UCLA
UCLA Previously Published Works

Title

Presurgical language fMRI: Mapping of six critical regions

Permalink

Journal

Human Brain Mapping, 38(8)

ISSN

1065-9471

Authors

Benjamin, Christopher F
Walshaw, Patricia D
Hale, Kayleigh
et al.

Publication Date

2017-08-01

DOI

10.1002/hbm.23661

Peer reviewed

Presurgical Language fMRI: Mapping of Six Critical Regions

Christopher F. Benjamin 🆔,[1,2]* Patricia D. Walshaw,[3] Kayleigh Hale,[4]
William D. Gaillard,[5] Leslie C. Baxter,[6] Madison M. Berl,[5]
Monika Polczynska,[3,7] Stephanie Noble,[8] Rafeed Alkawadri,[1]
Lawrence J. Hirsch,[1] R. Todd Constable,[8] and Susan Y. Bookheimer[3]

[1]Department of Neurology, Comprehensive Epilepsy Center, Yale School of Medicine,
New Haven, Connecticut
[2]Department of Neurosurgery, Yale School of Medicine, New Haven, Connecticut
[3]UCLA Department of Psychiatry and Biobehavioral Sciences, Los Angeles, California
[4]U.S. Department of Veterans Affairs, War Related Illness and Injury Study Center, Washington, DC
[5]Center for Neuroscience Research, Children's National Health System, Washington, DC
[6]Department of Neuroimaging Research, Barrow Neurological Institute, Phoenix, Arizona
[7]Faculty of English, Adam Mickiewicz University, Poznań, Poland
[8]Department of Radiology and Biomedical Imaging, Yale School of Medicine,
New Haven, Connecticut

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

Abstract: Language mapping is a key goal in neurosurgical planning. fMRI mapping typically proceeds with a focus on Broca's and Wernicke's areas, although multiple other language-critical areas are now well-known. We evaluated whether clinicians could use a novel approach, including clinician-driven individualized thresholding, to reliably identify six language regions, including Broca's Area, Wernicke's Area (inferior, superior), Exner's Area, Supplementary Speech Area, Angular Gyrus, and Basal Temporal Language Area. We studied 22 epilepsy and tumor patients who received Wada and fMRI (age 36.4[12.5]; Wada language left/right/mixed in 18/3/1). fMRI tasks (two × three tasks) were analyzed by two clinical neuropsychologists who flexibly thresholded and combined these to identify the six regions. The resulting maps were compared to fixed threshold maps. Clinicians generated maps that overlapped significantly, and were highly consistent, when at least one task came from the same set. Cases diverged when clinicians prioritized different language regions or addressed noise differently. Language laterality closely mirrored Wada data (85% accuracy). Activation consistent with all six language regions was consistently identified. In blind review, three external, independent clinicians rated the individualized fMRI language maps as superior to fixed threshold maps; identified the majority of regions significantly more frequently; and judged language laterality to mirror Wada lateralization more often. These data provide initial

validation of a novel, clinician-based approach to localizing language cortex. They also demonstrate clinical fMRI is superior when analyzed by an experienced clinician and that when fMRI data is of low quality judgments of laterality are unreliable and should be withheld. *Hum Brain Mapp 38:4239–4255, 2017.* © 2017 The Authors Human Brain Mapping Published by Wiley Periodicals, Inc.

INTRODUCTION

Epilepsy is a debilitating neurological condition for which neurosurgery can be curative. Language impairment is a key risk of surgery, however, with naming decline occurring in a third of left temporal patients [Sherman et al., 2011]. Accurate mapping of language cortex is therefore a key goal of surgical planning.

fMRI in Presurgical Language Mapping

Language mapping via functional Magnetic Resonance Imaging ("language fMRI") can be seen as having three distinct but related goals: language lateralization; language localization; and prediction of postsurgical language change.

Language fMRI is valid for language lateralization, a proxy for prediction of post-surgical outcome. Wada testing has been the gold standard for language lateralization since the 1960s though study of its relationship to surgical outcome has been limited [Loring and Meador, 2015]. Its attendant cost, discomfort, and risks–including stroke [Loddenkemper et al., 2008] are key factors mitigating its use. From the late 1970s and 1980s Positron Emission Tomography (PET), a foundational non-invasive technique for language mapping, was then applied to map the language system [e.g., Hunter et al., 1999; Petersen et al., 1988] and for surgical planning [Theodore, 1989]. PET was shown to lateralize language effectively and have good correspondence with the results of Wada testing [Pardo and Fox, 1993] and direct cortical stimulation mapping [e.g., Bookheimer et al., 1997]. The methods developed here were then employed to the less invasive functional MRI, which has now been shown to determine language lateralization with a similar degree of success as Wada testing without that procedure's risks [see Binder, 2011; Szaflarski et al., 2017]. Wada testing and language fMRI have been shown to predict outcomes with a similar degree of success, with an extensive review and systematic analysis of one site's experience concluding ∼86% concordance [Janecek et al., 2013]. With the protocols used at that site, 92–94% concordance between Wada and fMRI was observed when one method indicated left language dominance. Clear discordance occurred in only 4/229 cases (2%) when, in each instance, fMRI indicated right lateralization and Wada, left.

The ability to predict language outcome using fMRI, particularly relative to Wada testing, has received less attention. In a key paper [Sabsevitz et al., 2003] language fMRI was shown to be at least as sensitive and specific as Wada testing in predicting postsurgical decline in naming. The authors examined 56 temporal lobe cases who underwent presurgical language mapping and used a laterality index based on temporal lobe activation to predict postoperative changes in naming (Boston Naming Task raw score). There was an 81% probability that a decline would occur when predicted, following tailored anterior temporal lobectomy (sensitivity/specificity/Positive Predictive Power 100%/73%/81%). The results for Wada were less impressive (sensitivity/specificity/PPP= 92%/43%/67%). A larger, valuable multi-site NIH study is ongoing (FMRI in Anterior Temporal Epilepsy Surgery).

There is not an established protocol for using language fMRI to localize all relevant regions of language cortex and guide surgical margins. This is a difficult goal for many reasons. fMRI protocols used across sites differ markedly in nearly all aspects (e.g., experimental design; image processing pipeline). Such details can seem minor, but markedly influence the fMRI results [Binder, 2011] and therefore the ability to localize language areas. Further, professional guidelines recommend activation maps be "reviewed over multiple statistical thresholds with attention to both the voxel-wise statistical significance and the anatomic extent of activations" [American College of Radiology, 2014]. This approach is valuable, but means that the resulting map will reflect the operator's model of the language system–the areas that are expected to be active. Other areas are likely to be excluded during thresholding, or not interpreted in the clinical report and therefore not considered by the surgical team. The model of the language system used clinically focuses on Broca's and Wernicke's areas [Ropper and Brown, 2005] and in a recently completed survey of epilepsy centers we found evidence that this model forms the focus of presurgical language fMRI. Clinical and cognitive neurosciences research have now shown, however, that this model is incomplete.

A Current Model of the Language System for Clinical Language fMRI

This historic model defining Broca's and Wernicke's areas as "expressive" and "receptive" is a useful heuristic, but does not fully reflect our current theoretical knowledge or the clinical picture of aphasias [Tremblay and Dick, 2016].
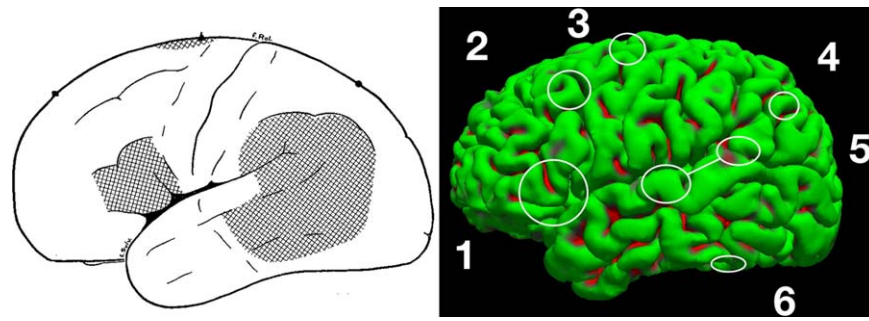
Figure 1.

Historic (left) and current (right) models of the language system. Left: "Speech areas: Evidence from stimulation," modified from Penfield & Roberts, 1959. Right: A model reflecting more recent knowledge (circles are approximate). (1) Broca's Area, in the posterior third of the inferior frontal gyrus. (2) Exner's Area, in the posterior middle frontal gyrus. (3) Supplementary motor area. (4) Angular gyrus. (5) Wernicke's area, inferior (mid to anterior STG) and superior (posterior STG and supramarginal gyrus) components. (6) Basal temporal language areas. Note that anterior temporal cortex also appears critically involved in auditory naming (not highlighted). [Color figure can be viewed at wileyonlinelibrary.com]

Based on our clinical experience using lexico-semantic tasks, at least six core clinically relevant language areas can be identified using fMRI (Fig. 1).

Broca's area

Our understanding of the role of Broca's area, perhaps most frequently defined as constituting pars opercularis and triangularis [Tremblay and Dick, 2016], in language is evolving. Our understanding of the clinical syndromes accompanying Broca's lesions evolved with research in the 1990s which highlighted the key role the insula plays in motor programming for speech articulation [Dronkers, 1996]. Specifically, clinical syndromes previously grouped as "Broca's Aphasia" can be understood as reflecting a spectrum spanning damage to the lenticular zone (anarthria), insula (dyspraxia), and inferior frontal cortex (Broca's-type agrammatic aphasia) [Donnan et al., 1999]. The specialization of subregions within Broca's continues to be elaborated. Discrete regions are likely involved in processing grammar, semantics, and phonology [Bookheimer, 2002; Rodd et al., 2015] in both expressive and receptive speech.

Exner's area ("graphemic motor frontal area")

This region, located at the posterior extent of the middle frontal gyrus, is critically involved in transforming phonological representations of words into the motor commands for producing their written forms [Roux et al., 2010]. Stimulation in this area has been found, for instance, to selectively disrupt handwriting (6/12 patients) [Roux et al., 2009]. Stimulation also showed Exner's area was separate from Broca's area, hand motor cortex, and frontal eye fields. In five further patients, writing disturbance was combined with reading and/or naming deficits on stimulation and two tumor patients who had the region partially resected suffered postoperative handwriting deficits. Of the numerous publications on this area, Anderson et al. [1990] case study of a secretary who underwent focal resection of a posterior MFG tumor and suffered postoperative alexia and agraphia, to which she adapted by, for example, using a rubber stamp of her signature and drawing grocery items for her shopping list, is of particular interest. This region falls within the dorsolateral prefrontal cortex, an area that is key in other cognitive processes such as working memory and executive skill, which are also recruited during use of language skills.

Supplementary speech area ("supplementary motor area," speech component; pre-SMA)

This area, located in posterior superior and medial frontal cortex, is critically engaged in initiating and sequencing motor movements for speech. It is unique among the listed language regions since its unilateral resection typically results in a transient (and in some instances, profound) aphasia that recovers over weeks to months postoperatively. For instance, Krainik et al. [2003] observed that six of 12 patients who had regions of the SMA resected suffered postoperative speech deficits varying from mutism (3 cases) to globally decreased output (3 cases), which fully or largely recovered by 8 months postoperatively (range 3–8 months). The degree of (transient) deficit was proportional to the extent of fMRI activation resected, and the authors concluded "unilateral SMA removal can be accomplished without resulting in significant permanent deficits" [Zentner et al., 1996]. Recovery appears to reflect recruitment of contralateral SMA, so that if communication between language areas and contralateral SMA is impaired (e.g., bilateral lesions, damage to corpus callosum) deficits may persist to various degrees [Endo et al., 2014].

Angular gyrus

This area is particularly involved in reading and transitioning between written and spoken forms of language. A meta-analysis by Binder et al. [2009] showed the left angular gyrus is consistently involved in semantic processing. Stimulation here can interrupt handwriting, causing unintelligible writing or repetition of words. In a series of 6 patients with Gerstmann syndrome and angular-inferior parietal lesions, Roux et al. [2003] noted agraphia in all six, with reading and object naming (ON) disturbances in a further five cases.

Wernicke's area

In contrast to Broca's area, defined anatomically, Wernicke's was defined functionally as an area where damage or stimulation causes comprehension deficits. Early work attributed Wernicke's large swathes of posterior cortex [Figure 1, Penfield and Roberts, 1959] and attempts to clarify its location and nature have been long-standing [Bogen and Bogen, 1976]. Multiple varied definitions remain in use [Binder, 2015; Tremblay and Dick, 2016]. In more recent decades, it has been appreciated that Wernicke's is not simply a receptive language region [e.g., Kagan and Saling, 1988]. An excellent recent overview [Binder, 2015] outlines evidence for (i) an anterior section of the superior temporal gyrus (STG), which is bilaterally engaged in phonological perception [see also Price, 2012], (ii) a posterior section extending to the supramarginal gyrus (SMG) involved in phonological retrieval (and processing), and (iii) evidence that damage to the posterior segment alone impairs phonological retrieval but not comprehension. In our experience, the inferior portion centers on BA 22, the superior, BA 22/39/40. fMRI tasks involving the processing and comprehension of auditory and written speech engage this region [Gaillard et al., 2004]. Damage causes impairment of phonological processing (posterior STG, SMG) and comprehension (temporal lobe more broadly) while disruption of phonological perception (pure word deafness) typically requires bilateral lesions to the anterior STG region [Binder, 2015].

The "basal temporal language area"

An area critical to language in the basal temporal region has long been recognized [e.g., Mills and McConnell, 1895]. In the epilepsy literature Lüders reported a case in 1986, one of a broader series [Lüders et al., 1985; Lüders et al., 1986], who suffered a dense anomia on stimulation of a discrete (2 × 2 mm) fusiform region. A number of series were subsequently reported in which stimulation in the fusiform, inferotemporal, and parahippocampal gyri caused disruption of a range of language tasks, for example, visual and auditory naming, auditory comprehension, reading, repetition and spontaneous speech [Burnstine et al., 1990; Krauss et al., 1996; Schaffler et al., 1994]. While the traditional view of this region is one associating semantics with names, this data suggests a less clear—although still critical—role. Separately, the cognitive neuroscience literature has focused on a more posterior fusiform region selectively engaged in word processing, which tunes across development to process text [Baker et al., 2007] (although the precise function of the "Basal Temporal Language Area" and its relationship to the "Visual Word Form Area"—that is, whether the two are synonymous—is an outstanding issue). This area has been mapped in surgical patients with fMRI and stimulation mapping, and when deafferented after surgery alexia without agraphia may ensue [Gaillard et al., 2006]. The extent to which these deficits are enduring are a topic of ongoing debate.

This Study

In our experience, these regions are consistently activated in presurgical language fMRI, and it is possible for clinicians such as neuropsychologists with the above knowledge of the language system and a detailed knowledge of fMRI to reliably identify them. The goal of this study was to provide an initial evaluation of the reliability of a standardized, clinician-generated (CG) approach to localizing language cortex using fMRI. We hypothesized that in a series of unselected patients, two independent clinicians would generate equivalent maps of the language system. Secondly, we hypothesized overall language laterality would match laterality determined by Wada testing, as the ability to lateralize the language system is a basic test that can be validated against a readily-available gold standard. Third, we expected that if this approach were reliable, the previously described regions would be consistently identified, as determined by independent reviewers. Fourth, we hypothesized CG maps with individualized thresholding would be superior (subjective quality; lateralization of language) to maps generated with a fixed threshold.

MATERIALS AND METHODS

Sample

The sample included all English-speaking patients who had undergone both Wada testing and clinical language fMRI at University of California, Los Angeles' (UCLA's) epilepsy and/or tumor programs from March 2012 through October 2013. Data were acquired retrospectively through review of existing records. The study was IRB-approved. Sample details are provided in Table I. 22 patients were included (a 23rd, blind, patient was excluded) with a mean age 36.4 years (SD 12.5, 16.7–63.6), including eight females and 16 right-handed individuals. Handedness was determined by clinical evaluation and epilepsy team opinion. Wada testing found left language

TABLE I. Sample demographics

Case	fMRI Age	Sex	Hand	Wada	Pathology laterality	Lobe	Febrile seizures	Birth diff.	Other notes	Onset (years)	Lesion/focus
A	22	M	L	L	L	T	–	–	–	22 [7]	Left anterior hippocampal nodule
B	36	F	R	L	L	T	18m[10]	–	–	32	Left amygdala, anterior temporal cortical dysplasia (MRI, PET)
C	54	M	R	L	L	TO	–	–	–	36	Left PCA infarct, historic, primarily temporal, more limited occipital
D	50	M	R	L	L	T	–	–	–	50	Left anterior and mesial temporal tumor
E	32	M	R	R	R	T	–	–	6/7y [2]	22	Right hippocampal lesion, suspected neoplasm
F	37	M	L	L	L	T	–	–	§	16	Past left anterior temporal resection, historic astrocytoma
G	31	F	R	L	L	T	–	–	–	26	Left middle temporal gyrus cavernous malformation & mesial temporal sclerosis
H	33	M	L	R	L/Bi	T	–	Y [1]	–	1	Left MTS, left caudate encephalomalacia, right thalamus and pons lacunar infarcts
I	48	F	R	L	R	T	–	–	5y [9]	30	MRI nonlesional, EEG R>L, proceeded to right anterior temporal lobectomy (ATL)
J	16	M	R	L	R	T	–	–	10y [3]	10	Right posterior temporal sclerosis (post-surgical report)
K	56	M	R	L	L	FT	–	–	29/30 [4]	48	Left fronto-temporal encephalomalacia
L	41	M	R	L	L	T	–	–	[5] §	27	Past left hippocampal-sparing ATL, probable recurrent tumor.
M	36	M	L	M	L	F	–	–	–	14	Left orbitofrontal on EEG, PET, Chaslin's gliosis & arteriosclerosis post phase III
N	43	M	R	L	L	T	–	–	–	41 [8]	Left MTL low grade neoplasm (inc. hippocampus)
O	23	F	R	L	L	T	–	–	–	20	Left amygdala tumor (biopsy: pleomorphic xanthoastrocytoma)
P	21	F	R	L	L	T	–	–	–	15	Left temporal pole cavernous malformation
Q	33	F	R	L	L	T	–	–	–	12	Left temporal pole abnormality and MTL sclerosis
R	63	F	R	L	L	T	36m	–	–	3	Left hippocampal atrophy and gliosis.
S	25	M	L	R	R	T	3m	–	3m [6]	1	Left anterior temporal cortical dysplasia and MTS
T	38	M	R	L	L	T	–	–	§	15	Past left ATL (astrocytoma), residual HS.
U	23	M	L	L	L	T	–	–	§	7	Past partial left ATL, residual amygdala distortion
V	29	F	R	L	L	TP	–	–	–	9	Right temporo-parietal encephalomalacia (lateral, mesial), extensive

T, temporal; FT, fronto-temporal; F, frontal; TP, temporo-parietal. § Prior resection as at time of fMRI. [1] Hypoxia, heart failure in utero (cord around neck). [2] Traumatic brain injury. [3] Mild TBI a few months prior to onset. [4] Severe TBI with 2 weeks' coma age 29 or 30. [5] Neurofibromatrosis type 1. [6] Meningits at 3 months, developmental delay (spoke at 4). [7] Tumor symptom onset. [8] Seizures are psychotic phenemoena; "onset" is first documented admission. [9] Single seizure at 5 years of age when hypoglycaemic (diabetes diagnosed age 2). [10] Definite fever with possible seizure. Data per MRI and other available records. Patients E, H, and L had multiple Wada tests (2, 3, and 2, respectively) with the same language finding, and the final result and conclusion is used. Age is relative to fMRI date.

dominance in 18 cases, right in three, and mixed in one. To evaluate this protocol in a standard clinical setting, patients were accepted regardless of pathology. Pathology was primarily left hemisphere (18; right in four) and affected lobes primarily included temporal (18); frontal (1); temporo-occipital (1); fronto-temporal (1); and temporo-parietal (1), and three patients had prior resections.

Protocol

Clinical fMRI (below) and Wada testing with sodium amytal [Połczyńska et al., 2014] were completed on different days. A clinical neuropsychologist (PDW or SYB) supervised acquisition.
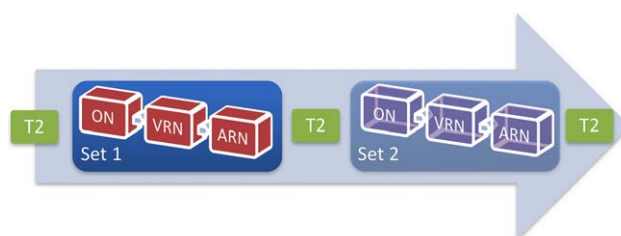
Wada Testing

A modified Montreal protocol was used [e.g., Połczyńska et al., 2014]. A catheter was placed within the carotid artery (transfemorally; radiologist). Positioning and flow were confirmed via X-Ray. EEG was monitored throughout the procedure (neurologist). Injection of the ipsilateral hemisphere occurred first. The contralateral hemisphere was later injected (>=30 min break between injections) if clinically requested. The patient counted aloud while gripping the neurologist's hand. Typically, 125 mg of amytal in 10 cc of saline were injected in a bolus over 4 s, modified if needed based on patient size or anesthetization level after initial injection. Anesthetization was considered acceptable when both (i) EEG slowing was observed with preserved contralateral awake background; and (ii) the patient's contralateral hand grip dropped from baseline to flaccid (0/5) as judged by the neurologist.

Encoding and recovery phases

Six physical objects were presented and repeatedly named aloud (neuropsychologist SB, PW or CB) while strength remained at 0 and EEG was slowed. Two commands (e.g., "wiggle your toes") were given. Recovery: language (comprehension; repetition; auditory naming) was tested using standard cognitive measures (e.g., Boston Diagnostic Aphasic Exam items) and conversational speech through to recovery. Recovery occurred when all of: (i) strength had fully returned; (ii) EEG was at baseline; (iii) language was at baseline; and (iv) at least 12 min had passed post-injection.

Recall phase

Memory was then tested using progressive cueing. (1) Spontaneous: The patient was asked to recall presented objects and commands. (2) Cued: For items not yet recalled, the patient was given a semantic cue ("I showed you something you might find in a toolbox..."). Correctly recalled items were noted. (3) Recognition: For items not yet recalled, the object (or command) was presented with three additional semantically related lures (e.g., a hammer might be shown with a paint brush; screwdriver and measuring tape). Correctly recognized items were noted.

Figure 2.
Overview of clinical imaging protocol. An MPRAGE and other images were also acquired. Two sets of task-related T2* images were typically acquired with each set including ON; VRN; and ARN. ON, object naming; VRN, verbal responsive naming; ARN, auditory responsive naming. [Color figure can be viewed at wileyonlinelibrary.com]

Wada data

The language score was percent of correctly named objects (6) and followed auditory commands (2) (maximum total: 8). The memory score was the percentage of objects identified in recall phases 1, 2, or 3 (scored out of 6 if language impaired; 8 [including recall of commands] if language was not impaired).

MRI Task Sequences

One to two sets of three functional tasks were completed (Fig. 2) on a 3T Siemens Allegra MRI scanner, as well as multiple coplanar matched bandwidth T2 images (which bookended the EPI sequences) for use in image interpretation. We aimed to acquire two sets of these tasks if time allowed, especially if movement, poor compliance, or image artifact occurred. The precise numbering and ordering of functional tasks for all patients is detailed in Supporting Information 3. These were followed by other sequences (e.g., MPRAGE, potentially other clinical or research sequences).

We used three lexico-semantic tasks. (1) ON [Bookheimer et al., 1995] with object use. A line drawn object was presented and the patient was asked to "silently name each object and an action you could perform with it." (2) Verbal responsive naming (VRN; i.e., word reading) [Gaillard et al., 2001, 2002, 2004]. A three-word written description of a concrete noun or adjective (e.g. "tall pink bird?") was presented. The patient's task was to read the description, and think of the object name. (3) Auditory responsive naming (ARN) [Bookheimer et al., 1997; Gaillard et al., 2004]. A three-word description of a concrete noun or adjective (e.g., "Color of snow?") was presented through headphones. The patient's task was to mentally generate the described object name silently. All tasks

included 12 paired blocks of task and rest (10 s per block) and began with 15 s rest (total 4′15″). All blocks began with a 1s cue (e.g., "rest now"). Rest involved the patient lying still and viewing a blank screen. Patients of low cognitive ability, determined by neuropsychological evaluation and/or interview with a clinical neuropsychologist, completed the same task though at a slower pace; 20 blocks of task (12.5s) and rest (10s) for ON and VRN (totaling 4′); and 7 task (20s)/8 rest (10s) blocks for ARN (also totaling 4′15″) (see Supporting Information 3). The T2* protocol included 96 T2* images; TE/TR/Voxel Size/FOV = 35/2.5/3.1 × 3.1 × 3/200 mm, with Fat saturation and began with two discarded images to address B0 effects. The anatomical image used in analysis and interpretation, the High Resolution matched bandwidth T2, was acquired at TE/TR/Voxel Size/FOV = 33/5.0/1.6 × 1.6 × 4/200 mm, 28 slices. Position of patients was typically adequate with the exception of cases E, H, J, N, and V being slightly tilted (accentuated in Case V given patient pathology).

The tasks used here are freely available for download at www.cogneuro.net/hbm2017 or from the authors. Please note that ON stimuli (black and white, line drawn objects from the Boston Naming Task 2 and other tests) have been replaced with non-copyrighted, black and white line-drawn alternatives and audio stimuli have been re-recorded.

Analysis

Variables

Wada testing yielded a judgment of language laterality based on naming (0–6 items), comprehension (0–2 items) scores, and qualitative language evaluation. The average interval between Wada testing and fMRI was 3.7 months (mean 112 days, SD 99, range 2–322) with the exception of an individual who had fMRI during surgical workup at age 23, and Wada at prior workup aged age 15 (Case U; 2913 day interval). This patient was left language dominant on Wada and fMRI, was left handed and had left hemisphere pathology.

fMRI analysis yielded 3 language maps per patient: two CG and one fixed threshold map (further to the below, see Supporting Information 1).

fMRI preprocessing was minimal to keep the data as close as possible to its original form. Data were spatially smoothed (2 mm kernel) and the correlation between actual and predicted BOLD (canonical HRF, block design) was computed (Pearson's r). A correlation coefficient allows for acquisition of fewer images than a General Linear Model (GLM)-based approach, and simple removal of movement-contaminated images (although this was not required for the patients in this dataset). Discussion of this approach, and comparison of one case also analyzed using GLM, is presented in Supporting Information 4. Each task's correlation map (r) was used for subsequent

analysis. This processing was completed in custom software and is detailed further in Supporting Information 1.

Two CG maps

Clinicians (SYB, PDW) independently generated one thresholded map per patient, representing that individual's language system. These maps were derived from the r maps for each of the 3–6 tasks and represent a conjunction of these tasks (areas common across maps). In order, preference was for a single map based on (1) all three tasks; (2) conjunction of one auditory and one visual task; (3) a single task. Thresholds were selected by the clinician subjectively as per American College of Radiology guidelines [2014] through a process of iteratively (i) evaluating sequence quality (e.g., evaluating r maps for signs of movement); (ii) selecting and thresholding candidate sequences, and viewing regions' time courses; and (iii) forming conjunction maps. Maps were overlaid on a matched bandwidth T2 anatomical reference. Clinicians were blind to patient and clinical data, and Wada findings. A detailed working description of the six areas as used in analysis is provided in Supporting Information 2.

One "fixed threshold" map

This was generated using the "common threshold-dependent" method [Suarez et al., 2009] without input from a clinician. Central to this approach is the use of an inflexible threshold, frequently set at $P < 0.001$ uncorrected or $P < 0.05$ corrected [Seghier, 2008]. This map was a conjunction of the three language maps from Set 1 (ON, ARN, VRN) thresholded at the voxel level with a joint probability of $P < 0.001$ (each individual map $P < 0.1/r> = 0.17$). The first acquired T2 was used as a reference image.

Quantitative analysis

The spatial overlap of maps was directly evaluated using average percentage overlap and a similarity coefficient [e.g., DICE coefficient; Dice, 1945]. To this end, each patient's three correlation maps were rigidly aligned (nearest neighbor). Transformations were derived using each map's collinear T2 image. fMRI language laterality was judged using active voxels in the left (L) and right (R) hemispheres to calculate an asymmetry index (L-R/L + R) [Binder et al., 1996]. Values varied from +1 (left) to −1 (right) with the lower 5% of rightward and leftward lateralizing values (i.e., −0.05 to +0.05) considered bilateral. CG maps for one case (S) were used for qualitative but not quantitative analysis; they could not be accurately aligned in spite of varied attempts to do so.

Qualitative analysis

The 22 cases yielded a total of 62 maps (22 fixed threshold; 20 clinician 1; 20 clinician 2), as the clinicians

independently considered the same two cases too poor for mapping. Three clinicians not involved in analysis reviewed all images (WDB, LCB, and MMB). All reviewers are faculty who complete language fMRI for their epilepsy programs and publish in the field. Specifically, WDG is an epileptologist and neurologist with over 20 years' experience in language fMRI; LCB is a practicing clinical neuropsychologist with 18 years' experience in language fMRI; MMB is a practicing clinical neuropsychologist with 14 years' experience in fMRI. Each reviewed 31 images (50% of rater one's images, 50% of rater two's, 50% of the fixed threshold). Inter-rater reliability of laterality judgment was good, with map laterality agreeing 89, 86, and 84% of the time for raters WDG, LCB, and MBB, respectively ($n = 18$, $n = 21$, $n = 19$ cases).

Reviewers were asked to label all visible regions, judge overall laterality, and rate image quality on a five-point scale (Supporting Information 2) from (1) small, questionable activations to (5) very good activation with the "location and extent of the majority of identified regions being consistent with your expectations". To ensure consistency, before reviewing any study data reviewers read definitions of the areas and four sample cases. This detailed definition and description of each region is provided in Supporting Information 2. Reviewers were blind to patient and clinical data, Wada findings, and whether maps were generated by a clinician or automated approach (images were randomly ordered).

Evaluation of hypotheses

(1) That independent clinicians will generate equivalent maps using this method. The degree to which maps overlapped was evaluated: (i) using the average percent overlap of each map with the other, and (ii) using the Dice coefficient, a more conservative measure comparing the (doubled) union of the two maps by the total area of each map ($2[A\cap B]/[A + B]$) [Dice, 1945]. Patient correlates of high vs. low overlap (handedness, language laterality, prior resection, movement, number of image sets acquired) were also evaluated. The runs selected for analysis by each clinician were also compared. (2) That language laterality using this method will closely match Wada language laterality. Language lateralization using laterality indices from CG maps for each patient was contrasted with that derived from Wada testing [as per e.g., Suarez et al. 2009], and compared with the average fMRI-Wada agreement for fMRI paradigms in the field [~85%; Janecek et al., 2013]. Judgment of language laterality was also evaluated qualitatively (by external reviewers not involved in analysis) to more accurately reflect how laterality is often determined in a clinical setting. (3) That the afore-mentioned language regions will be consistently identified. External reviewers also independently reviewed each image to identify Broca's Area; Wernicke's Area inferior; Wernicke's Area superior (including supramarginal gyrus); Basal Temporal Language Area; Angular Gyrus;

Supplementary Motor Area/Supplementary Speech Area; and Exner's Area, and overall accuracy was reviewed. (4) That CG maps will differ from fixed threshold maps generated without clinician input. Subjective map "quality" (see above) of CG and fixed threshold maps, determined by external reviewers blind to map type, was compared using t tests. The spatial extent of each map type was contrasted. Estimated language laterality and the frequency with which each identified the language regions were evaluated using Chi^2.

RESULTS

Equivalence of CG Maps

Overall, the average spatial overlap between clinicians' maps for any given case was moderate at 61% (SD = 0.31; Dice = 0.58, SD = 0.33; $n = 19$). In the majority of cases where clinicians selected at least one of the six sequences from the same set ($n = 12$ patients), average overlap was high, at 78% (SD = 0.26; Dice = 0.74, SD = 0.31). This degree of overlap (78%) was significantly above that when analysis only included sequences from different sets ($n = 7$) (average overlap 0.78 vs. 0.32, $t_{(17)} = -4.353$; $P < 0.000$; Dice = 0.74 vs. 0.31, $t_{(17)} = -3.475$; $P = 0.003$). Two cases, which both clinicians independently considered too poor for analysis, were set aside, here. In comparing cases with lower ($<50\%$, $n = 8$) versus higher ($>50\%$, $n = 11$) overlap, a second set of sequences were more likely to be acquired if overlap was low (Fisher's exact, $P = 0.045$), though this did not reflect greater movement within the two sets (mean displacement = 0.24 vs. 0.26 mm; $t(99) = -0.5135$, $P = 0.61$; see also Supporting Information 3). Patient factors, including handedness (Fisher's exact, $P = 0.338$); laterality of pathology on MRI ($P = 0.603$); Wada language laterality ($P = 1$); and the presence of prior resection ($P = 1$), did not differ in low as compared with high overlap cases. A sample map is shown in Figure 3.

Qualitative examination of divergent cases (average overlap <0.5) suggested differing results when clinicians made different decisions in dealing with image noise or prioritizing representation of language structures. When dealing with noise, core regions tended to be well-represented by both clinicians but activation extent and artifactual activation varied. For instance, case E (average overlap 0.34) contained significant noise (Fig. 4). Clinician 1 (red) initially formed a conjunction from two tasks, though then elected to reduce the noise further with a conjunction of all three tasks. In contrast, Clinician 2 (yellow) used two runs (VRN, ARN) to ensure a more liberal representation of the language regions, noting they may use global normalization in clinic to further reduce noise. The images were closely aligned and activations were similar (and approximately colocated), although extent was limited by the three-task conjunction. Similar representations
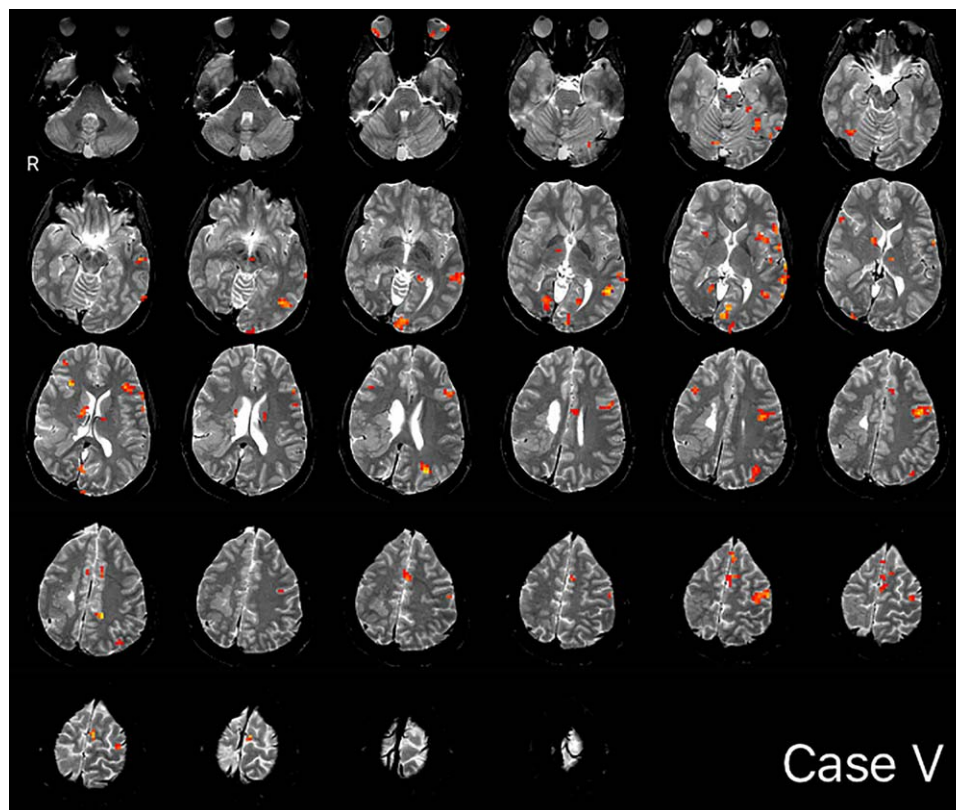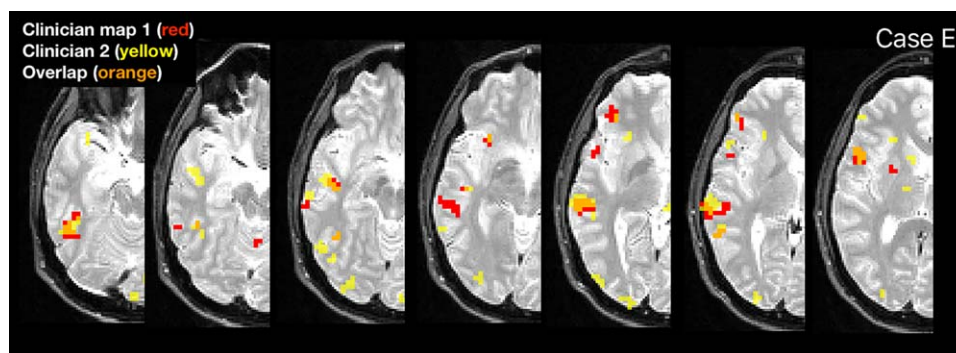
Figure 3.

Example map, Case V. Conjunction of ON, visual responsive naming and ARN language maps. ON, object naming; ARN, auditory responsive naming. [Color figure can be viewed at wileyonlinelibrary.com]

of most areas were obtained (e.g., Broca's area, Basal Temporal Language Area), and laterality indices from both maps indicated right hemisphere language (–0.38; −0.27), consistent with Wada. The clinicians similarly struggled with noise in cases D, K, and T.

In other instances, maps differed due to clinicians prioritizing representation of different language regions. In case O (average overlap 0.16; Fig. 5), both clinicians analyzed VRN and ARN runs but from different sets, and the images were in close alignment. Clinician 1 focused on

Figure 4.

Example divergent case with significant noise (Case E) showing maps generated by clinician 1 (red) and two (yellow), and the overlap (orange). [Color figure can be viewed at wileyonlinelibrary.com]
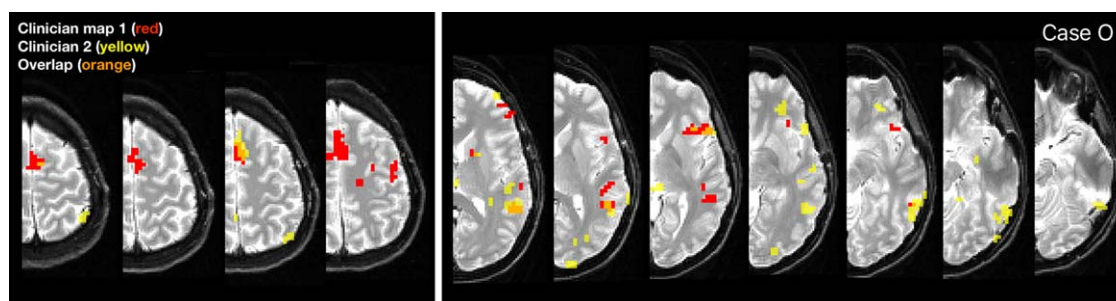
Figure 5.

Example divergent case where clinicians prioritized representations of different language regions (case O). Clinician 1 prioritized SMA (left, red) while clinician 2 better emphasized basal temporal language areas (right, yellow). Clinically this frequently occurs when the clinical question relates to different anatomical regions (e.g., frontal tumor vs. anterior temporal lobectomy). In this instance, the fixed threshold approach identified minimal temporal activation. [Color figure can be viewed at wileyonlinelibrary.com]

optimally representing SMA, while Clinician 2 prioritized activation consistent with the basal temporal language and inferior Wernicke's areas at the expense of superior language regions. Activation in the region of Broca's and Exner's areas was represented similarly in the maps. Here, Clinician 2 noted during analysis that they would typically create two maps to best represent anterior and posterior language areas. Other cases (e.g., L and Q) also reflected prioritization of different regions.

Language Laterality as Compared with Wada Testing

Language laterality indices determined from CG maps closely followed Wada-determined language laterality (Table II). With Wada laterality as reference, 29/32 (91%) of left-dominant cases were found left dominant by the clinician; 4/4 (100%) of right dominant cases were similarly categorized, as were 1/2 mixed cases. These data indicate an overall concordance of ~89%.

Clinical judgment of laterality by external reviewing clinicians (blind to analysis) typically matched Wada findings (85% of cases; 51/60). When cases of mixed language on fMRI (12%; $n = 7$) were not considered (in clinic these would elicit a Wada test), discordance occurred in 2/60 cases (3%).

For one of these cases, discordance appeared to reflect both poor quality fMRI data and possible mixed dominance on Wada testing (Case L; Wada left). For the fMRI data, both reviewers had judged the fixed threshold map unusable and the CG maps as low quality (1.5/5, 2/5). The discordant reviewer had considered the map right dominant (the other had concluded mixed). Further review of this patient's Wada (left injection only) revealed they had followed auditory commands, suggesting intact Wernicke's function. As such, the report had concluded both overall left language dominance and that the "right hemisphere likely supports at least some basic comprehension."

For the second discordant case (Case S; Wada right dominant) fMRI was again judged to be of poor quality and Wada findings were not clear-cut. For fMRI, four evaluations were completed: one review of the fixed threshold map; and three (total) reviews of the two CG maps (one

TABLE II. Language laterality by method

| Case | Wada | Language laterality (LI) by method | | |
		Clinician 1	Clinician 2	Fixed Threshold
A	L	L (0.64)	L (0.59)	L (0.55)
B	L	L (0.35)	L (0.35)	L (0.2)
C	L	L (0.46)	L (0.67)	L (0.44)
D	L	L (0.63)	L (0.49)	L (0.57)
E	R	R (−0.38)	R (−0.27)	R (−0.29)
F	L	L (0.55)	L (0.55)	L (0.19)
G	L	L (0.36)	L (0.36)	L (0.14)
H	R	R (−0.39)	R (−0.48)	R (−0.41)
I	L	L (0.37)	L (0.64)	L (0.43)
J	L	L (0.07)	M (0.03)	L (0.08)
K	L	L (0.32)	L (0.30)	M (0.04)
L	L	R (−0.07)	M (−0.03)	L (0.33)
M	M	M (0)	R (−0.06)	L (0.08)
N[a]	L	–	–	–
O	L	L (0.52)	L (0.57)	L (0.33)
P	L	L (0.30)	L (0.3)	L (0.38)
Q	L	L (0.12)	L (0.45)	L (0.09)
R[a]	L	–	–	–
S	R	–	–	–
T	L	L (0.28)	L (0.09)	L (0.2)
U	L	L (0.06)	L (0.12)	L (0.08)
V	L	L (0.65)	L (0.65)	L (0.62)

[a]Each clinician independently determined fMRI data quality was too poor (noise) to be used for language mapping.
Laterality index is in brackets. Shaded values represent discordance from Wada result.

review by each external reviewer). The fixed-threshold map was judged unusable. The reviewer whose judgment was Wada-discordant (left) also judged the CG map as low quality (2/5) and they could identify only one of the six language areas (Wernicke's–inferior and superior; left). The other two reviewers independently judged their maps (one generated by clinician 1; the other by clinician 2) as being of low quality (2/5) and of mixed dominance (in one instance, trending left). Review of the Wada report showed that while the patient was considered right dominant, they did name 1/6 items using the left hemisphere alone (right injection), again suggesting possible mixed dominance.

While these data suggest good consistency between fMRI and Wada, they suggest that when fMRI data are of low quality they are more likely to be read as differing from Wada findings. For seven maps (five patients), external reviewers considered maps to be of the poorest quality (uninterpretable). All were fixed-threshold (not CG) maps. For two of these patients clinicians had considered data unusable and not generated maps (Cases N, R). For two further patients, CG maps were read discordant with Wada (Cases L, S, above), while both maps for the fifth (Case V) were read consistent with Wada (left).

Localization of Language Regions

As judged by independent external reviewers, clinicians identified activation consistent with specific language regions with a high degree of consistency. Broca's Area was most frequently labeled (97% of cases), as were Wernicke's Area inferior (98%) and, to a lesser extent, Wernicke's Area superior (including supramarginal gyrus; 83%), Supplementary Speech Area (90%), Basal Temporal Language Area (82%), and Exner's Area (77%). Activation consistent with the Angular Gyrus was judged present least consistently, being labeled in 70% of cases.

Comparison of CG and Fixed Threshold Maps

The independent external reviewers' (blind) evaluation of CG and fixed-threshold map quality was compared (range: 1–5, see above). CG maps generated by different clinicians did not differ from one another in subjective quality (clinician 1: mean 3.5, SD = 0.9; clinician 2, mean = 3.4, SD = 1.1; $t_{(58)} = 0.379$, $P = 0.706$). For the same cases, maps generated using a fixed threshold were rated as being significantly worse (mean = 2.6, SD = 1.1; $t_{(87)} = 3.252$, $P = 0.002$). This held true even when only cases where the same clinician evaluated (blindly) both the CG and fixed threshold maps for each patient ($n = 29$ cases; $t_{(56)} = 2.607$, $P = 0.012$).

The voxel-by-voxel overlap between the CG and fixed threshold map for any given patient was ~52% (SD = 0.17; DICE = 0.37, SD = 0.21). Overlap between the CG and fixed threshold was equivalent to (did not differ from) the overlap

between the two clinicians' CG maps (average overlap $t_{(38)} = 1.419$, $P = 0.164$; Dice $t_{(38)} = 1.63$, $P = 0.110$). In the instances where the clinicians relied on at least one sequence from the same set ($n = 12$, per 3.1 above), their CG maps' overlap with the other CG map was significantly greater than the CG maps' overlap with the fixed threshold map (average overlap, $t_{(31)} = 3.746$, $P = 0.001$; Dice $t_{(31)} = 3.371$, $P = 0.002$). Specifically, CG maps were highly similar in spatial extent (78% overlap) while overlapping significantly less with fixed-threshold maps (52% overlap).

Laterality indices generated using fixed-threshold maps were similar in their correspondence with Wada data to those based on CG maps (Table II). Fixed threshold map laterality indices matched Wada lateralization in 15/16 (94%) left dominant cases, 2/2 right (100%) and 0/1 mixed (0%) cases.

Separately, external reviewers' judgments of the language laterality of fixed threshold maps was evaluated and compared with their review of CG maps. As noted, overall, reviewers read CG maps consistent with Wada in 85% of cases (51/60 maps) while fixed threshold maps matched Wada laterality in 69% (20/29). When reviewers (blindly) reviewed both the fixed threshold map and a CG map for the same patient, their determination of laterality differed a third of the time (31%; 9/29 cases). When their determination differed, the CG map most often matched Wada laterality (6/9 cases; all Wada left) while fixed threshold maps were either read as bilateral (5/9) or judged unusable (4/9).

Identification of activation consistent with specific language regions occurred more frequently in CG as compared with fixed threshold maps. Areas identified more frequently included Broca's Area (97 vs. 79%, $\chi^2 = 7.199$, $P < 0.05$); Wernicke's Area inferior (98 vs. 69%, $\chi^2 = 16.907$, $P < 0.05$); Wernicke's Area–superior (including supramarginal gyrus; 83 vs. 55%, $\chi^2 = 8.09$, $P < 0.05$); Angular Gyrus (70 vs. 48%, $\chi^2 = 3.955$, $P < 0.05$); and Supplementary Motor Area/Supplementary Speech Area (90 vs. 72%, $\chi^2 = 4.561$, $P < 0.05$). There was no difference in the rate of identification of activity consistent with Basal Temporal Language Area (82 vs. 69%, $\chi^2 = 1.81$, n.s.) or Exner's Area (77 vs. 66%, $\chi^2 = 1.234$, n.s.).

DISCUSSION

The hypothesis that experienced clinicians could use this novel method to generate equivalent language maps was supported. While voxel-by-voxel overlap was moderate when clinicians used data from across two sets of three runs, when clinicians generated maps using at least some data from the same set of sequences, maps were significantly more similar, averaging 78% voxel-wise overlap (Dice = 0.74). Our findings supported the hypothesis that language laterality determined from these maps was equivalent to laterality determined by Wada test. Independent clinicians typically read CG map laterality as

consistent with the Wada result (85% of cases) and activation consistent with all six described regions was identified in a majority of cases. There was also consistent evidence that language maps generated using this CG approach differ from those generated using a simple fixed threshold. In comparing CG with fixed threshold maps, laterality indices from CG maps and fixed threshold maps demonstrated similar correspondence with Wada data. When generated using data with at least one sequence from the same set, CG maps' spatial extent differed significantly from that of equivalent fixed threshold maps. Expert reviewers blind to map type concluded two different clinicians using this approach generated maps that were of equivalent quality. Further, they judged these maps of superior quality to those created using a fixed threshold. Finally, these maps localized activation consistent with known language areas more frequently than fixed threshold maps.

Of note is the fact that the two clinicians' maps were more similar when each included at least one run from the same task set. While this could suggest movement was an issue, within-sequence movement was not found to be reduced in these patients. Visual inspection also suggested the images were collinear and that the final conjunction maps were aligned. It is possible, however, that the presence of between-run movement led to slight misalignment of the raw data used to form the conjunction images, and that this is partly responsible. This is unlikely to account fully for this difference, however; clinicians' reports during analysis documented deliberate decisions about which runs should be selected to represent specific language regions.

This finding emphasizes the key role of the clinical question, that is, the subjective goal of the person analyzing data, in defining the results of clinical fMRI. This point is often under-appreciated, and may be less significant when simple lateralization of language, rather than localization, is the goal. Variation as a function of the clinical question can occur through simple alteration of the thresholds used and/or the tasks used in conjunction (e.g., [ON∩VRN∩ARN] as opposed to [VRN∩ARN], or [ON∩ARN]). In the context of different presumed clinical questions, in one instance maps were generated with markedly different temporal and frontal activation (Case O, Fig. 5). In this case, clinicians' laterality indices were equivalent though nearly double the magnitude of those based on a fixed threshold map (Table II). Moreover, in this patient, external reviewers concluded activation consistent with the basal temporal language area was present in only one of the two CG maps. This is consistent with the American College of Radiology's recommendation that "the physician supervising and interpreting fMRI must be clinically informed and understand the specific questions to be answered prior to the procedure" [American College of Radiology, 2014].

This study thus also highlights the importance of the individual who analyzes clinical fMRI data clearly articulating what regions they did and did not seek to identify during analysis. These points are ideally conveyed by the analyst in a report or, better, presentation to the surgical team. The latter is particularly useful when a map is generated to answer one referral question (e.g., possible lateral temporal resection) and at case conference the team unexpectedly considers an alternate procedure. For a similar reason, it can also often be useful to provide a map highlighting basal temporal dropout, to ensure the absence of activation in this region is not misinterpreted as an absence of eloquent cortex.

Comparison with Wada

These data reveal a good overall correspondence between this approach and language laterality determined through Wada testing (85–89%), consistent with the ~85–86% concordance rate reported in reviews of the existing literature and larger case series [Janecek et al., 2013]. Further, absolute discordance–laterality being read as left on Wada, right on fMRI or vice versa–was also similar here (3%) [Janecek et al., 2013]. The relative consistency of laterality indices, paired with the findings on specific language region identification, supports the notion that fMRI-based language lateralization is less sensitive to changes in data analysis choices than language localization. The consistency with Wada data is also congruent with prior reports that use of a panel of multiple language tasks aids in classification [Gaillard et al., 2004].

Of interest is the fact that laterality indices for the two Wada right cases were compellingly right lateralized (−0.27 to −0.48) regardless of method. While lateralization of the mixed dominance case was less clear, laterality indices clustered close to 0 and it is very likely that in the clinical setting such a case would be referred for Wada testing. As such categorizations are highly dependent on the thresholds used to determine laterality, it is of interest that when independent clinicians not involved in analysis reviewed a subset of the clinical maps their judgment was consistent between raters (84–89% agreement) and typically consistent with Wada (85% agreement).

The meaning of fMRI–Wada discordance in the two instances here remains unclear. As with fMRI, there is to some extent a need to artificially dichotomize Wada findings (left v right) and review of these cases' Wada results suggested the lateralization may not have been as clear as at first appeared. The ultimate test of accuracy is the task's ability to predict postsurgical language outcome. Prior work has shown both that fMRI and Wada results diverge when lateralizing language and that fMRI can predict outcomes more accurately [Sabsevitz et al., 2003]. Therefore, although concordance with the Wada procedure is informative, lack of complete concordance does not necessarily mean that the fMRI results would not predict outcome better. Further work validating the current protocol in predicting language change postsurgery would be highly valuable.

Language Localization

We emphasize that the method studied here is not proposed for use in creating a simple template for resection by outlining regions that are critical or indispensable. While the location of the activation and review by external clinicians suggests at least some of these areas represent known language-critical regions, this evidence is correlational and observational. Concurrently, the consistent location of these activations across multiple language tasks argues strongly for their forming part of the language system. If further research supports our findings, this form of language fMRI for localization would be a valuable addition to currently available approaches. The shortcomings of existing methods means novel approaches to localize language are needed. Wada's prediction of postsurgical naming change is suboptimal [Sabsevitz et al., 2003]. Further, 41% of sites report at least one instance of postoperative language decline in spite of the team preserving all eloquent areas identified during intracranial EEG for seizure localization and functional mapping [Hamberger et al., 2014]. These data suggest that while fMRI has its own limitations, it may prove a valuable complement to traditional approaches. While the widespread adoption of fMRI to guide surgical margins is less likely, the use of an fMRI-based approach that is highly (rigidly) standardized to help guide electrode placement in stimulation mapping, quantify risk and predict outcome, would appear reasonable and useful.

The language model used here builds on the work of countless others, and is more heavily clinical than theoretical. It seeks to map areas required for language function; that is, areas that will lead to language deficits if resected, regardless of hypothesized function. Patients and epilepsy teams can then weigh a possible decline in a function (e.g., handwriting) against the probability of postsurgical seizure freedom. The model presented here does not include all clinically relevant regions, however. There is good evidence that the anterior temporal lobe, an area we do not explicitly map, is language-critical with resection resulting in auditory naming deficits [Hamberger et al., 2001], and a protocol for mapping this region has previously been developed [Binder et al., 2011]. Protocols such as this may form a supplement to the approach presented here. It may also be that some of the numerous other language protocols not used here—for example, phonological processing tasks—may be comparable or even superior to this approach.

This latter point touches on a key limitation of studies such as this. When a method such as that used here is used by others, or the results are weighed, it is critical to note that the findings (and their reproducibility) depend fundamentally on the methods used. Paradigms that require different cognitive skills can result in starkly different activation profiles, and any paradigm that is to be used should be validated for that purpose [Binder et al., 2008; Jayakar et al., 2014; Pillai and Zaca, 2011; Zacà et al.,

2012]. The comparison approach we used–a "fixed threshold" approach–must be interpreted with the caveat that the threshold used is just one possible threshold; the use of different thresholds or approaches (e.g., a threshold-independent approach [Branco et al., 2006]) may yield different results. Estimates of reliability will also vary as a function of the measures evaluated. While alternate approaches such as comparing the center of mass could be used, we elected to compare the maps' spatial extent as this is the property of the map that many neurosurgical teams consider when formulating a surgical plan and considering intracranial electrode placement. Similarly, alternative approaches to analysis could have been used. Our goal was to evaluate an approach we have tested extensively and found useful, using a correlation coefficient and minimal preprocessing to keep the data as close to its original form as possible. GLM is another, alternate and widely used approach for image analysis, which has the benefit of easily modeling a large number of effects on the data (e.g., conditions, sessions, movement parameters) but with the side effect of progressively decreasing sensitivity to task-related effects [Taylor et al., 1996]. As such, approaches to addressing movement in GLM-based analysis, which include adding regressors to model movement, can reduce explanatory power and sensitivity. Greater data smoothing is also typically required for clinically feasible sequences. The correlation analysis used here considers only two variables (actual and [HRF-based] predicted blood flow), allowing greater sensitivity to task-related effects. To address movement, contaminated images can simply be removed; while this does not remove all movement effects [Friston et al., 1996] in our experience it can deal with the worst of these without dramatically reducing power. It is important to note that residual effects will remain, however. Further development of our approach with more recent analytic approaches such as GLM may strengthen the protocol described here, as may complementary methods shown to boost reliability such as the normalization of activation maps (AMPLE) [Voyvodic, 2012]. Finally, it is also important to note that a number of the regions mapped here may be considered by various clinicians as part of different structures, or with different names. For instance, dorsolateral prefrontal cortex may be treated as synonymous with Exner's Area, and specifically mapped by some clinicians in this context.

Less is also currently known about the ability to predict deficits in some higher-order aspects of language function, such as grammar, postsurgically. There is evidence that non lexico-semantic aspects of language such as grammar may have a partially distinct neuroarchitecture in epilepsy and tumor patients [Ojemann and Mateer, 1979] and preliminary evidence that considering grammar as part of evaluation may improve the utility of clinical mapping paradigms [Połczyńska et al., 2014; Rofes and Miceli, 2014]. More broadly, this fMRI-based approach focuses on cortical structures and does not explicitly consider the

connections between these regions, or subcortical involvement, and it is imperative that these be considered in planning. In clinical practice, diffusion maps and tractography are typically considered with language fMRI to ensure accurate estimation of risk in surgical planning. Resting state fMRI estimates of connectivity are being investigated and validated, and may prove a useful addition to presurgical planning. This is of particular interest in patients who cannot comply with standard mapping approaches, such as young children. Cortex proximal to key language areas is also vital for and can be recruited with language, such as dorsolateral prefrontal cortex (working memory). Such regions are also vulnerable during surgery and should be considered (and when possible mapped) to avoid postsurgical deficits. Further, language is a more complex process than is reflected by the single-word based tasks used in many forms of clinical language mapping [e.g., Polczynska et al., 2014] and determining whether mapping higher-order language constructs improves prediction of surgical outcome is an important, outstanding question.

The Role of the Clinician in Clinical fMRI

The finding that an experienced clinician generates language maps that both differ from the simple application of a threshold and formation of a conjunction, and that this improves the quality of the map, is of particular importance. These data show that when language fMRI is analyzed by an experienced clinician such as a neuropsychologist, results are both qualitatively (image quality) and quantitatively (identification of activation consistent with specific language regions) improved relative to an automated analysis approach using the same methods (application of a threshold; conjunction) without review and adjustment by a skilled clinician. Further, in this sample when analysis applied a fixed threshold rather than relying on a clinician's expertise, laterality was less often read as consistent with Wada laterality (69 vs. 85%). Errors most often included incorrect judgments of mixed dominance (5/29 cases), though data was also often determined unusable data (4/29). Of note, when a fixed threshold map was read as unusable (5 cases) a clinician analyzing the same data often concluded the data was unusable (40%) or generated a map read as inconsistent with Wada results (40%). This suggests that when data are unusually poor, clinicians should not attempt to make a decision regarding language laterality using the data that they have, but should instead suggest repeating the fMRI or using an alternative method (e.g., Wada testing). Using such data to generate maps regardless may be misleading and yield unreliable findings.

The benefit of clinician-based analysis held for language localization as well as lateralization. There was a consistent pattern for clinician-based analysis to identify more probable language regions (as per external reviewers' judgments). This was true not only for Broca's (97 vs.

79%) and Wernicke's areas (inferior, 98 vs. 69%; superior 83 vs. 55%), but also for less frequently appreciated areas including the Angular Gyrus and Supplementary Speech Area. These findings are not the result of rater bias or the scale used. The external raters, who were independent and blind to map type during data review, reviewed both the fixed threshold and CG maps for each patient. That is, any biases in region definition, quality, or determination of lateralization were applied (blindly) to the same patients' CG and fixed threshold maps. It is likely that the benefit of having clinicians complete analysis came from factors including their knowledge of the language system; clinical considerations; the strengths and weaknesses of fMRI; and the cognitive structure of the tasks being used. Regardless: confirming activation in these areas truly represents language cortex will be vital.

A possible criticism of studies such as this is that they may validate an approach that can only be used by (for example) the two individuals who completed analysis. While this method is complex and requires skill in a number of areas, we have developed and provided additional material (included here as Supplements) so that they might be evaluated and improved by others. We are also happy to provide versions of the tasks used. This approach seeks to identify areas known and identifiable to other clinicians in the field (e.g., LCB, WDG, MBB). We expect it can be used by clinicians (e.g., neuropsychologists, neurologists, radiologists) who have detailed training in and a detailed understanding of areas including cognition, neuroanatomy, functional MRI, image analysis, and epilepsy and others have been trained in this method previously (e.g., CFB).

Limitations

A widely appreciated issue evident in our data was the degree to which noise impacted data, guiding analysis decisions and the resulting language maps. As detailed, we achieved minimal movement and consistent good task engagement by training patients prior to testing, modifying the task version to match the patient's cognitive baseline (e.g., slower stimuli in low IQ/slow processing speed patients), and ensuring immediate (sensory) feedback on movement (provided by tape placed across head coil touching patient's forehead). Other measures influencing signal to noise ratio include the amount of data (greater with more images acquired), extent of signal change (greater with longer block duration), and the removal of random noise through data smoothing and image realignment. Our protocol was sufficient to allow good correspondence with Wada findings and relatively high image quality on independent review. Increasing the number of images acquired during imaging, as with recent subsecond TR multiband sequences, would increase signal. Data were smoothed to decrease noise and accentuate signal. While, ideally, smoothing would not be necessary (signal would be perfectly sampled), the time available to

acquire multiple language tasks in clinic (as per clinical guidelines) is limited. We smooth by a minimal amount (2 mm), which has the benefit of keeping the data as close to its raw spatial resolution (and accuracy) as possible while minimizing attenuation of relevant signal and the smoothing of irrelevant structures (e.g., vessels) into cortex. The downside of smoothing with a smaller kernel is that significant noise will remain, and the spatial extent of some activations may be under-represented. Also of note is that we chose to include a brief, one-second cue at the start of each 10s block reminding patients of the task. While this will influence task-related BOLD signal, our experience is that such a cue is highly valuable, particularly with patients with poor working memory.

CONCLUSIONS

In sum, these results suggest that clinicians experienced in this method can form conjunctions from two sets of three language tasks to generate equivalent maps, and that these maps differ from and are rated as of better quality than those generated with a fixed threshold. This difference did not markedly alter laterality indices, but it did improve the correspondence of clinicians' judgments of language laterality so that they were more consistent with Wada laterality. The caveat here was that when the data were of low quality–to the point where a fixed threshold and conjunction revealed little or no activation–the maps clinicians generated were less likely to be read by others as consistent with Wada results. For localization, however, these data show a standardized, clinician-based approach can use a suite of tasks to identify probable language regions that are both vulnerable in epilepsy surgery and frequently overlooked. These data suggest a fixed threshold without input from a clinician with expert knowledge of cognition and fMRI may lead to inaccurate findings in presurgical planning. Finally, with additional study and refinement, the method described here may improve the yield of other invasive and imperfect methods of language localization, such as extraoperative stimulation mapping, and improve patient outcomes from surgery.

ACKNOWLEDGMENTS

REFERENCES

American College of Radiology (2014): ACR–ASNR–SPR Practice Parameter for the Performance of Functional Magnetic Resonance Imaging (fMRI) of the Brain. Res 39.

Anderson S, Damasio A, Damasio H (1990): Troubled letters but not numbers. Brain 113:749–766.

Baker CI, Liu J, Wald LL, Kwong KK, Benner T, Kanwisher N (2007): Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. Proc Natl Acad Sci USA 104:9087–9092.

Binder JR (2011): Functional MRI is a valid noninvasive alternative to Wada testing. Epilepsy Behav 20:214–222.

Binder JR (2015): The Wernicke area. Neurology 85:2170–2175.

Binder JR, Swanson SJ, Hammeke TA, Morris GL, Mueller WM, Fischer M, Benbadis S, Frost JA, Rao SM, Haughton VM (1996): Determination of language dominance using functional MRI: A comparison with the Wada test. Neurology 46:978–984.

Binder JR, Swanson SJ, Hammeke TA, Sabsevitz DS (2008): A comparison of five fMRI protocols for mapping speech comprehension systems. Epilepsia 49:1980–1997.

Binder JR, Desai RH, Graves WW, Conant LL (2009): Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex 19:2767–2796.

Binder JR, Gross WL, Allendorfer JB, Bonilha L, Chapin J, Edwards JC, Grabowski TJ, Langfitt JT, Loring DW, Lowe MJ, Koenig K, Morgan PS, Ojemann JG, Rorden C, Szaflarski JP, Tivarus ME, Weaver KE (2011): Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. Neuroimage 54:1465–1475.

Bogen JE, Bogen GM (1976): "Wernicke's region — Where is it?" Ann NY Acad Sci 280: 834–843.

Bookheimer S (2002): Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. Annu Rev Neurosci 25:151–188.

Bookheimer S, Zeffiro TA, Blaxton T, Gaillard W, Theodore W (1995): Regional cerebral blood flow during object naming and word reading. Hum Brain Mapp 3:93–106.

Bookheimer SY, Zeffiro TA, Blaxton T, Malow BA, Gaillard WD, Sato S, Kufta C, Fedio P, Theodore WH (1997): A direct comparison of PET activation and electrocortical stimulation mapping for language localization. Neurology 48:1056–1065.

Branco DM, Suarez RO, Whalen S, O'Shea JP, Nelson AP, da Costa JC, Golby AJ (2006): Functional MRI of memory in the hippocampus: Laterality indices may be more meaningful if calculated from whole voxel distributions. Neuroimage 32:592–602.

Burnstine TH, Lesser RP, Hart J, Uematsu S, Zinreich SJ, Krauss GL, Fisher RS, Vining EPG, Gordon B (1990): Characterization of the basal temporal language area in patients with left temporal lobe epilepsy. Neurology 40:966–966.

Dice L (1945): Measures of the Amount of Ecologic Association Between Species. Ecology 26:297–302.

Donnan GA, Carey LM, Saling MM (1999): More (or less) on Broca. Lancet 353:1031–1032.

Dronkers NF (1996): A new brain region for coordinating speech articulation. Nature 384:159–161.

Endo Y, Saito Y, Otsuki T, Takahashi A, Nakata Y, Okada K, Hirozane M, Kaido T, Kaneko Y, Takada E, Okazaki T, Enokizno T, Saito T, Komaki H, Nakagawa E, Sugai K, Sasaki M (2014): Persistent verbal and behavioral deficits after resection of the left supplementary motor area in epilepsy surgery. Brain Dev 36: 74–79.

Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R (1996): Movement-related effects in fMRI time-series. Magn Reson Med 35:346–355.

Gaillard R, Naccache L, Pinel P, Clémenceau S, Volle E, Hasboun D, Dupont S, Baulac M, Dehaene S, Adam C, Cohen L (2006): Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading. Neuron 50:191–204.

Gaillard WD, Pugliese M, Grandin CB, Braniecki SH, Kondapaneni P, Hunter K, Xu B, Petrella JR, Balsamo L, Basso G (2001): Cortical localization of reading in normal children: An fMRI language study. Neurology 57:47–54.

Gaillard WD, Balsamo L, Xu B, Grandin CB, Braniecki SH, Papero PH, Weinstein S, Conry J, Pearl PL, Sachs B, Sato S, Jabbari B, Vezina LG, Frattali C, Theodore WH (2002): Language dominance in partial epilepsy patients identified with an fMRI reading task. Neurology 59:256–265.

Gaillard WD, Balsamo L, Xu B, McKinney C, Papero PH, Weinstein S, Conry J, Pearl PL, Sachs B, Sato S, Vezina LG, Frattali C, Theodore WH (2004): fMRI language task panel improves determination of language dominance. Neurology 63:1403–1408.

Hamberger MJ, Goodman RR, Perrine K, Tamny T (2001): Anatomic dissociation of auditory and visual naming in the lateral temporal cortex. Neurology 56:56–61.

Hamberger MJ, Williams AC, Schevon CA (2014): Extraoperative neurostimulation mapping: Results from an international survey of epilepsy surgery programs. Epilepsia 55:933–939.

Hunter KE, Blaxton TA, Bookheimer SY, Figlozzi C, Gaillard WD, Grandin C, Anyanwu A, Theodore WH (1999): 15O water positron emission tomography in language localization: A study comparing positron emission tomography visual and computerized region of interest analysis with the Wada test. Ann Neurol 45:662–665.

Janecek J, Swanson SJ, Sabsevitz DS, Hammeke TA, Raghavan M, Rozman M, Binder JR (2013): Language lateralization by fMRI and Wada testing in 229 epilepsy patients: Rates and predictors of discordance. Epilepsia 54:314–322.

Jayakar P, Gaillard WD, Tripathi M, Libenson MH, Mathern GW, Cross JH (2014): Diagnostic test utilization in evaluation for resective epilepsy surgery in children. Epilepsia 55:507–518.

Kagan A, Saling MM (1988): An introduction to Luria's aphasiology: Theory and application. Johannesburg: Witwatersrand University Press.

Krainik Lehéricy S, Duffau H, Capelle L, Chainay H, Cornu P, Cohen L, Boch A-L, Mangin J-F, Le Bihan D, Marsault C (2003): Postoperative speech disorder after medial frontal surgery: Role of the supplementary motor area. Neurology 60: 587–594.

Krauss GL, Fisher R, Plate C, Hart J, Uematsu S, Gordon B, Lesser RP (1996): Cognitive effects of resecting basal temporal language areas. Epilepsia 37:476–483.

Loddenkemper T, Morris HH, Möddel G (2008): Complications during the Wada test. Epilepsy Behav 13:551–553.

Loring DW, Meador KJ (2015): The Wada Test: Current Perspectives and Applications. In: Barr, WB, Morrison, C, editors. Handbook on the Neuropsychology of Epilepsy, 1st ed. New York: Springer. pp 123–137.

Lüders H, Lesser R, Hahn H, Dinner D, Morris H, Harrison M (1985): Language Disturbances Produced by Electrical Stimulation of the Basal Temporal Region. Ann Neurol 18:151.

Lüders H, Lesser RP, Hahn J, Dinner DS, Morris H, Resor S, Harrison M (1986): Basal temporal language area demonstrated by electrical stimulation. Neurology 36:505–510.

Mills C, McConnell I (1895): The Naming Center, with the report of a case indicating its location in the temporal lobe. J Nerv Ment Dis 22 1–7.

Ojemann G, Mateer C (1979): Human language cortex: Localization of memory, syntax, and sequential motor-phoneme identification systems. Science (80) 205:1401–1403.

Pardo JV, Fox PT (1993): Preoperative assessment of the cerebral hemispheric dominance for language with CBF PET. Hum Brain Mapp 1:57–68.

Penfield W, Roberts L (1959): Speech and brain mechanisms. Princeton, NJ: Princeton University Press.

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988): Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331:585–589.

Pillai JJ, Zaca D (2011): Relative utility for hemispheric lateralization of different clinical fMRI activation tasks within a comprehensive language paradigm battery in brain tumor patients as assessed by both threshold-dependent and threshold-independent analysis methods. Neuroimage 54 Suppl 1:S136–S145.

Połczyńska M, Benjamin CFA, Moseley BD, Walshaw P, Eliashiv D, Vigil C, Jones M, Bookheimer SY (2014): Role of the Wada test and functional magnetic resonance imaging in preoperative mapping of language and memory: two atypical cases. Neurocase 1–14.

Połczyńska M, Curtiss S, Walshaw P, Siddarth P, Benjamin C, Moseley BD, Vigil C, Jones M, Eliashiv D, Bookheimer S (2014): Grammar tests increase the ability to lateralize language function in the Wada test. Epilepsy Res 108:1864–1873.

Price CJ (2012): A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage 62:816–847.

Rodd JM, Vitello S, Woollams AM, Adank P (2015): Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. Brain Lang 141:89–102.

Rofes A, Miceli G (2014): Language mapping with verbs and sentences in awake surgery: A review. Neuropsychol Rev 24:185–199.

Ropper AH, Brown RH (2005): Principles of neurology, 8th ed. New York: McGraw-Hill.

Roux FE, Boulanouar K, Lotterie JA, Mejdoubi M, LeSage JP, Berry I, Berger MS, McKhann GM, Hirsch J, Schramm J (2003): Language functional magnetic resonance imaging in preoperative assessment of language areas: Correlation with direct cortical stimulation. Neurosurgery 52:1335–1347.

Roux FE, Dufor O, Giussani C, Wamain Y, Draper L, Longcamp M, Démonet JF (2009): The graphemic/motor frontal area Exner's area revisited. Ann Neurol 66:537–545.

Roux FE, Draper L, Köpke B, Démonet JF (2010): Who actually read Exner? Returning to the source of the frontal "writing centre" hypothesis. Cortex 46:1204–1210.

Sabsevitz DS, Swanson SJ, Hammeke TA, Spanaki MV, Possing ET, Morris GL, Mueller WM, Binder JR (2003): Use of preoperative functional neuroimaging to predict language deficits from epilepsy surgery. Neurology 60:1788–1792.

Schaffler L, Luders HO, Morris HH, Wyllie E (1994): Anatomic distribution of cortical language sites in the basal temporal language area in patients with left temporal lobe epilepsy. Epilepsia 35:525–528.

Seghier ML (2008): Laterality index in functional MRI: methodological issues. Magn Reson Imaging 26:594–601.

Sherman EMS, Wiebe S, Fay-Mcclymont TB, Tellez-Zenteno J, Metcalfe A, Hernandez-Ronquillo L, Hader WJ, Jetté N (2011): Neuropsychological outcomes after epilepsy surgery: Systematic review and pooled estimates. Epilepsia 52:857–869.

Suarez RO, Whalen S, Nelson AP, Tie Y, Meadows ME, Radmanesh A, Golby AJ (2009): Threshold-independent functional MRI determination of language dominance: A validation study against clinical gold standards. Epilepsy Behav 16:288–297.

Szaflarski JP, Binder JR, Gaillard WD, Golby AJ, Holland SK, Ojemann J, Spencer DC, Swanson SJ, French JA, Theodore WH (2017): Practice guideline summary: Use of fMRI in the presurgical evaluation of patients with epilepsy Report of the Guideline Development, Dissemination, and Implementation

Subcommittee of the American Academy of Neurology. Neurology 88(4):395–402.

Taylor JMG, Siqueira AL, Weiss RT (1996): The cost of adding parameters to a model. J R Stat Soc Ser B 593–607.

Theodore WH (1989): The role of fluorodeoxyglucose-positron emission tomography in the evaluation of seizure disorders. Semin Neurol 9:301–306.

Tremblay P, Dick AS (2016): Broca and Wernicke are dead, or moving past the classic model of language neurobiology. Brain Lang 162:60–71.

Voyvodic JT (2012): Reproducibility of single-subject fMRI language mapping with AMPLE normalization. J Magn Reson Imaging 36:569–580.

Zacà D, Nickerson JP, Deib G, Pillai JJ (2012): Effectiveness of four different clinical fMRI paradigms for preoperative regional determination of language lateralization in patients with brain tumors. Neuroradiology 54:1015–1025.

Zentner J, Hufnagel A, Pechstein U, Wolf HK, Schramm J (1996): Functional results after resective procedures involving the supplementary motor area. J Neurosurg 85:542–549.