performance: an fMRI study. *Neuroimage Clin*. 2013;3:73-83.
12. Sabsevitz DS, Swanson SJ, Hammeke TA, et al. Use of preoperative functional neuroimaging to predict language deficits from epilepsy surgery. *Neurology*. 2003;60(11):1788-1792.
13. Audrain S, Barnett AJ, McAndrews MP. Language network measures at rest indicate individual differences in naming decline after anterior temporal lobe resection. *Hum Brain Mapp*. 2018;39:4404-4419.
14. You X, Zachery AN, Fanto EJ, et al. fMRI prediction of naming change after adult temporal lobe epilepsy surgery: activation matters. *Epilepsia*. 2019;60(3):527-538.
15. Szaflarski JP, Gloss D, Binder JR, et al. Practice guideline summary. *Neurology*. 2017;88:395-402.
16. Schmid E, Thomschewski A, Taylor A, et al. Diagnostic accuracy of functional magnetic resonance imaging, Wada test, magnetoencephalography, and functional transcranial Doppler sonography for memory and language outcome after epilepsy surgery: a systematic review. *Epilepsia*. 2018;59(12):2305-2317.
17. Binder JR, Tong JQ, Pillay SB, et al. Temporal lobe regions essential for preserved picture naming after left temporal epilepsy surgery. *Epilepsia*. 2020;61(9):1939-1948.
18. Drane DL, Loring DW, Voets NL, et al. Better object recognition and naming outcome with MRI-guided stereotactic laser amygdalohippocampotomy for temporal lobe epilepsy. *Epilepsia*. 2015;56:101-113.
19. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996;29:162-173.
20. Binder JR, Gross WL, Allendorfer JB, et al. Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. *NeuroImage*. 2011;54(2):1465-1475.
21. Friedman L, Glover GH, Krenz D, Magnotta V, BIRN TF. Reducing inter-scanner variability of activation in a multicenter fMRI study: role of smoothness equalization. *Neuroimage*. 2006;32(4):1656-1668.
22. Binder JR, Swanson SJ, Hammeke TA, Sabsevitz DS. A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia*. 2008;49(12):1980-1997.
23. Binder JR, Swanson SJ, Hammeke TA, et al. Determination of language dominance using functional MRI: a comparison with the Wada test. *Neurology*. 1996;46(4):978-984.
24. Janecek JK, Swanson SJ, Sabsevitz DS, et al. Language lateralization by fMRI and Wada testing in 229 patients with epilepsy: rates and predictors of discordance. *Epilepsia*. 2013;54(2):314-322.
25. Binder JR, Sabsevitz DS, Swanson SJ, Hammeke TA, Raghavan M, Mueller WM. Use of preoperative functional MRI to predict verbal memory decline after temporal lobe epilepsy surgery. *Epilepsia*. 2008;49(8):1377-1394.
26. Wilke M, Schmithorst VJ. A combined bootstrap/histogram analysis approach for computing a lateralization index from neuroimaging data. *NeuroImage*. 2006;33(2):522-530.
27. Springer JA, Binder JR, Hammeke TA, et al. Language dominance in neurologically normal and epilepsy subjects: a functional MRI study. *Brain*. 1999;122(pt 11):2033-2046.
28. Knecht S, Flöel A, Dräger B, et al. Degree of language lateralization determines susceptibility to unilateral brain lesions. *Nat Neurosci*. 2002;5(7):695-699.
29. Knecht S, Dräger B, Deppe M, et al. Handedness and hemispheric language dominance in healthy humans. *Brain*. 2000;123(pt 11):2512-2518.
30. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. 1991;59:12-19.
31. Hermann BP, Seidenberg M, Schoenfeld J, Peterson J, Leveroni C, Wyler AR. Empirical techniques for determining the reliability, magnitude, and pattern of neuropsychological change after epilepsy surgery. *Epilepsia*. 1996;37(10):942-950.
32. Sawrie SM, Chelune GJ, Naugle RL, Lüders HO. Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. *J Int Neuropsychol Soc*. 1996;2:556-564.
33. Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Rao SM, Cox RW. Conceptual processing during the conscious resting state. A functional MRI study. *J Cogn Neurosci*. 1999;11:80-95.
34. Binder JR. FMRI of language systems. In: Filippi M, ed. *Neuroimaging: Methods and Protocols*, 2nd ed. Humana Press; 2016:355-386.
35. Baxendale S, Thompson PJ, Duncan JS. The role of the Wada test in the surgical treatment of temporal lobe epilepsy: an international survey. *Epilepsia*. 2008;49(4):715-5.
36. Rausch R, HS, H-GW, CB D, KJ M, MJ-G. Intraarterial amobarbital procedures. In: Engel J, ed. *Surgical Treatment of the Epilepsies*, 2nd ed. New York: Raven Press; 1993:341-357.
37. Kovac S, Möddel G, Reinholz J, et al. Visual naming performance after ATL resection: impact of atypical language dominance. *Neuropsychologia*. 2010;48(7):2221-2225.
38. Janecek JK, Swanson SJ, Sabsevitz DS, et al. Naming outcome prediction in patients with discordant Wada and fMRI language lateralization. *Epilepsy Behav*. 2013;27(2):399-403.