

RESEARCH ARTICLE

Presurgical language fMRI: Clinical practices and patient outcomes in epilepsy surgical planning

Christopher F. A. Benjamin¹  | Alexa X. Li² | Hal Blumenfeld¹ |
 R. Todd Constable¹ | Rafeed Alkawadri¹ | Stephan Bickel³ |
 Christoph Helmstaedter⁴ | Stefano Meletti⁵  | Richard Bronen¹ |
 Simon K. Warfield⁶ | Jurriaan M. Peters⁶ | David Reutens⁷ | Monika Połczyńska⁸ |
 Dennis D. Spencer¹ | Lawrence J. Hirsch¹

¹Yale University School of Medicine, 333 Cedar Ave, New Haven, CT, USA

²Quinnipiac University School of Medicine, 370 Bassett Rd, North Haven, CT, USA

³Northwell Health, Great Neck, NY, USA

⁴University of Bonn, Regina-Pacis-Weg 3, Bonn, Germany

⁵University of Modena and Reggio Emilia, Via Università, 4, Modena, MO, Italy

⁶Boston Children's Hospital, 360 Longwood Avenue Boston, MA, USA

⁷The University of Queensland, St. Lucia QLD, Australia

⁸University of California Los Angeles, Westwood Plaza, CA, USA

Correspondence

Christopher F. A. Benjamin, 800 Howard Ave, LL, New Haven, CT 06511, USA.
 Email: christopher.benjamin@yale.edu

Funding information

Yale CTSA, Grant/Award Number: UL1TR000142; National Center for Advancing Translational Science (NCATS); National Institutes of Health USA; Swebilius Foundation

Abstract

The goal of this study was to document current clinical practice and report patient outcomes in presurgical language functional MRI (fMRI) for epilepsy surgery. Epilepsy surgical programs worldwide were surveyed as to the utility, implementation, and efficacy of language fMRI in the clinic; 82 programs responded. Respondents were predominantly US (61%) academic programs (85%), and evaluated adults (44%), adults and children (40%), or children only (16%). Nearly all (96%) reported using language fMRI. Surprisingly, fMRI is used to guide surgical margins (44% of programs) as well as lateralize language (100%). Sites using fMRI for localization most often use a distance margin around activation of 10mm. While considered useful, 56% of programs reported at least one instance of disagreement with other measures. Direct brain stimulation typically confirmed fMRI findings (74%) when guiding margins, but instances of unpredicted decline were reported by 17% of programs and 54% reported unexpected preservation of function. Programs reporting unexpected decline did not clearly differ from those which did not. Clinicians using fMRI to guide surgical margins do not typically map known language-critical areas beyond Broca's and Wernicke's. This initial data shows many clinical teams are confident using fMRI not only for language lateralization but also to guide surgical margins. Reported cases of unexpected language preservation when fMRI activation is resected, and cases of language decline when it is not, emphasize a critical need for further validation. Comprehensive studies comparing commonly-used fMRI paradigms to predict stimulation mapping and post-surgical language decline remain of high importance.

KEYWORDS

epilepsy, fMRI, language, presurgical

1 | INTRODUCTION

Neurosurgery is an effective and potentially curative treatment for temporal lobe epilepsy (Wiebe, Blume, Girvin, & Eliasziw, 2001). Surgical risk to language and memory can exclude a patient from treatment.

As 34%–41% of left temporal patients undergoing focal resections experience a decline in naming (Busch et al., 2016; Sherman et al., 2011), determining the surgical risk to language remains essential.

While the Intracarotid Amobarbital Test ("Wada" testing) has been the gold standard for determining the language dominant hemisphere,

.....
 This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors Human Brain Mapping Published by Wiley Periodicals, Inc.

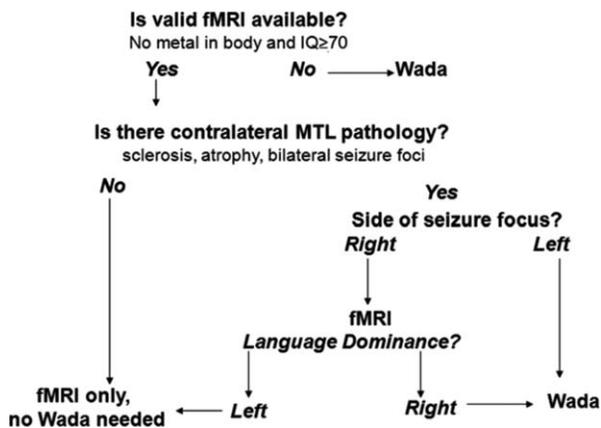


FIGURE 1 An algorithm for determining when to conduct fMRI or Wada testing. This heuristic may be modified, for example, based on baseline memory function, and is based on a particular language task (Semantic Decision Making task; Binder et al., 1995) Source. Republished with permission of Springer New York from Handbook on the Neuropsychology of Epilepsy, William B. Barr and Chris Morrison (Eds.), Edition 1 (2015), Chapter 8, “Functional MRI in the Presurgical Epilepsy Evaluation” by Swanson, S., Binder, J.R., Raghavan, M., & Euler, M. Permission conveyed through Copyright Clearance Center, Inc.

functional magnetic resonance imaging (fMRI), which combines cognitive assessment with MRI, is non-invasive and can be accurate (Sabsevitz et al., 2003) and less costly (Medina, Aguirre, Bernal, & Altman, 2004). The evidence supporting fMRI’s validity was recently outlined (Szaflarski et al., 2017), with the conclusion that language fMRI is a valid alternative to Wada testing in most patients. One approach to incorporating fMRI in clinical decision making developed by Swanson and colleagues is shown in Figure 1 (Swanson, Binder, Raghavan, & Euler, 2015). In short, if language fMRI shows left hemisphere dominance and a patient has right hemisphere pathology, Wada testing may be deemed unnecessary for language lateralization. Given the significance of a right hemisphere language finding and the increased probability of discordance with Wada (Janecek et al., 2013), an argument can be made for repeating language fMRI for right dominance findings.

While different language fMRI tasks will yield differing maps and accuracy in predicting post-operative decline (Binder, Swanson, Hammeke, & Sabsevitz, 2008), good estimates of fMRI’s validity in a tertiary epilepsy setting are available for a semantic decision making task. With this protocol, a key study found decline of more than two standard deviations in naming skill (relative to controls) could be predicted with 100% sensitivity and 73% specificity, with prediction superior to that using Wada (92% and 45%, respectively; Sabsevitz et al., 2003). Overall, 41% of variance in post-operative language skill was predicted by fMRI and the positive predictive value was 81% for fMRI and 67% for Wada. In 229 epilepsy patients, using a slightly different analysis, 80% were classified as left dominant using fMRI, of which 92% were also Wada left dominant (167/182; 15 bilateral; Janecek et al., 2013). fMRI bilateral cases ($n = 28$) were typically left (46%) or bilateral (36%) on Wada, though occasionally (18%) right. fMRI right cases were most

often also right on Wada (53%), though could be markedly discrepant (21% [$n = 4$] left Wada; all right handed, right seizure foci). Explicitly examining cases of fMRI–Wada disagreement, this protocol has been shown to more accurately predict naming decline than Wada testing (Sabsevitz et al., 2003). BOLD fMRI language maps are not validated for the routine drawing of boundaries around indispensable cortex (the removal of which is associated with language decline), however.

Importantly, fMRI is not validated for comprehensive language mapping to tailor surgical margins to avoid post-surgical language decline (Giussani et al., 2010; Szaflarski et al., 2017). Clinical guidelines for fMRI use are available within radiology (American College of Radiology, 2014) and neuropsychology (Bobholz et al., 2004). Recent surveys from the European Union’s E-PILEPSY project showed 82% of European epilepsy programs use language fMRI (Mouthaan et al., 2016), with Wada being used more judiciously. The questions clinical teams ask of language fMRI, how the results are interpreted, and whether they improve clinical care, however, remain unclear. Whether current clinical practices are in keeping with current best evidence is also unknown. The goals of this study were to characterize how epilepsy programs use language fMRI in surgical planning and aggregate reports of patient outcomes that are otherwise likely to go unreported, to help inform future validation and study of clinical language fMRI.

2 | MATERIALS AND METHODS

This study was approved and overseen by Yale Medical School’s Institutional Review Board. All participants provided informed consent (Supplement A).

2.1 | Survey

A survey focused on clinicians’ use of and experience with presurgical language fMRI was developed (Supplement A). Questions centered on identifying the dominant hemisphere (“lateralization”); identifying language areas to guide surgical margins (“localization”); how programs use fMRI; their confidence in results; and patient outcomes. A past survey of extraoperative mapping was used as a reference in design (Hamberger, Williams, & Schevon, 2014). Clinicians from neurology, radiology and neurosurgery provided feedback on clarity and length. The research design was reviewed by Yale’s Center for Analytic Sciences. A second, related survey of those acquiring fMRI data will be reported separately. The survey was presented via www.qualtrics.com. Questions were organized hierarchically with all respondents answering key questions which could elicit related questions. If a question was not answered a warning appeared, but the respondent could elect to continue without answering.

2.2 | Procedure

2.2.1 | Site identification

Within the US the major body formally accrediting epilepsy programs, the National Association of Epilepsy Centers (NAEC), provided details of all epilepsy programs. We invited all programs completing surgery

(levels 3 or 4) and Wada or fMRI. Worldwide, prominent epilepsy organizations (e.g., ILAE members) and researchers were contacted. We asked individuals to invite other programs they knew, to increase sample size, using a modified “snowball sampling” approach (Goodman, 1961).

2.2.2 | Data collection

The survey was open from July 17, 2015 to January 15, 2016. We emailed invitations to all NAEC sites and followed up by telephone. We also contacted researchers, epilepsy organizations, and ILAE member committees worldwide requesting they forward an email invitation to relevant contacts. In 11.2015 we sent reminders and emailed the American Epilepsy Society (AES) listserv.

2.2.3 | Final sample

Eighty-two surveys were received from respondents involved in selecting patients for epilepsy surgery. The number of responses per question is indicated in square brackets; for example, ($n = X$). Responses vary due to the study's hierarchical design; some specific responses elicited additional relevant questions. Data were cleaned; of note, two respondents answered only initial questions on their confidence in methods for language lateralization, and three reported no patients received fMRI. Detailed review of the 79% of responses for which an institution was identifiable showed one (1.5%) could reflect a duplicate response from a program. As it may have reflected a unique response and survey instructions emphasized provision of one clinical response (Supporting Information A), the data were not removed. A second pair was identified, but it was apparent that the respondents work in different programs, but with fMRI executed by the same individual. Respondent's country was estimated from stated location (per the survey or email) or IP address. Descriptive statistics are presented and compared using t and Fisher's Exact tests.

Programs were typically university-affiliated (85%), evaluating a mean of 106 (SD 67; 10–300) patients annually and surgically treating a mean of 34 (SD 21, 0–100; $n = 80$ respondents). Overall 84% evaluated adults and 56% children; specifically—44% evaluated predominantly adults; 40% predominantly both adults and children; and 16% predominantly children (<18 years; $n = 80$). Respondents who identified their background included neurologists (89%), neurosurgeons (6%) and neuropsychologists (5%; $n = 66$). They were typically clinician-researchers (52%) or clinicians (45%; $n = 65$), and 93% reported direct involvement in deciding whether patients are offered surgery ($n = 68$). Just under half (46%) were surgical program directors ($n = 68$). Nine respondents reported they also collect, analyze and interpret the fMRI data.

Almost two thirds of respondents were from the USA (50); with further responses from Australia (6); Canada (3); France (3); Italy (3); Turkey (3); England (2); Germany (2); Denmark (1); Egypt (1); Georgia (1); Japan (1); Netherlands (1); Norway (1); Portugal (1); Scotland (1); Sweden (1); and Switzerland (1).

3 | RESULTS

Through the following results, the number of responses available for each question is indicated in square brackets at the end of the relevant sentence (e.g., $n = X$). See Section 2.2.3 for discussion of hierarchical survey structure, data cleaning, and variable response numbers.

3.1 | Methods used clinically for language lateralization

Nearly all respondents reported using fMRI (96% of programs), neuropsychological assessment (99%) and extraoperative stimulation mapping (93%) in evaluating language preoperatively (total respondents: $n = 80$; Figure 2). The majority of programs also use intraoperative stimulation mapping (ISM; 83%), with other methods each used by fewer than half of programs. Patients are most likely to receive neuropsychological assessment (93% of patients), fMRI (58%) or Wada testing (43%; $n = 80$) at programs using these methods. The three programs not using fMRI (two US, one worldwide) reported they did not have fMRI capabilities, or felt it was not necessary given, for example, availability of Wada. Clinicians reported high confidence that, with all methods at their disposal, they can identify a patient's language dominant hemisphere (mean 92%, SD 8; 60–100) and localize language regions to guide surgical margins (mean 84%, SD 14, 30–100; $n = 81$).

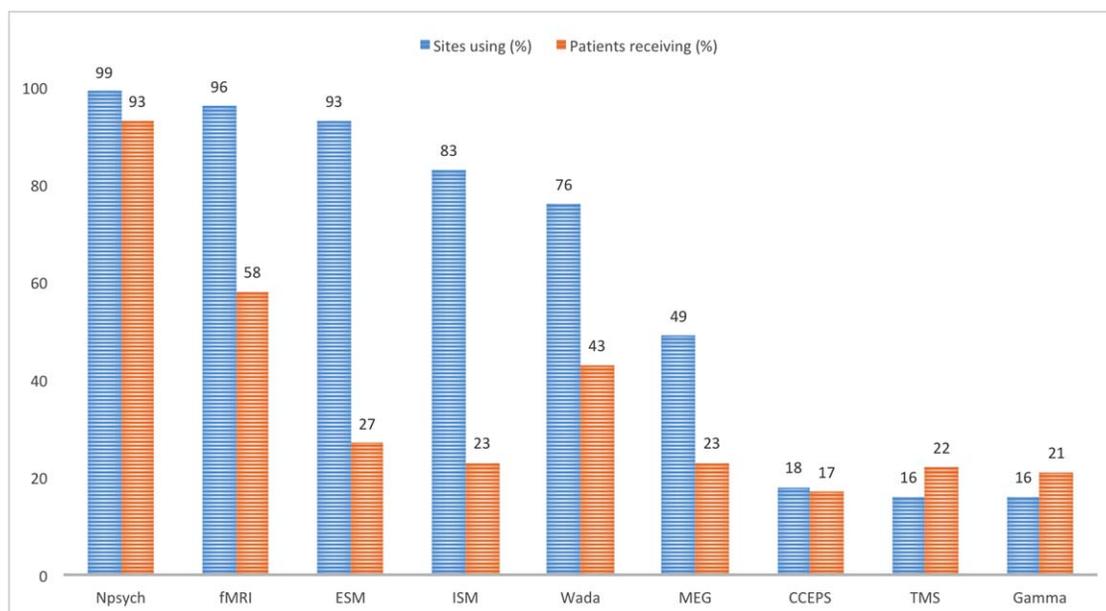
3.2 | Clinicians' opinions of language fMRI

Programs using fMRI reported confidence (73%, SD 18, range 1–100) in their program's ability to identify a patient's language dominant hemisphere using fMRI alone ($n = 78$). They reported low confidence (45%, SD 27, 0–92) in their ability to use fMRI to identify specific language areas to guide surgical boundaries (i.e., localization; $n = 77$). Accordingly, when considering the technique of fMRI generally they considered it reliable for language lateralization (81% confidence, SD 15 19–100) but less so for localization (i.e., identifying language areas to guide surgical boundaries; 48%, SD 25 0–93; $t_{(74)} = 11.471$ $p < .001$; $n = 76, 78$). When asked if language fMRI completed at different centers yields equivalent results, their response was neutral (score of 0 SD 2.5; confidence scale: strongly disagree to strongly agree [–5 to +5]).

3.3 | Language lateralization with fMRI

All programs administering fMRI use it to identify the language dominant hemisphere ($n = 71$). Just under half of programs (46%) reported fMRI has never disagreed with other measures of laterality. A third of programs (34%) reported at least one instance of disagreement with Wada; a quarter (24%) with stimulation mapping; and 15% with other methods (e.g., MEG, semiology, TMS, neuropsychological testing; $n = 74$). Multiple programs noted fMRI yields bilateral or equivocal findings more often than other methods, particularly in children.

In cases of disagreement fMRI was most often judged to have been incorrect (55%); a third reported cases where true lateralization remained unknown (29%); and a fifth (18%) reported cases where fMRI had been correct and the other modality was not ($n = 38$). Five



| | Npsych | fMRI | ESM | ISM | Wada | MEG | CCEPS | TMS | Gamma |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Worldwide (n=80) | | | | | | | | | |
| Sites using | 99% | 96% | 93% | 83% | 76% | 49% | 18% | 16% | 16% |
| Patients receiving* | 93% (15) | 58% (27) | 27% (22) | 23% (21) | 43% (35) | 23% (25) | 17% (26) | 22% (30) | 21% (23) |
| – Median | 100% | 60% | 20% | 12% | 31% | 10% | 9% | 10% | 10% |
| – Range | [33-100] | [9-100] | [1-93] | [1-80] | [2-100] | [1-100] | [1-100] | [1-94] | [1-70] |
| United States (n=48) | | | | | | | | | |
| Sites using | 98% | 96% | 96% | 81% | 92% | 63% | 19% | 17% | 25% |
| Patients receiving* | 94% [13] | 57% [25] | 32% [24] | 23% [21] | 54% [34] | 22% [25] | 20% [33] | 24% [37] | 22% [24] |
| Worldwide, excluding United States (n=32) | | | | | | | | | |
| Sites using | 97% | 94% | 84% | 81% | 50% | 28% | 16% | 16% | 3% |
| Patients receiving* | 91% [18] | 60% [30] | 18% [14] | 22% [23] | 16% [20] | 25% [28] | 12% [7] | 20% [17] | 7% |

FIGURE 2 Language mapping methods: Proportion of programs using (left bars; blue) and patients receiving (right bars; orange) each method (overall and by geographic region). The proportion of patients receiving a measure is estimated based only on programs using that method—that is, at the 83% of sites using ISM, on average 23% of patients receive ISM. Neuropsychological assessment (Npsych), functional MRI (fMRI), extraoperative stimulation mapping (ESM), intraoperative stimulation mapping (ISM), Wada testing (Wada), Magnetoencephalography (MEG), Cortico-cortical evoked potentials (CCEPS), transcranial magnetic stimulation (TMS), Gamma-activation mapping (Gamma). Question: “Please estimate the proportion of surgical candidates at your center who receive the following prior to surgery to clarify language organization” [Color figure can be viewed at wileyonlinelibrary.com]

programs reported cases of Wada-fMRI disagreement where fMRI was judged correct. None had been published, though one was included “in general in wider publications.” Of all programs reporting disagreement between fMRI and another measure, two (5%) had published these cases ($n = 39$).

3.4 | Guiding surgical margins (localization) with fMRI

Forty four percent of programs reported using fMRI to localize language ($n = 71$). Most centers using fMRI to localize language (guide surgical margins and preserve language cortex) seek to map known language areas (86%; $n = 29$). These include Broca’s (86%) and Wernicke’s areas (76%). Other areas which can be language critical, “basal temporal language area” (BTLA; 17%) and other regions (21%), are mapped infrequently.

Programs that do not use fMRI to localize language (i.e., programs using fMRI only for language lateralization) less often seek to map

known language regions (58%; Fisher’s exact $p < .001$; $n = 40$). Specifically, these programs map Broca’s (43%; $p < .001$) and Wernicke’s Areas (38%; $p = .003$) less frequently than programs using fMRI for localization, and also rarely map BTLA (5%, $p = .122$) or other regions (5%; $p = .062$).

Clinicians using fMRI to guide surgical margins report success in mapping Broca’s 75% of the time ($SD 15\%$, 39%–93%; $n = 25$); in mapping Wernicke’s 71% of the time ($SD 16\%$, 30%–92%; $n = 22$); BTLA 62% of the time ($SD 28\%$, 19%–90%; $n = 5$) and other regions 52% of the time ($SD 31\%$, 3%–80%; $n = 6$). The few who use fMRI for localization, but do not seek to identify specific language regions ($n = 4$), find fMRI is rarely successful (27% of cases) in guiding margins.

Direct electrical stimulation at fMRI-positive cortex typically confirms fMRI findings (74% of the time; $SD 15\%$, 30%–95%), with most of these respondents (86%) reporting they specifically seek to map certain language regions ($n = 28$). Instances of post-operative language decline were reported with equivalent frequency (26%) by both

programs using fMRI for lateralization only ($n = 27$) and those using it for both lateralization and localization ($n = 23$).

3.5 | Cognitive outcomes when fMRI activation is preserved

When no fMRI-positive language cortex is resected, half of responding programs (49%) reported no cases of persisting language decline three months post-surgery ($n = 75$). Seventeen percent reported at least one otherwise unexplained instance of language decline following temporal (67%), frontal (50%) or less often parietal (33%) or occipital (8%) surgery ($n = 12$). Of note, most patients (70%) receive postoperative neuropsychological testing at most (96%) programs ($n = 79$). None of these cases of decline had been published. A third of programs did not know if cases of decline had occurred.

Programs noting cases of decline most often related these to surgical variables (100% reported at least one such case) and 23% reported instances where fMRI-related factors were considered responsible. Surgically, resection of white matter language tracts was most often noted (85%). Resection too close to fMRI-positive areas was also often reported (38%), as were surgical complications (23%); multiple subpial transections over eloquent cortex (15%); and surgical injury outside the planned resection area (15%). Resection in language association cortex and post-surgical inflammation was also noted, and 31% reported cases where the cause was unknown. With respect to fMRI-related factors, different programs attributed at least one instance of unanticipated deficits to imaging (e.g., artifact, insufficient resolution; $n = 1$); analysis ($n = 1$); and patient-related factors (e.g., seizures, movement; $n = 1$).

In exploratory analysis, programs reporting cases of cognitive decline in spite of fMRI-active areas being preserved did not differ from those which did not on a range of variables (Supplement C). While a nonsignificant difference, in this sample programs reporting unexpected decline more often predominantly evaluated children (31% vs. 11%); were less likely to review the actual fMRI images at conference (69% vs. 89%); and less often had the individual involved in fMRI data analysis interpret data at surgical conference (38% vs. 59%).

3.6 | Cognitive outcomes when fMRI activation is resected

Half of the programs using fMRI to guide surgical margins (54%) reported cases where patients maintained function after resection of language-positive sites ($n = 26$). These followed frontal (62%), temporal (54%), or parietal (23%) resections. Select sites noted surgery in the left operculum (bilaterally active), posterotemporal cortex (unspecified); right posterotemporal cortex (bilateral activation) and middle frontal gyrus. None had been published.

In exploratory analysis, these programs were less likely to use the Wada (in at least some instances) to evaluate language organization (Supplement C), while use of language fMRI was equivalent (61% [SD 25] vs. 64% [SD 24; $n = 14, 12$]). Other nonsignificant differences in

this sample included programs reporting unexpected preservation more often loading images into an intraoperative system (64% vs. 33%); using fMRI to localize language and guide surgical margins (57% vs. 33%); and completing more surgeries annually (42 [SD 21] vs. 34 [SD 16]).

Language outcomes at sites using (76%) and not using (24%) Wada testing were also compared post-hoc ($n = 80$). There was no difference in the proportion of sites reporting cases of persistent (>3 months) post-operative decline in function when fMRI-positive language sites were preserved (sites not using Wada: 20%, sites using Wada 29%; Fisher's exact $p = .728$; $n = 15, 35$). There was also no difference in cases of maintained pre-operative ability when fMRI-positive cortex was resected, though there was a trend towards sites not using the Wada to more often report preserved post-operative function (not using Wada: 86%, using Wada 42%; Fisher's exact $p = .081$; $n = 7, 19$).

3.7 | Geographic variation in methods

Based on previous data showing geographic differences in Wada use within and outside the United States, we completed post-hoc analyses evaluating geographic differences in key study findings. Significantly fewer respondents from non-US programs reported using the Wada (non-US 50%, US 92%; Fisher's exact $p < .000$; $n = 32, n = 48$), MEG (non-US 28%, US 63%, $p = .003$), and Gamma mapping (non-US 3%, US 25%, $p = .012$) to clarify language organization (Figure 2). Programs outside the US were significantly more likely to use fMRI for language localization (61%) as compared with those within the US (33%; $\chi^2 = 5.465, p < .05$; $n = 28, n = 43$). There were no such differences in the proportion of sites reporting fMRI had disagreed with other methods of determining language laterality (non-US 48%, US 58%; $\chi^2 = 0.641, p = .423$; $n = 29, n = 45$); in reports of patients experiencing persistent post-operative language decline when all fMRI-positive language sites were preserved (non-US 33%, US 19%; $\chi^2 = 1.29, p = .256$; $n = 24, n = 26$); in unexpected language preservation when fMRI-positive language cortex was resected (non-US 64%, US 42%; Fisher's exact $p = .431$; $n = 14, n = 12$); or in the number of programs identifying any language areas beyond Broca's and Wernicke's cortex (non-US 24%; US 25%; Fisher's exact $p = 1.0$) ($n = 17, n = 16$).

3.8 | How clinicians interpret fMRI maps

Teams who routinely use fMRI to guide surgical margins most often constrain resections to a given distance from fMRI activation (59% of programs; $n = 29$). The distance threshold used by programs varied from 3 to 50 mm (average 15 mm, SD 12). The most commonly-used margin was 10 mm (42% of programs). Programs often (28%) will not operate in the same gyrus as activation or will use other criteria (28%), most often routine investigation of activation with direct stimulation. These programs most often report they would resect fMRI-positive cortex in at least some circumstances (79%), for instance if cleared by direct stimulation (62%); if activation was not anatomically consistent

with a language area (38%); if deficits would likely be temporary (e.g., uncomplicated unilateral SMA resection; 38%); or if the patient was willing to accept post-surgical deficits (24%; $n = 29$). One site noted they would resect any fMRI-positive activation “not (in) a primary language site (Broca/Wernicke)” (this site reported both unpredicted decline and unexpected preservation), while one site was uncertain,

and 17% reported they would never resect cortex that was fMRI language-positive.

To interpret fMRI, teams most often review the images (visually) at surgical conference (86% of programs) and the team or referrer reviews a written report (71%; $n = 79$). The individual involved in analysis often interprets the data for the team (63%). Laterality indices are rarely used (whole brain, 28%; specific region, 18%) and a third of programs load data into an intraoperative mapping system (35%).

Respondents expressed confidence that fMRI's ability to lateralize language will improve with further technical advances (mean = 89%, $SD 12$) though were less confident about future advances improving fMRI's utility for localization (mean = 64%, $SD 26$; $n = 77$) ($t_{(75)} = 8.831, p < .001$).

4 | DISCUSSION

These results confirm fMRI is well established for language lateralization, and in contrast to expectations is already used in many epilepsy programs to guide surgical margins. Consistent with the research literature, most clinical programs using language fMRI (54%) reported instances where fMRI language lateralization had conflicted with other methods including the Wada (34%). In these situations it is rarely clear which method is “correct,” though when they could do so respondents judged fMRI incorrect more often (55%). These findings may suggest the fMRI tasks used by different programs vary in their ability to predict language outcome; reflect a decline in language not captured by

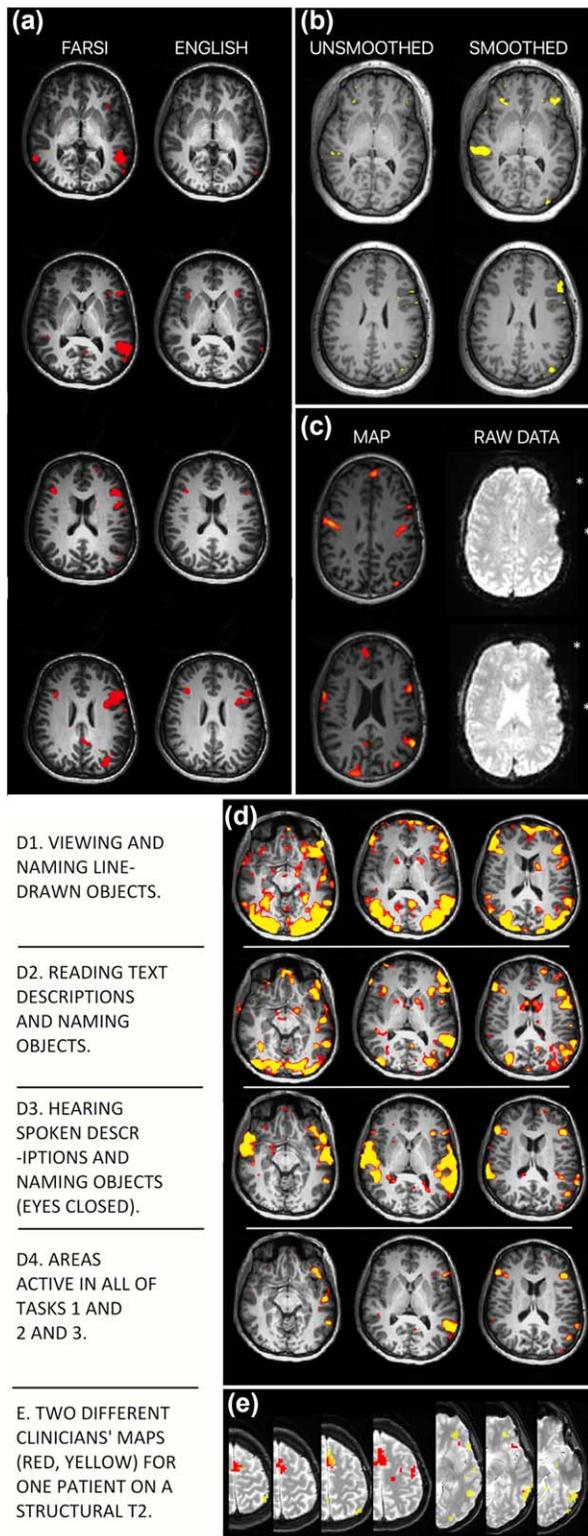


FIGURE 3 Language maps will vary in clinically meaningful ways due to multiple variables. Surgical teams can manage these factors by using experts in both imaging (e.g., radiology) and cognition (e.g., neuropsychology) in clinical fMRI design, analysis and interpretation. (a) Language skill. A patient's language ability will change their activation map. Maps using the same tasks in Farsi and English from a patient who reported fluency in, and made medical decisions in English. (b) Data analysis. Each analysis step changes the map. Data “smoothing” removes noise. Whether it is appropriate, and to what degree, is debated. The degree of smoothing in commercial software may be unspecified. Identical analysis without (left) and with (right) smoothing (8 mm kernel). (c) Data quality. A statistical map (left) does not show where raw data are missing (right, asterisks). These areas will not be active even if they are language critical. This map (left) was presented to a surgical team for surgical planning without caveat. (d) Cognitive task. Different language tasks give different maps. Subtle changes in task instructions, patient motivation and cognitive strategy change language maps. (D1) Visual object, (D2) text reading and (D3) auditory tasks are shown as well as (D4) the intersection of these. (e) Analyst expectations. The analyst's perceived goal will change the activation map. Two overlaid maps (red; yellow) generated independently by two clinicians for the same patient (see Benjamin et al., 2017). Analysts were blind to case details. One prioritized frontal (red) and the other temporal (yellow) regions, as when mapping frontal tumor versus temporal lobectomy cases. Overlap in orange.

Source. (e) is reprinted from Benjamin et al., “Presurgical language fMRI,” (2017); Creative Commons Attribution-Non-Commercial License [Color figure can be viewed at wileyonlinelibrary.com]

visual naming tasks (Hamberger & Seidel, 2003); or represent recall bias and clinicians' uncertainty regarding fMRI.

While fMRI is not validated for language localization, programs have already begun to cautiously use it for this purpose (44%), typically removing fMRI-positive cortex only after confirmation via direct stimulation, counseling patients on possible decline, or never removing fMRI-positive cortex. This is consistent with clinicians' reports in the literature that they consider language fMRI appropriate "to guide surgical margins and preserve language cortex" and that "fMRI... can be used with the same results as awake craniotomy" (Ganslandt, Nimsky, Buchfelder, & Grummich, 2016). This emphasizes the importance of further validating highly standardized and easily replicable forms of fMRI (Mouthaan et al., 2016). Cases of unexpected preservation and unexpected decline are equally concerning; the former may increase post-surgical cognitive impairment while the latter may increase surgical failure rates (Jakola, Unsgård, & Solheim, 2011). Until validated, any criteria used to "preserve" functional cortex relative to fMRI map boundaries in surgery are arbitrary. Excluding patients based on fMRI results regardless of direct cortical stimulation findings, as reported by 17% of programs, may unnecessarily restrict resections and decrease the probability that patients achieve seizure freedom.

Unlike researchers, clinicians interpreting fMRI often appear unaware that rather than showing the immutable boundaries of language-critical cortex, fMRI language maps reveal the probable location of language-related areas. They are likely also unaware that language maps reflect assumptions made by the individuals who selected the cognitive task and imaging sequences, and analyzed and reported the data. These assumptions can, of course, dramatically change map boundaries. Examples are shown in Figure 3. Researchers presenting findings to clinical teams may improve patient care by helping their colleagues understand that the apparent boundaries of language areas will vary with a patient's language skill (Figure 3a); analysis parameters (Figure 3b); signal loss not evident in the final map (Figure 3c); the task and control conditions used (Figure 3d); and the areas the analyst is asked (or not asked) to identify (Figure 3e). Just as importantly, the fact that language fMRI represents task-correlated changes in blood oxygenation (rather than neural activation) and may show non-critical regions may be emphasized Polczynska et al., 2017).

The role of the clinician in defining the results is of particular importance given clinicians' current focus on Broca's and Wernicke's but not other known language-critical regions (e.g., BTLA, Exner's area, SMA; Anderson, Damasio, & Damasio, 1990; Benjamin et al., 2017; Krainik et al., 2003; Krauss et al., 1996; Roux et al., 2009). More sophisticated models of language have been suggested for well over a century in the clinical literature (e.g., see Benjamin et al., 2017 for a review) and have been a recent focus in the cognitive literature (e.g. Tremblay & Dick, 2016). Of particular relevance to clinicians is a compelling recent reanalysis and reinterpretation of the nature of Wernicke's area (Binder, 2015), which suggests discrete bilateral and unilateral components in typically organized individuals (see also Price, 2012), and the suggestion that more sophisticated tasks (e.g., evaluating grammar) may better map language regions such as the angular gyrus (Polczynska et al., 2017).

In a clinical setting, many of these critical factors are obscured when tasks and analysis packages from commercial MR vendors are used. US clinicians are not mandated to use FDA approved software, and delegating these considerations is risky when perhaps the best validated language protocols and analysis software to date are open-source and freely available (Supplement B). The clinician is ultimately responsible for their results. The skills required for clinical fMRI (selecting sequences and cognitive paradigms; analyzing and interpreting data) do not fit within one existing discipline, and this represents an opportunity for research neuroscientists to improve clinical care. When a "gold standard" for language fMRI develops, it will likely approximate the interdisciplinary Wada protocol with a team of at least two qualified professionals from a selection of radiology, neuropsychology, neuroscience, and/or engineering. Expertise in the clinical, imaging and cognitive skills required for fMRI, rather than an individual's discipline, will likely best predict quality of care. This is already acknowledged in US billing codes for fMRI, that allow for psychologists and medical doctors to complete and bill fMRI.

The above discussion fits with the observation that, in our sample, clinical programs that reported unexpected decline were less likely to review the data in surgical conference (69% vs. 89%) and less likely to have the individual who completed analysis interpret data at conference (38% vs. 59%; though note that neither reached statistical significance). The greater use of the Wada for language lateralization by programs not reporting unexpected decline, despite their using language fMRI as regularly as other programs, may reflect a more cautious supplementation of fMRI with Wada when results are equivocal or uncertain.

4.1 | Limitations

A limitation of this report is our inability to relate these findings to the specific language tasks used. We attempted to make this link by asking respondents to have those acquiring and analyzing fMRI at their program complete a paired survey (see Supplement A). Respondents rarely did so, potentially due to the time burden of this survey (often 15–20 min). Thus this project provides a valuable broad overview of how clinicians are using fMRI, and future large-scale research with primary data and multiple paradigms remains critical to inform the relationship between different tasks, language outcomes, and the results of direct cortical stimulation mapping. Of note, variation in protocols is likely to influence overall network lateralization less than regional localization. It is possible respondents' recall and reports are imperfect, and it is not possible for us to judge what reports of unexpected decline or preservation are based on. Of particular note, in considering patient outcomes some respondents may have used formal outcome data while others will have relied on their memory, so that some responses may variably reflect objective outcomes or clinicians' beliefs. These data better represent American (61%) and academic (85%) epilepsy programs. To increase our response rate, collaborators forwarded the survey to numerous colleagues to maximize our sample size. This meant that we were not able to identify specifically how many sites were offered the opportunity to participate, however, obscuring the true

response rate. As an estimate, we contacted 221 US NAEC programs and received 50 U.S. responses, suggesting a 23% response rate (using this approach Hamberger et al. (2014) received 39 U.S. responses). Note also that our prioritization of respondents' anonymity makes it difficult to identify duplicate responses, though fewer than 2% of identifiable responses were duplicates, and any duplicates will likely both be from neurologists (89% of respondents). A major limitation of this work, however, is that the results here include data from only one primarily Asian site (Japan), and do not include data from other Asian programs. Future work might specifically target programs in these regions to ensure their representation. The survey length will also have decreased responses. Regardless, with these flaws in mind, our respondents' data provides valuable, detailed data for the first time on how fMRI is applied in the clinic and moves that can be taken to improve its use.

5 | CONCLUSIONS

Clinical fMRI is widely used to predict language laterality and post-surgical language change. In contrast to research evidence, it is also used for guiding surgical margins, which may lead to either unnecessary caution or over-confidence in surgery. The caveats documented in the literature—for example, occasional disagreement with Wada—are seen in the clinic. Outcomes can likely be improved through use of existing, well validated tasks. Our data, from centers using a range of tasks and methods, emphasize that cautious use of language fMRI for lateralization is warranted and that fMRI maps cannot simply be treated as representing language-critical cortex. Without standardization and explicit validation, any criteria for preserving language-critical cortex relative to fMRI map boundaries are arbitrary. We suggest an initial minimum for clinical care might involve ensuring those who analyze a program's language fMRI, and understand the task's cognitive design, interpret the data in 3D in discussion with the team at conference. They might begin by reviewing the task itself (expected activation), the best estimates of its sensitivity and specificity, and the limitations of the specific results (patient factors, areas of signal loss). This will reduce opportunity for misinterpretation and likely improve patient care.

ACKNOWLEDGMENTS

This article is dedicated to our respondents; thank you for donating the time to make this work possible.

CONFLICTS OF INTEREST

None of the authors has any conflict of interest to disclose.

ORCID

Christopher F. A. Benjamin  <http://orcid.org/0000-0002-2000-9016>

Stefano Meletti  <http://orcid.org/0000-0003-0334-539X>

REFERENCES

- American College of Radiology (2014). ACR–ASNR–SPR practice parameter for the performance of functional magnetic resonance imaging (fMRI) of the brain. *Amended*, 2014, (Resolution 39).
- Anderson, S. W., Damasio, A. R., & Damasio, H. (1990). Troubled letters but not numbers. *Brain*, 113(3), 749–766.
- Benjamin, C. F., Walshaw, P. D., Hale, K., Gaillard, W. D., Baxter, L. C., Berl, M. M., . . . Bookheimer, S. Y. (2017). Presurgical language fMRI: Mapping of six critical regions. *Human Brain Mapping*, 38(8), 4239.
- Binder, J. R. (2015). The Wernicke area: Modern evidence and a reinterpretation. *Neurology*, 85(24), 2170–2175.
- Binder, J. R., Rao, S. M., Hammeke, T. A., Frost, J. A., Bandettini, P. A., Jesmanowicz, A., & Hyde, J. S. (1995). Lateralized human brain language systems demonstrated by task subtraction functional magnetic resonance imaging. *Arch Neurol*, 52, 593–601.
- Binder, J. R., Swanson, S. J., Hammeke, T. A., & Sabsevitz, D. S. (2008). A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia*, 49(12), 1980–1997.
- Bobholz, J., Bilder, B., Bookheimer, S., Cole, M., Mirsky, A., Pliskin, N., . . . Sweeney, J. others. (2004). Official position of the division of clinical neuropsychology (APA division 40) on the role of neuropsychologists in clinical use of fMRI. *Clinical Neuropsychology*, 18, 349–351.
- Busch, R. M., Floden, D. P., Prayson, B., Chapin, J. S., Kim, K. H., Ferguson, L., . . . Najm, I. M. (2016). Estimating risk of word-finding problems in adults undergoing epilepsy surgery. *Neurology*, 87(22), 2363–2369.
- Ganslandt, O., Nimsky, C., Buchfelder, M., & Grummich, P. (2016). fMRI in Neurosurgery. In: Filippi, M, editor. *fMRI Techniques and Protocols* 2nd ed. New York: Springer. pp 801–815.
- Giussani, C., Roux, F. E., Ojemann, J., Sganzerla, E., Pietro, Pirillo, D., & Papagno, C. (2010). Is preoperative functional magnetic resonance imaging reliable for language areas mapping in brain tumor surgery? Review of language functional magnetic resonance imaging and direct cortical stimulation correlation studies. *Neurosurgery*, 66(1), 113–120.
- Goodman, L. A. (1961). Snowball sampling. *The Annals of Mathematical Statistics*, 32(1), 148–170.
- Hamberger, M. J., Williams, A. C., & Schevon, C. A. (2014). Extraoperative neurostimulation mapping: Results from an international survey of epilepsy surgery programs. *Epilepsia*, 55(6), 933–939.
- Hamberger, M. J., & Seidel, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society: Jins*, 9(3), 479–489.
- Jakola, A. S., Unsgård, G., & Solheim, O. (2011). Quality of life in patients with intracranial gliomas: The impact of modern image-guided surgery. *Journal of Neurosurgery*, 114(6), 1622–1630.
- Janecek, J., Swanson, S. J., Sabsevitz, D. S., Hammeke, T. A., Raghavan, M., Rozman, M., & Binder, J. R. (2013). Language lateralization by fMRI and Wada testing in 229 epilepsy patients: Rates and predictors of discordance. *Epilepsia*, 54(2), 314–322.
- Krainik, A., Lehericy, S., Duffau, H., Capelle, L., Chainay, H., Cornu, P., . . . Marsault, C. (2003). Postoperative speech disorder after medial frontal surgery: Role of the supplementary motor area. *Neurology*, 60(4), 587–594.
- Krauss, G. L., Fisher, R., Plate, C., Hart, J., Uematsu, S., Gordon, B., & Lesser, R. P. (1996). Cognitive effects of resecting basal temporal language areas. *Epilepsia*, 37(5), 476–483.
- Medina, L. S., Aguirre, E., Bernal, B., & Altman, N. R. (2004). Functional MR imaging versus Wada test for evaluation of language lateralization: Cost analysis. *Radiology*, 230(1), 49–54. <http://www.ncbi.nlm.nih.gov/pubmed/14695386>.

- Mouthaan, B. E., Rados, M., Barsi, P., Boon, P., Carmichael, D. W., Carrette, E., ... Braun, K. P. J. et al. (2016). Current use of imaging and electromagnetic source localization procedures in epilepsy surgery centers across Europe. *Epilepsia*, 57(5), 770–776.
- Połczyńska, M., Japardi, K., Curtiss, S., Moody, T., Benjamin, C., Cho, A., ... Bookheimer, S. (2017). Improving language mapping in clinical fMRI through assessment of grammar. *NeuroImage: Clinical*, 15, <http://dx.doi.org/10.1016/j.nicl.2017.05.021>.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2), 816–847.
- Roux, F. E., Dufor, O., Giussani, C., Wamain, Y., Draper, L., Longcamp, M., & Démonet, J. F. (2009). The graphemic/motor frontal area Exner's area revisited. *Annals of Neurology*, 66(4), 537–545.
- Sabsevitz, D. S., Swanson, S. J., Hammeke, T. A., Spanaki, M. V., Possing, E. T., Morris, G. L., ... Binder, J. R. (2003). Use of preoperative functional neuroimaging to predict language deficits from epilepsy surgery. *Neurology*, 60(11), 1788–1792.
- Sherman, E. M. S., Wiebe, S., Fay-McClymont, T. B., Tellez-Zenteno, J., Metcalfe, A., Hernandez-Ronquillo, L., ... Jetté, N. (2011). Neuropsychological outcomes after epilepsy surgery: Systematic review and pooled estimates. *Epilepsia*, 52(5), 857–869.
- Swanson, S. J., Binder, J. R., Raghavan, M., & Euler, M. (2015). Functional MRI in the presurgical epilepsy evaluation. In: Barr, W., Morrison, C., editors. *Handbook on the Neuropsychology of Epilepsy*. New York: Springer. pp 169–194.
- Szaflarski, J. P., Gloss, D., Binder, J. R., Gaillard, W. D., Golby, A. J., Holland, S. K., ... Theodore, W. H. et al. (2017). Practice guideline summary: Use of fMRI in the presurgical evaluation of patients with epilepsy. *Neurology*, 88(4), 395–402.
- Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, 162, 60–71.
- Wiebe, S., Blume, W. T., Girvin, J. P., & Eliasziw, M. (2001). A randomized, controlled trial of surgery for temporal-lobe epilepsy. *The New England Journal of Medicine*, 345(5), 311–318.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Benjamin CFA, Li AX, Blumenfeld H, et al. Presurgical language fMRI: Clinical practices and patient outcomes in epilepsy surgical planning. *Hum Brain Mapp.* 2018;00:1–9. <https://doi.org/10.1002/hbm.24039>