Published in final edited form as:

Ann Neurol. 2017 April ; 81(4): 526–537. doi:10.1002/ana.24899.

Language fMRI and Direct Cortical Stimulation in Epilepsy Preoperative Planning

Alison Austermuehle, BA¹, John Cocjin, BS², Richard Reynolds, MS³, Shubhi Agrawal, MD¹, Leigh Sepeta, PhD¹, William D. Gaillard, MD¹, Kareem Zaghloul, MD PhD², Sara Inati, MD⁴, and William H. Theodore, MD¹

¹Clinical Epilepsy Section, National Institute of Neurological Disorders and Stroke

²Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke

³Scientific and Statistical Computing Core, National Institute of Mental Health

⁴Electroencephalography Section, National Institute of Neurological Disorders and Stroke

Abstract

Objective—Presurgical language assessment can help minimize damage to eloquent cortex during resective epilepsy surgery. Two methods for presurgical language mapping are functional MRI (fMRI) and direct cortical stimulation (DCS) of implanted subdural electrodes. We compared fMRI results to DCS to help optimize non-invasive language localization and assess its validity.

Methods—We studied 19 patients referred for presurgical evaluation of drug-resistant epilepsy. Patients completed 4 language tasks during preoperative fMRI. After subdural electrode implantation, we used DCS to localize language areas. For each stimulation site, we determined whether language positive electrode pairs intersected with significant fMRI activity clusters for language tasks.

Results—Sensitivity and specificity depended on electrode ROI radii and statistical thresholding. For patients with at least one language positive stimulation site, an auditory description decision task provided the best tradeoff between sensitivity and specificity. For patients with no language positive stimulation sites, fMRI was a dependable method of excluding eloquent language processing.

Interpretation—Language fMRI is an effective tool for determining language lateralization prior to electrode implantation, and is especially useful for excluding unexpected critical language areas. It can help guide subdural electrode implantation and narrow the search for eloquent cortical areas via DCS.

CORRESPONDING AUTHOR: William H. Theodore, Building 10 Room 7D-43, Bethesda MD 20892-1408, Phone: 301-496-1505, Fax: 301-402-2871, theodorw@ninds.nih.gov.

Author Contributions

AA and WHT designed the study, conducted the main analysis, and were the primary composers of the manuscript. JC, SA, SKI, RR and WDG assisted with portions of data analysis and assisted in drafting the manuscript. LS, SKI and KZ participated in data acquisition.

Conflicts of Interest

The authors have no conflicts of interest to disclose.

Introduction

Presurgical language assessment can help minimize perioperative damage to eloquent cortex. Two common methods for language mapping are functional magnetic resonance imaging (fMRI) and direct cortical stimulation (DCS) through subdural electrodes. Extraoperative mapping with DCS is the clinical ‘gold standard’ for localizing critical language areas, but is invasive, costly and adds additional risk from bleeding and infection.¹ fMRI provides reliable, non-invasive language lateralization for adults and children with epilepsy.^{2,3} Current clinical practice parameters indicate that fMRI may be considered for lateralizing language function in place of IAP in temporal lobe epilepsy patients; however, the use of fMRI for localization in clinical settings is not well established.^{4,5}

Several small studies compared fMRI with intraoperative and extraoperative cortical stimulation, with little agreement about concordance between language maps generated by fMRI and stimulation mapping, and wide variation in reported fMRI sensitivity and specificity. Most studies agree that optimal and standardized statistical thresholds for evaluating fMRI activation are uncertain.^{6,7,8}

We examined concordance between fMRI and DCS in patients undergoing epilepsy surgery evaluation. By co-registering implanted subdural electrodes and fMRI results to preoperative MRI, we compared fMRI language activation with extraoperative DCS language mapping. We sought to identify the best approach for comparing these tools, and determine whether standardized methods could be applied across subjects’ fMRI data to detect the presence and absence of DCS-defined critical language areas. We evaluated the utility of various language fMRI paradigms as screening tools for detecting eloquent cortex, guiding electrode implantation, and narrowing the search for language via DCS.

Methods

Participants

Nineteen patients with medically refractory epilepsy who underwent presurgical evaluation and intracranial EEG monitoring at the National Institutes of Health between 2012 and 2016 were included (age at date of implantation: range = 22.6–59.9 years, median = 33.3 years; 14 females) (Table 1). All patients underwent fMRI using echo-planar imaging (EPI) blood oxygen level-dependent (BOLD) techniques during 4 language tasks, obtained 15 to 218 days prior to electrode implantation (mean = 109.2 days). Semiology, neurologic examination, ictal video electroencephalography, high-resolution MRI and [18F]-FDG PET were used to localize seizure foci and guide subdural electrode implantation. T1-weighted MPRAGE images were collected preoperatively and CT scans following implantation. All patients underwent electrocortical stimulation mapping to localize language, motor, and sensory function depending on their seizure focus and implant location. Sixteen patients underwent a subsequent resection. Approval by the NIH Combined Neurosciences Institutional Review Board was obtained. Informed consent was obtained from all patients.

fMRI acquisition and analyses

Eighteen patients completed four tasks: ADDT, AUDCAT, listening, and reading. One did not complete AUDCAT. Each paradigm used a block design composed of five epoch cycles. Each cycle consisted of an active alternating with a condition condition; each hemicycle lasted 30 seconds. The paradigms have been described previously.⁹ Briefly:

- A.** Auditory Description Decision Task (ADDT): Participants hear an auditory clue (5–6 word sentences) and are instructed to push a button when the description is true. Twenty of 30 items are correct targets (“Something that hangs in a museum is a painting”); 10 of 30 are foils (“Something that parts your hair is a refrigerator”).
- B.** Auditory Categories (AUDCAT): Participants are presented with categories (“Food”) followed by a list of nouns, and are instructed to designate whether each presented noun falls into the previously presented category (“pancake”) or not (“manpower”) by pushing a button.
- C.** Listening: Participants listen to stories and are instructed to push a button on hearing a tone, presented intermittently throughout the story.
- D.** Reading: Participants read a story presented on a screen and are instructed to press a button when they reach the period following each sentence (active condition) and for white, but not black, squares presented on the screen (control condition).

The baseline conditions were designed to match the experimental conditions for primary audition, vision, motor response, length of utterance, and volume. During each auditory task, the control condition consisted of reverse speech where patients were instructed to push a button on hearing a tone. Task performance was evaluated by the overall accuracy for the ADDT task (true positives/sum of correct answers possible + number of commissions $\times$ 100)¹⁰. Data were available for six of 19 subjects: overall accuracy was 85.8% ($\pm$ 13.2%). Average response time was 3014.4 ms ($\pm$ 195.9 ms).

All patients were scanned on a 3.0 Tesla GE scanner using echo-planar imaging (EPI) blood oxygen level-dependent (BOLD) techniques. Gradient echoplanar images were collected using TE=30 msec, flip angle = 65°, FOV=22 $\times$ 22 cm, acquisition matrix=64 $\times$ 64, and interscan interval (TR)=2000 msec. Voxel sizes were either: 3.4 $\times$ 3.4 $\times$ 4.0 mm with brain volumes consisting of 30 $\times$ 4-mm-thick axial slices or 3.0 $\times$ 3.0 $\times$ 3.0 mm with brain volumes consisting of 40 $\times$ 3-mm-thick axial slices. Images were collected parallel to the anterior commissure-posterior commissure plane.

Functional data were analyzed within the framework of the publicly available software package AFNI (Analysis of Functional NeuroImages)¹¹ in each patient’s native space. For preprocessing, echo-planar images were resampled to 4.0 mm isotropic voxels for consistency across scan types (performed along with registration, avoiding additional interpolation). Each of 4 runs was input as a single 4D dataset of 150 time points. The pre-steady state volumes, collected before equilibrium magnetization was reached, were discarded automatically during acquisition. Slice timing correction was performed to

synchronize timing across slices. Next, the EPI registration base volume was aligned to a skull-stripped, deobliqued preoperative T1 weighted MPRAGE using a Local Pearson correlation absolute value (LPA) cost function. For three subjects, the LPA cost function failed and a Local Pearson Correlation (LPC) cost function was used. The remaining volumes in the EPI dataset were aligned to the MPRAGE base volume with a concatenated transformation. Voxels without data at every time point were masked out.

Each volume of each run was spatially blurred using a 4-mm full-width at half-maximum (FWHM) Gaussian kernel within the EPI extent mask. De-measured motion parameters were computed for use in regression and motion was regressed per individual task run. Motion above 0.3 mm was censored, based on the Euclidean norm of motion derivatives. We excluded runs during which the fraction of time points censored due to motion exceeded 0.2.

A generalized least squares time series fit with restricted maximum likelihood estimation (REML) of temporal autocorrelation was used for regression using a voxel-wise ARMA(1,1) model. The regression model provided a single estimate of the response magnitude to each stimulus type in each voxel for each subject. Correction for multiple comparisons was based on Monte Carlo simulation using AFNI's 3dClustSim.¹² The estimated probability of false positive detection per image, based on the frequency count of cluster sizes, was used to establish cluster size thresholds for three uncorrected p values, in order to achieve corrected $p < 0.05$.

Laterality Index

Hemispheric laterality indices were calculated by determining the number of voxels included in left and right hemisphere fMRI masks for each threshold. LI was calculated and classified using the following conventions¹³:

$$LI = \frac{Voxels_{left} - Voxels_{right}}{Voxels_{left} + Voxels_{right}}$$

$$\begin{aligned} LI < -0.2 &= Right \\ -0.2 < LI < 0.2 &= Bilateral \\ LI > 0.2 &= Left \end{aligned}$$

If a subject's laterality classification differed across thresholds for a task, the more common classification was chosen (e.g. if a subject was bilateral at P1, but left-lateralized at P2 and P3, they were classified as left) (Table 2).

Electrode localization

For each patient, electrode centers identified in the post-implantation CT scan were brought to the T1 MPRAGE space via a 12 parameter affine transform implemented in AFNI, and corrected for brain shift using a modified elastic algorithm^{14,15} which brought constrained subdural electrode locations to a reconstructed outer-smoothed pial surface using Freesurfer suite routines (<http://surfer.nmr.mgh.harvard.edu>) and custom scripts in the MATLAB

environment. The algorithm was constrained by a subset of electrode locations identified via intraoperative photograph and designated in 3D-Slicer (<http://www.slicer.org>). Gyral anatomy and vascular patterns were used as landmark references¹⁵ (Fig 1). Electrode coordinates were transformed into ROIs to be overlaid on fMRI activation maps.

Stimulation mapping

Stimulation parameters were: pulse width 0.5ms, pulse rate 50Hz, duration 3-5 seconds. Current amplitude increased from 2 mA to a maximum of 16 mA in 1 mA increments. Adjacent electrodes (1 cm apart) were stimulated in a bipolar manner to screen for language function. In some instances, electrode pairs that elicited positive responses were tested individually using a distant, cleared electrode as reference. The following tasks were used:

- A. Countinous counting or alphabet recitation until either a reproducible functional change occurred (such as a pause or dysarthria), after-discharges observed on EEG, or no response elicited at 16 mA. Subsequent tasks were completed using this threshold (or 1 mA lower when after-discharges appeared).
- B. Visual naming: patients looked at pictures of objects and named them aloud.
- C. Reading: patients read short sentences aloud.
- D. Token test: patients pointed at objects given the experimenter's verbal description (*Ex*: "Point to the small, red circle").

Electrode pairs which reproducibly disrupted either language production or comprehension at any stimulation amplitude, during any task, were defined as *language positive* and considered to be overlying a critical language area. In cases where electrodes were tested monopolarly using a reference electrode, the single electrode causing a disruption was defined as language positive, and modeled as a single ROI.

Electrode pairs cleared of language function during all tasks were defined as *clear*. This occurred when either no response was elicited at 16 mA or the stimulation was pain-limited but caused no language deficits at a lower threshold. Electrodes limited by pain at a low threshold were excluded, as the language task could not produce reliable results. We eliminated electrodes producing sensory or motor effects, after-discharges, or electrographic seizures upon stimulation; language disruption occurring during that time could not be attributed to stimulation alone. We excluded subtemporal electrodes, as this region is subject to distortion and signal loss during fMRI.^{16,17}

Comparison algorithm

In order to compare the results of DCS with activation maps produced by fMRI, we created ROIs based on projected electrode coordinates to represent both cleared and language positive electrode pairs. To account for uncertainty in the effective radius of DCS propagation,^{18,7} we evaluated ROI sizes of 4, 6, 8, and 10 mm radii. Ten mm radii are commonly used in the existing literature;^{19–20} this radius is chosen based on reports that more permanent language deficits occur if the distance of the resection margin from the nearest language site is less than 1 cm.²¹ Electrodes stimulated bipolarly were modeled as a single ROI, composed of two intersecting spheres of the four possible radius sizes.

Next, clusters from the ARMA regression results were detected: at a given uncorrected voxelwise threshold, a list of voxel clusters was made, keeping only those as large as those achieving the above corrected $p < 0.05$ from the Monte Carlo simulations. Different minimum cluster sizes were computed for each uncorrected p-value ($p=0.05$, 0.01 and 0.001), though with each achieving the corrected $p < 0.05$ for the clusters (Fig 2). This range was chosen to evaluate the tradeoff between sensitivity of a liberal threshold (minimizing false negatives) and specificity of a strict threshold (minimizing false positives) when making comparisons between fMRI and DCS. We refer to the thresholds as P1 ($P=0.05$), P2 ($P=0.01$), and P3 ($P=0.001$).

The clustered fMRI activity mask at each uncorrected p value was systematically overlaid on ROI masks of each size and stimulation site (resulting in 12 permutations: 3 thresholds $\times$ 4 ROI sizes). The algorithm determined if any intersections between the two datasets existed and the number of intersecting voxels.

To characterize further fMRI utility for localizing language critical areas, we created task activation maps combining tasks: ADDT $\times$ AUDCAT $\times$ Listening $\times$ Reading ($n=10$), ADDT $\times$ AUDCAT ($n=14$), and ADDT $\times$ Reading ($n=12$) for patients showing significant activation across all thresholds for each task used. These task combinations were chosen based on the sensitivity, specificity, and target activation region of the individual tasks. The cluster corrected statistical overlays for each task included in a mask were multiplied using AFNI's 3dcalc command. For a given combination mask, only voxels that were contained in a significant cluster across tasks were included; clusters present in one task but not another were eliminated from the mask (Fig 2).

We considered intersections between significant fMRI clusters and language positive electrode ROIs to be true positives (TP), and intersections between fMRI clusters and cleared electrode ROIs false positives (FP). When a language positive electrode ROI did not intersect with a cluster, it was considered a false negative (FN). When a cleared electrode ROI did not intersect with an fMRI cluster, it was considered a true negative (TN). We calculated sensitivity, specificity, accuracy, and odds ratio for each combination of threshold and ROI parameters to evaluate fMRI test performance compared to DCS. fMRI and DCS were acquired and processed independently without reference to each other by investigators blinded to the alternative dataset.

Results

Patients

We divided our patients into two subgroups: those with at least one language positive stimulation site on DCS ($n = 12$), and those with none ($n = 7$). Ten language-positive patients had left temporal electrode coverage, one right temporal, and one bilateral. Four of 7 patients with no positive language findings had right temporal coverage, one right central/parietal, one left central/parietal, and one bilateral coverage. Across all patients, we considered 190 electrode pairs and 47 single electrodes (427 electrodes total; 278 in the DCS+ group, 149 in the DCS- group). Of those, 47 electrodes produced positive language responses on stimulation (during any task), and were considered language positive.

For all patients in the language positive group, at least one of 4 fMRI tasks showed lateralization matching the hemisphere where DCS identified eloquent cortex. In one patient, an fMRI task (listening) lateralized to the opposite hemisphere (right) as DCS localized; this patient was ambidextrous and showed left activation of AUDCAT and bilateral activation for reading. Detailed laterality index results are presented in Table 2.

Ten of 16 patients who underwent a resection completed a postoperative neuropsychology battery. For 8 patients, postoperative testing was completed at a one year follow-up (average = 375.1 days); 2 patients completed the testing at a two-year follow-up (811 and 1000 days postoperatively). To assess changes in language function, we compared preoperative and postoperative scores on the 60-item Boston Naming test (BNT), F-A-S total corrected score, and Verbal Comprehension IQ (VCI) and Full Scale IQ (FSIQ) from the Wechsler Adult Intelligence Scale III (WAIS III). Change scores (postoperative minus preoperative) were calculated and classified as either improved, no change, or declined, using established reliable change indices for epilepsy.²² One patient showed significant declines on F-A-S, VIC, and FSIQ. One patient showed a decline on BNT, VCI, and FSIQ. One patient showed a decline on the BNT. All other patients showed either improvement or no change across tests. Findings are summarized in Table 1.

Exclusions

We developed an algorithm to detect intersections between a series of overlapping masks: functional MRI activation masks and electrode ROI masks for stimulation sites. We created activation masks at three thresholds. We excluded patients for whom no clusters reached statistical significance across all three thresholds for a given task. If significance was reached at any threshold, the patient was included for that task, and the runs for which no clusters reached significance were considered “misses” or “true negatives” depending on the electrode designation. For example, if significant clusters were achieved at P1 for the ADDT task, but not at P2 or P3, any language positive electrodes were considered “missed” at P2 and P3.

Of the DCS positive subgroup, two patients were excluded from ADDT, one from AUDCAT, and one from listening because they did not show significant activation at any threshold. Two patients were excluded from reading based on motion exclusion criteria. One DCS negative subgroup patient was excluded from AUDCAT because she did not complete the task.

Test performance of fMRI

Supplementary table 1 shows true positives, false negatives, false positives, and true negatives for each task at each electrode radius and thresholds. After exclusions, the number of electrodes considered for ADDT, AUDCAT, listening, and reading were 361, 375, 391, and 382 respectively.

To evaluate the utility of each task for identifying critical language areas defined by DCS, we calculated sensitivity, specificity, and accuracy for each combination of parameters (sensitivity was calculated as $TP/(TP+FN)$, specificity was calculated as $TN/(TN+FP)$, and accuracy was calculated as $TP+TN/(TP+FN+FP+TN)$). Supplementary table 2 shows values

obtained at each combination of parameters for each task. As expected, sensitivity increased with increasing radius size and decreased with stricter thresholding; the opposite was true for specificity. In addition, we used the diagnostic odds ratio (DOR) as a single indicator of test performance (DOR was calculated as $TP \cdot TN / (FP \cdot FN)$) (supplementary table 2). The DOR represents the ratio of the odds of positivity (of the test = fMRI) in diseased (diseased = a positive DCS stimulation site) relative to the odds of positivity (fMRI) in the nondiseased (DCS-).²³ Higher values indicate better test performance. A value of 1 indicates that the test (fMRI) does not discriminate between positive and negative DCS outcomes.

Because we wished to evaluate fMRI as a screening tool, we placed value on test parameters that provided the highest sensitivities. Across tasks and thresholds, the 10 mm radius provided the highest sensitivity. For both ADDT and AUDCAT, sensitivity was maximal at the 10 mm radius and intermediate threshold, P2. This combination offered 80.6% sensitivity, 72.7% specificity, 73.4% accuracy, and DOR of 11.1 for ADDT. For the AUDCAT task, a sensitivity of 78.4%, specificity of 67.8%, accuracy of 68.8%, and DOR of 7.6 were found with the same parameters. Our observation of improved performance at P2 was two-fold: as expected, fewer clusters survived P3 due to the stricter threshold. At P1, the number of voxels required for a cluster to reach significance was higher. A P1 activation mask could miss electrode intersections if clusters did not meet the minimum voxel size.

Overall, our listening and reading paradigms were less robust in lateralizing frontal and temporal activation on fMRI, and often showed bilateral activation (Table 2). Listening and reading sensitivities were both optimized at radius=10 mm: listening sensitivity was 59.5% (specificity 64.4%) at P1; reading had sensitivity 95.1% (specificity 35.8%) at P2.

For patients with at least one DCS site, reading provided the highest sensitivities overall, with low specificities. ADDT produced the best diagnostic odds ratio overall (Fig 3). For patients with no language positive stimulation sites, ADDT provided the highest overall specificity, ranging from 85.9% (at the largest radius and P1), to 97.3% (at the smallest radius and P3), indicating that fMRI may be useful for confirming that certain regions are not eloquent. Reading was relatively nonspecific overall in both subgroups.

For the 10 patients included in the combined activation masks of all four tasks, the 10 mm radius and fMRI threshold P1 offered sensitivity of 70.6% and specificity of 89.2%. The optimal set of parameters for ADDT $\times$ AUDCAT masks was 10 mm radius and P2, showing a sensitivity/specificity tradeoff of 74.1/82.4%. The ADDT $\times$ Reading mask, when thresholded at P1 and evaluated with a 10 mm radius, had sensitivity of 81.0% and specificity of 70.6%. All three combination masks provided high specificity, with values ranging from 70.6% to 99.1% (Fig 4).

Discussion

For fMRI to be an effective screening tool, high sensitivity is critical. If fMRI fails to detect any critical language region found on DCS, it provides an inaccurate or incomplete picture for surgical decisions independent of DCS.

We found only moderate sensitivity, but relatively high specificity, when evaluating fMRI as a tool for detection of electrode locations associated with language impairment. Because the number of language positive electrodes was small (47) compared to the total number tested (427), we did not place too much emphasis on the high specificity of fMRI, especially in patients with no language areas on DCS, as the large number of true negatives could inflate test performance. Additionally, high sensitivity (TPs compared to FNs) was often accompanied by low specificity (FPs compared to TNs). We used a diagnostic odds ratio to evaluate fMRI performance based on a single statistic. This statistic aggregated sensitivity and specificity, providing a single number to evaluate each set of parameters.

Previous investigators used varying methods to compare fMRI language mapping with intraoperative stimulation mapping, with mixed results. Sensitivity and specificity were reported to be 100%/68% for Broca's area and 64%/85% for Wernicke's area when DCS sites and 3D fMRI images were considered concordant if separated by less than 1 cm.²⁴ Other studies reported similarly high sensitivities and specificities for these regions.^{25,26,7} One study compared co-registered fMRI activation maps to intraoperative images, and showed sensitivity 75% and specificity 68% for 5 patients who underwent language mapping.⁸

Other reports found less reliable concordance between fMRI and intraoperative mapping. Coregistering intraoperative photographs of language tags with fMRI, using four language tasks in 11 patients, led to sensitivity ranging from 33%–92% and specificity ranging from 40–80% depending on task; using a battery of tasks improved results.¹⁹ One study reported sensitivity ranging from 22% (naming) to 36% (verb generation), improving to 59% sensitivity/97% specificity when the two tasks were combined.²⁷ In 27 patients, visual comparisons were made between fMRI activation areas and positive intraoperative sites; a visual correlation was found in only 6 of 14 patients.²⁸ Obstacles to comparing intraoperative mapping and fMRI include obtaining reliable intraoperative mapping results, coregistering modalities, and making visual comparisons.

Studies comparing extraoperative mapping and fMRI have also used varied approaches.^{29,30} fMRI results in 4 children with language positive DCS were compared on a lobar basis without direct coregistration; concordance varied depending on whether fMRI was assessed by visual inspection or laterality index.³¹ Studies examining concordance between extraoperative subdural electrode mapping report sensitivity/specificity trade-offs ranging from 50%/80% (9 patients)⁶ to 64%/77% (6 patients).²⁰

We studied a larger sample of adult epilepsy patients with an extensive battery of fMRI language tasks. We used stimulation results obtained extraoperatively, allowing retesting of ambiguous sites over several days. We performed testing in several small sessions so fatigue did not confound patients' performance. Our electrode localization algorithm corrected for edema and tissue shift post implantation, leading to improved coregistration. This method eliminates the need for visual comparisons between fMRI and DCS based on anatomical landmarks, which many previous studies depended on.

We aimed to define the parenchymal volume affected by DCS by constructing ROIs based on stimulation fields, and assessed overlap with positive voxels identified on fMRI. Given incomplete understanding of the extent of spread of electrical stimulation and standardized thresholding of functional data, we tested concordance at different thresholds for both modalities. We believe this provides a better measure of inter-modality concordance.

fMRI activation levels varied across patients and tasks, agreeing with previously reported observations.¹⁹ Several studies reported that concordance rates depend heavily on individualized statistical thresholds.^{6,8} To account for this variation, we developed a method for evaluating fMRI activation that did not depend on visual inspection or subjective thresholding, determining significant clusters for each patient at a standardized corrected p-value across three thresholds.

One limitation of our study, and much existing literature, is the use of disparate tasks to compare the two methods. The difference in fMRI and DCS tasks may help to explain our findings. Previous reports indicate that various fMRI tasks generally identify limited and specific aspects of language processing. During DCS, object naming was conducted by visually presenting objects for the patient to name. ADDT required patients to listen to aurally presented sentences, retrieve semantic knowledge, and make decisions.³² It is a complex task that has been shown to activate both frontal and temporal language areas.^{33,34,35} Auditory and visual object naming may be anatomically distinct.³⁶

AUDCAT required patients to make semantic decisions; semantic decision tasks activate frontal language regions robustly with less consistent temporal activation.^{37,38,32,2,39} Neither of these tasks, or the cognitive functions they target, was directly replicated during DCS mapping sessions.

Our listening task, requiring auditory comprehension, was most similar to the token task patients performed during DCS. The reading task completed during fMRI was well aligned with the reading task patients completed during DCS. Both paradigms activate temporal cortical receptive regions.^{33,40,41,42,9,43,44}

Positive DCS areas were not differentiated by specific language functions (if stimulation disrupted any of 4 tasks, the site was considered positive). This may explain why some fMRI tasks were not sensitive compared to DCS. We attempted to account for these differences by creating combined activation masks. The ADDT $\times$ AUDCAT mask was expected to provide robust frontal language activation maps. Because we saw uniformly “strong” activation for reading across patients, we expected the ADDT $\times$ Reading mask to account for both frontal (ADDT) and temporal (ADDT, reading) language areas. However, the combined task mask may have obscured eloquent areas active during only one of the included tasks; therefore, the combination masks should be considered alongside the individual task results at each of the thresholds. Using identical tasks for both mapping modalities will likely improve test concordance.

Conclusion

We demonstrated that for stimulation sites cleared on DCS, certain fMRI tasks have a low false positive rate. For patients whose seizure foci and prospective resections are in regions atypical for language (such as the parietal lobe or nondominant hemisphere), fMRI may help confirm absence of critical language areas. Our results confirm that language fMRI is a useful screening tool for identifying critical language areas prior to surgery, and complements invasive language mapping.

Our results suggest fMRI is not sensitive enough to be used independently as a localization tool for surgical decisions. The observed discordance between DCS and fMRI can be explained by several factors. First, there is a sampling limitation in DCS. Electrodes cannot be implanted uniformly across patients, and do not cover (and therefore cannot test) all fMRI positive regions. In addition, there is an inherent difference in what the modalities test: while DCS determines areas that are necessary for language function via inhibition, fMRI identifies regions that contribute to language function during a given task via activation. There are differences in the specific facets of language function that each task tests; fMRI and DCS tasks targeted specific regions, some of which do not overlap. The majority of DCS language deficits we observed were induced at the maximally tested current amplitude (16 mA), possibly affecting underlying white matter tracts in addition to cortex,^{18,45,46} not identifiable by fMRI. Incorporation of diffusion tensor imaging with fMRI may provide insight into why fMRI does not detect all language areas defined by DCS, and may provide more sensitive preoperative language mapping.

Our fMRI lateralization results show close but imperfect concordance with the DCS ‘gold standard’, agreeing with previous studies comparing fMRI to IAP.⁵ We have shown that fMRI may be especially useful for excluding unexpected critical language areas prior to surgery. The possibility of fMRI providing precise language localization is of major interest for patients undergoing resective surgery for epilepsy, but technical and methodological improvements must be made before this is possible.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge the hard work of our multidisciplinary team that made this study possible, including Edythe Wiggs, Emily Whitehead, Michael Trotta, Timothy Sheehan, Ashley Zachery, Robert Joshua, Daniel Goldenholz, Aaliyah Hamidullah Thiam, Irene Dustin, and Tamika Mason. This study was supported by the National Institute of Neurological Disorders and Stroke Division of Intramural Research.

References

1. Wellmer J, Von Der Groeben F, Klarmann U, et al. Risks and benefits of invasive epilepsy surgery workup with implanted subdural and depth electrodes. *Epilepsia*. 2012; 53(8):1322–1332. [PubMed: 22708979]
2. 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. [PubMed: 8780076]

3. Fernandez G, Specht K, Weis S, et al. Intrasubject reproducibility of presurgical language lateralization and mapping using fMRI. *Neurology*. 2003; 60(6):969–975. [PubMed: 12654961]
4. 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. [PubMed: 23294162]
5. Szaflarski JP, Gloss D, Binder JR, et al. Practice guideline summary: Use of fMRI in the presurgical evaluation of patients with epilepsy. *Neurology*. 2017; doi: 10.1212/WNL.0000000000003532
6. Babajani-Feremi A, Narayana S, Rezaie R, et al. Language mapping using high gamma electrocorticography, fMRI, and TMS versus electrocortical stimulation. *Clin Neurophysiol*. 2015; 127(3):1822–1836. [PubMed: 26679420]
7. Rutten GJM, Ramsey NF, Van Rijen PC, et al. Development of a functional magnetic resonance imaging protocol for intraoperative localization of critical temporoparietal language areas. *Ann Neurol*. 2002; 51(3):350–360. [PubMed: 11891830]
8. Meier MP, Ilmberger J, Fesl G, Ruge MI. Validation of functional motor and language MRI with direct cortical stimulation. *Acta Neurochir (Wien)*. 2013; 155(4):675–683. [PubMed: 23385293]
9. Gaillard WD, Balsamo L, Xu B, et al. fMRI language task panel improves determination of language dominance. *Neurology*. 2004; 63(8):1403–1408. [PubMed: 15505156]
10. Sepeta LN, Berl MM, Wilke M, et al. Age-dependent mesial temporal lobe lateralization in language fMRI. *Epilepsia*. 2016; 57(1):122–130. [PubMed: 26696589]
11. Cox RW. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Comput Biomed Res*. 1996; 29:162–173. [PubMed: 8812068]
12. Ward DB. Simultaneous Inference for fMRI Data. *J Chem Inf Model*. 2000; 53(9):1689–1699.
13. Seghier ML. Laterality index in functional MRI: methodological issues. *Magn Reson Imaging*. 2008; 26(5):594–601. [PubMed: 18158224]
14. Dykstra AR, Chan AM, Quinn BT, et al. Individualized localization and cortical surface-based registration of intracranial electrodes. *Neuroimage*. 2012; 59(4):3563–3570. [PubMed: 22155045]
15. Cocjin, J., Damera, S., Saad, Z., et al. Standardized regions of interest for population analyses using electrocorticography (ECoG). Chicago, IL: 2015.
16. Olman CA, Davachi L, Inati S. Distortion and signal loss in medial temporal lobe. *PLoS One*. 2009; 4(12)
17. Ojemann JG, Akbudak E, Snyder AZ, et al. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage*. 1997; 6(3):156–67. [PubMed: 9344820]
18. Nathan SS, Sinha SR, Gordon B, et al. Determination of current density distributions generated by electrical stimulation of the human cerebral cortex. *Electroencephalogr Clin Neurophysiol*. 1993; 86(3):183–192. [PubMed: 7680994]
19. FitzGerald DB, Cosgrove GR, Ronner S, et al. Location of language in the cortex: A comparison between functional MR imaging and electrocortical stimulation. *Am J Neuroradiol*. 1997; 18(8): 1529–1539. [PubMed: 9296196]
20. Genetti M, Tyrand R, Grouiller F, et al. Comparison of high gamma electrocorticography and fMRI with electrocortical stimulation for localization of somatosensory and language cortex. *Clin Neurophysiol*. 2015; 126(1):121–130. [PubMed: 24845600]
21. Haglund MM, Berger MS, Shamseldin M, et al. Cortical localization of temporal lobe language sites in patients with gliomas. *Neurosurgery*. 1994; 34(4):567–576. [PubMed: 7516498]
22. Sawrie SM, Chelune GJ, Naugle RI, Luders HO. Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. 2017; (1996):556–564.
23. Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: A single indicator of test performance. *J Clin Epidemiol*. 2003; 56(11):1129–1135. [PubMed: 14615004]
24. Bizzi A, Blasi V, Falini A, et al. Presurgical Functional MR Imaging of Language and Motor Functions: Validation with Intraoperative Electrocortical Mapping 1. *Radiology*. 2008; 248(2): 579–589. [PubMed: 18539893]

25. Ruge MI, Victor J, Hosain S, et al. Concordance between Functional Magnetic Resonance Imaging and Intraoperative Language Mapping. *Imaging Meet Am Soc Stereotact Funct Neurosurg Stereotact Funct Neurosurg Stereotact Funct Neurosurg*. 1999; 7272
26. Hirsch J, Ruge MI, Kim KHS, et al. An integrated functional magnetic resonance imaging procedure for preoperative mapping of cortical aread associated with tactile, motor, language, and visual functions. *Neurosurgery*. 2000; 47(3):711–722. [PubMed: 10981759]
27. Roux FE, Boulanouar K, Lotterie JA, et al. Language functional magnetic resonance imaging in preoperative assessment of language areas: Correlation with direct cortical stimulation. *Neurosurgery*. 2003; 52(6):1335–1347. [PubMed: 12762879]
28. Spena G, Nava A, Cassini F, et al. Preoperative and intraoperative brain mapping for the resection of eloquent-area tumors. A prospective analysis of methodology, correlation, and usefulness based on clinical outcomes. *Acta Neurochir (Wien)*. 2010; 152(11):1835–1846. [PubMed: 20730457]
29. Suarez RO, Taimouri V, Boyer K, et al. Passive fMRI mapping of language function for pediatric epilepsy surgical planning: Validation using Wada, ECS, and FMAER. *Epilepsy Res*. 2014; 108(10):1874–1888. [PubMed: 25445239]
30. Ille S, Sollmann N, Hauck T, et al. Impairment of preoperative language mapping by lesion location: a functional magnetic resonance imaging, navigated transcranial magnetic stimulation, and direct cortical stimulation study. *J Neurosurg*. 2015; 123(August):314–324. [PubMed: 25884257]
31. Rodin D, Bar-Yosef O, Smith M Lou, et al. Language dominance in children with epilepsy: Concordance of fMRI with intracarotid amyltal testing and cortical stimulation. *Epilepsy Behav*. 2013; 29(1):7–12. [PubMed: 23911353]
32. 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. [PubMed: 18513352]
33. Gaillard W. Re: Atypical language in lesional and nonlesional complex partial epilepsy. *Neurology*. 2008; 70(23):1761–1771.
34. Berl MM, Zimmaro LA, Khan OI, et al. Characterization of atypical language activation patterns in focal epilepsy. *Ann Neurol*. 2014; 75(1):33–42. [PubMed: 24038442]
35. You X, Adjouadi M, Guillen MR, et al. Sub-patterns of language network reorganization in pediatric localization related epilepsy: A multisite study. *Hum Brain Mapp*. 2011; 32(5):784–799. [PubMed: 21484949]
36. Hamberger MJ, Goodman RR, Perrine K, Tamny T. Anatomic dissociation of auditory and visual naming in the lateral temporal cortex. *Neurology*. 2001; 56(1):56–61. [PubMed: 11148236]
37. Binder JR. Functional MRI is a valid noninvasive alternative to Wada testing. *Epilepsy Behav*. 2011; 20(2):214–222. [PubMed: 20850386]
38. Binder JR, Frost Ja, Hammeke Ta, et al. Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci*. 1997; 17(1):353–362. [PubMed: 8987760]
39. Gaillard WD, Balsamo L, Xu B, et al. Language dominance in partial epilepsy patients identified with an fMRI reading task. *Neurology*. 2002; 59(2):256–65. [PubMed: 12136067]
40. Ahmad Z, Balsamo LM, Sachs BC, et al. Auditory comprehension of language in young children: neural networks identified with fMRI. *Neurology*. 2003; 60(10):1598–605. [PubMed: 12771248]
41. Lehéricy S, Cohen L, Bazin B, et al. Functional MR evaluation of temporal and frontal language dominance compared with the Wada test. *Neurology*. 2000; 54(8):1625–1633. [PubMed: 10762504]
42. Schlosser MJ, Aoyagi N, Fulbright RK, et al. Functional MRI studies of auditory comprehension. *Hum Brain Mapp*. 1998; 6(1):1–13. [PubMed: 9673659]
43. Schlosser MJ, Luby M, Spencer DD, et al. Comparative localization of auditory comprehension by using functional magnetic resonance imaging and cortical stimulation. *J Neurosurg*. 1999; 91(4): 626–35. [PubMed: 10507385]
44. Gaillard WD, Pugliese M, Grandin CB, et al. Cortical localization of reading in normal children: an fMRI language study. *Neurology*. 2001; 57(1):47–54. [PubMed: 11445627]

45. Mandonnet E, Winkler PA, Duffau H. Direct electrical stimulation as an input gate into brain functional networks: Principles, advantages and limitations. *Acta Neurochir (Wien)*. 2010; 152(2): 185–193. [PubMed: 19639247]
46. Diehl B, Piao Z, Tkach J, et al. Cortical stimulation for language mapping in focal epilepsy: Correlations with tractography of the arcuate fasciculus. *Epilepsia*. 2010; 51(4):639–646. [PubMed: 20002151]

Figure 1. Electrodes projected to the cortical surface

Patients 16 and 25; green = cleared electrodes, red = language positive electrodes, grey = electrodes were not tested

Figure 2. Functional MRI masks and electrode ROIs

Patient 19, a right handed 28 year old female, with left lateralizing fMRI activation; Left = Right; 2A: Task A (ADDT), at thresholds P1, P2, and P3, from left to right; 2B: Task B (AUDCAT), at thresholds P1, P2, and P3, from left to right (color overlay represents value of T statistic); 2C: ADDT $\times$ AUDCAT combined activation map at P1; 2D: a DCS language positive electrode (red), and a DCS cleared electrode pair (green).

Figure 3. Sensitivities and specificities of each task for DCS+ Patients
 A: ADDT, B: AUDCAT, C: Listening, D: Reading

Figure 4. Sensitivities and specificities of combined activation maps for all patients

A: ADDT × AUDCAT × Listening × Reading, B: ADDT × AUDCAT, C: ADDT × Reading

Table 1**Patient Demographics**

Patient	Gender	Age at Implant	FSIQ (Preop)	Handed-ness	MRI Findings	Implant side	+ DCS	Resection type	Postop decline
1	F	44y	85	R	2	L	+	L ATL	*
2	M	25y	88	L	1	R		R posterior temporal neocortical resection	
3	M	31y	128	R	1	R		R parietal topectomy	
4	F	54y	91	R	3	R	+	R ATL	
5	F	25y	77	R	3	R		No resection	
6	F	33y	113	R	2	L	+	L Temporal Topectomy	
7	F	30y	127	Ambi	1	L	+	L SAH	
8	F	23y	102	R	1	R		R ATL	
9	F	59y	110	L	2	L	+	L ATL	
10	F	29y	98	R	4	BI		R ATL	
11	F	49y	130	R	2	L	+	L SAH	*
12	F	34y	80	R	2	L	+	L ATL	*
13	M	44y	125	R	1	R		R FL	
14	F	35y	93	R	1	BI	+	R ATL	
15	F	22y	100	L	1	L		No resection	
16	M	51y	74	R	5	L	+	L ATL	
17	M	33y	91	R	3	L	+	L ATL	
18	F	56y	92	R	5	L	+	No resection	
19	F	28y	86	R	1	L	+	L ATL	

MRI Findings: 1 = normal, 2 = mesial temporal sclerosis, 3 = lesional (tumors, focal cortical dysplasia), 4 = dual pathology, 5 = other (encephalomalacia, traumatic brain injury, nonspecific MRI); Resection types: ATL = anterior temporal lobectomy, FL = frontal lobectomy, SAH = selective amygdalohippocampectomy. Handedness was based on self-report during clinical evaluation. The 3 subjects who reported left-handedness scored in the ambidextrous range on the Edinburgh Handedness Inventory (EDHI). Full Scale IQ (FSIQ) scores were obtained during preoperative neuropsychological testing; all patients completed the WAIS-III except for patients 15, 17, and 18, who completed the WAIS-IV. If postoperative neuropsychological testing was completed, and a significant decline occurred (on 1 or more described language measures), the change is noted in the last column (*).

Table 2

Language Characteristics

Patient	Handedness	fMRI Laterality Index	Classification		Implant side	Location of + DCS site
			ADDT	AUDCAT	Listening	Reading
1	R	L	L	L	L	L
2	L	L	L	L	L	R
3	R	L	L	L	L	R
4	R	R	BI	R	BI	R
5	R	L	BI	BI	BI	R
6	R	L	L	L	BI	L
7	Ambi		L	R	BI	L
8		L	BI	L	BI	R
9	L	L	L	L	BI	L
10	R	L	L	L	L	BI
11	R	L	BI	BI	BI	L
12	R	BI	BI	BI	BI	L
13	R	L	L	L	L	R
14	R	L	BI	BI	BI	L
15	L	R	R	R	R	L
16	R				BI	L
17	R	BI	BI	BI	BI	L
18	R	L	L	L	BI	L
19	R	L	BI	BI	BI	L

A laterality index was calculated at 3 thresholds for each fMRI language task. If a subject's laterality classification differed across thresholds for a task, the more common classification was chosen (e.g. if a subject was bilateral at P1, but left-lateralized at P2 and P3, they were classified as left). Subjects without a laterality classification were excluded based on movement exclusions, activation exclusion, or simply not completing a given task.