**Language Mapping With fMRI**

**Current Standards and Reproducibility**

Shruti Agarwal, PhD,* Haris I. Sair, MD,* Sachin Gujar, MBBS,* and Jay J. Pillai, MD††

**Abstract:** Clinical use of blood oxygen level dependent (BOLD) functional magnetic resonance imaging (fMRI) is a relatively new phenomenon, with only about 3 decades of collective experience. Nevertheless, task-based BOLD fMRI has been widely accepted for presurgical planning, over traditional methods, which are invasive and at times perilous. Many studies have demonstrated the ability of BOLD fMRI to make substantial clinical impact with respect to surgical planning and preoperative risk assessment, especially to localize the eloquent motor and visual areas. Reproducibility and repeatability of language fMRI are important in the assessment of its clinical usefulness. There are national efforts currently underway to standardize language fMRI. The American Society of Functional Neuroradiology (ASFNRR) has recently provided guidelines on fMRI paradigm algorithms for presurgical language assessment for language lateralization and localization. In this review article, we provide a comprehensive overview of current standards of language fMRI mapping and its reproducibility.

**Key Words:** language fMRI, lateralization, presurgical mapping, repeatability, reproducibility.


**Clinical** use of blood oxygen level dependent (BOLD) functional magnetic resonance imaging (fMRI) is a relatively new phenomenon, with only about 3 decades of collective experience. Nevertheless, task-based BOLD fMRI has been widely accepted for presurgical planning in patients with brain tumors and other focal brain lesions such as brain neoplasms, arteriovenous malformations and other vascular malformations, cortical dysplasias, and other epileptogenic lesions. Preoperative fMRI has been very helpful in assisting neurosurgeons in making decisions of whether or not to attempt surgical resection of the lesion and in selection of patients for asleep vs awake craniotomy. The preoperative detection of functional areas around a lesion can guide the intraoperative cortical stimulation (ICS) mapping, thus reducing the total surgical time and planning the safest surgical trajectory. BOLD fMRI has recently been used in patients with trauma, vascular diseases, inflammations, multiple sclerosis, Alzheimer disease, developmental disorders, learning disabilities, and many other conditions, although most of these applications are still considered to be in the research realm.

**BACKGROUND ON BOLD fMRI**

For a long time, it was mistakenly believed that cerebral hemodynamics were solely controlled by the brain’s short-term metabolic requirements. Contrary to this popular belief, in 1986, Fox and Raichle demonstrated using 015-labeled radiotracers positron emission tomography (PET), the independence between brain blood flow and oxygen demand. In that study, they found that regional cerebral blood flow (CBF) increases during cerebral activation; however, the cerebral metabolic rate of oxygen consumption (CMRO2) does not increase proportionally. Consequently, more fresh (oxygenated) blood is supplied to that region of functioning brain than is required for its immediate metabolic needs. Effectively, this results in decreased relative concentration of deoxyhemoglobin compared with oxyhemoglobin. In 1992, Ogawa et al exploited different magnetic properties of oxygenated and deoxygenated blood to detect changes in regional CBF. This property, popularly known as BOLD (blood oxygen level dependent), provides contrast for the most commonly performed clinical functional MRI (fMRI). A decrease in the relative concentration of deoxyhemoglobin in active cortex reduces the T2/T2* shortening effects of deoxygenated globin with resultant net increase in the BOLD signal in activated areas.

**Imaging Strategies and Tradeoffs**

fMRI necessitates image acquisition methods, which are sensitive to changes in T2 and T2*; have sufficient spatial resolution to cover the entire brain, and have sufficient temporal resolution to detect changes in BOLD signal associated with specific tasks. Acquisition strategies are largely based on a trade-off between signal-to-noise ratio (SNR), spatial resolution, temporal resolution, and motion artifacts.

T2*-weighted gradient echo (GRE) sequences are the most commonly used BOLD sequences at fields of 3T or below. The high sampling rates necessary for fMRI are achieved utilizing “single-shot” sequences, which make use of fast gradient switching technologies allowing for acquisition of the data for an entire slice in one readout window after one excitation. Echo-planar imaging (EPI) and spiral imaging are the most widely used single-shot imaging sequences. Recently developed “parallel imaging (PI)” techniques reduce the readout time by acquiring data from multiple coils. For BOLD acquisitions at 1.5T and 3T, PI is not preferable because of high SNR loss at even the lowest acceleration factor. A recent development is multiband parallel imaging technique. In multiband EPI, multiple 2D slices are simultaneously excited (multiband excitation), and individual slices are reconstructed. Multiband is not hindered by the SNR reduction associated with reduced sampling in traditional parallel imaging, however, interslice leakage artifact occurs due to incomplete separation of slice signals over time. Artifacts such as noise, motion, aliasing, chemical shift, Gibbs, susceptibility, RF-interference, and so on, which are usually noticed in conventional MR imaging, also occur in PI with slightly different appearances because of the complex PI reconstruction process.
Language fMRI Paradigm Designs

 Appropriately designing the stimulus paradigm is as important as choosing imaging parameters. A valid stimulus paradigm should ideally localize only the areas pertaining to the considered function of interest. The location and extent of activated areas should not vary significantly in different trials during the same or different sessions, that is, the paradigm should be reproducible.

 During an fMRI scan session, the brain is repeatedly imaged while the subject is resting or performing the control task and while the subject is presented with a stimulus or is performing an active task. The paradigm timing depends on tradeoff between the level of attention of the subject during the task, avoiding motion and acquiring enough data to provide statistically significant mapping results. For presurgical planning, block designs are most commonly used that typically last between 2 and 4 minutes with epochs of 15 and 30 seconds. In the block design, there are regular epochs of stimulus and rest, labeled “on” (active) and “off” (control). The resulting difference between the signal acquired during the “on” period and the “off” period is the activation of interest. The “off” condition is not always the absence of any activity, but may also consist of another task that differs from the primary task performed during the “on” period and does not activate the functional network of interest.

 A large variety of paradigms have been reported in the literature for mapping the language network with fMRI. These paradigms can be divided into 3 main categories: expressive, receptive, and semantic. Expressive paradigms are designed to elicit language productive areas, particularly Broca area (BA) in the dominant hemisphere. Frontal brain regions, including precentral gyrus, superior frontal gyrus, middle frontal gyrus, inferior frontal gyrus (including pars opercularis, pars triangularis and pars orbitalis), and supplementary motor area, are expected to see activation during expressive language task performance. Verbal fluency tasks are very frequently used for presurgical expressive language brain mapping. Receptive paradigms are used to elicit areas related to comprehension of language, particularly Wernicke area (WA) in the dominant hemisphere. Temporal brain regions, including fusiform gyrus, temporal lobes (superior temporal gyrus, middle temporal gyrus, inferior temporal gyrus, temporal poles, opercularis, and parietal regions including supramarginal gyrus and angular gyrus) are expected to see activation during receptive language task performance. Receptive paradigms include mainly language reading or listening comprehension tasks. Semantic paradigms are more general purpose for language mapping than expressive and receptive paradigms. They are designed to be useful for language lateralization and localization of inferior frontal, posterior temporal, and inferior parietal speech areas in the dominant cerebral hemisphere. Figure 1 displays general regions of the brain where language related functional areas may be located in the Frontal, Temporal and Parietal lobes.

 Benjamin et al categorized 6 critical regions for presurgical language mapping: BA within the inferior frontal gyrus (pars opercularis and pars triangularis); Exner area (“graphemic motor frontal area”) in the posterior region of the middle frontal gyrus, critically involved in transforming phonological representations of words into motor commands for producing their written forms; Supplementary speech area (“pre-supplementary motor area,” speech component; pre-SMA) in the posterior superior and medial frontal cortex, critically engaged in initiating and sequencing motor movements for speech; Angular gyrus, particularly involved in reading and transitioning between written and spoken forms of language; WA in temporal lobe gyri (specifically, superior and middle temporal gyri) with multiple varied definitions in use however largely representing receptive language function. Basal Temporal language area, critical to language in the basal temporal region, a region thought to link semantics with names.

 The American Society of Functional Neuroradiology (ASFNR) hosts presentation files and parameters of the most common paradigms that are freely available for use on their website (https://www.asfnr.org/paradigms/). Additional unique paradigms utilized by institutions across the country are also described in a document hosted by the ASFNR website (https://www.asfnr.org/wp-content/uploads/ASFNR-BOLD-Paradigms.pdf): (1) Auditory Responsive Naming Task (Aural/Visual) in which the patient is asked to read a sentence and push a button corresponding to the correct choice (visual presentation) and silently generate a word that fits the description (aural presentation). The baseline is a resting baseline for the aural presentation and a nonsense baseline for the visual presentation. The task has been used to localize WA and may be a more sensitive measure than the typical object naming during electrocorticography. (2) Semantic Decision Task (Visual) in which word pairs (e.g., “fruit” and “apple”) are presented and the patient is asked to push a button when pairings are correct. The baseline is a visual presentation of pair of lines and patient is asked to push the button when the 2 sets of lines are identical. This semantic decision paradigm has been used for language lateralization and localization of inferior frontal, posterior temporal, and inferior parietal speech areas in the dominant cerebral hemisphere. (3) Text Reading versus Non-linguistic Symbols in which the patient is instructed to concentrate on reading and comprehending text during the active block and merely attend to the symbols during the control block. This task has been used to localize language comprehension areas in the dominant hemisphere region of posterior superior temporal gyrus/parietal angular gyrus (WA). (4) The Visual Language Comprehension in which sentence-question pairs are presented and patient is asked to push a 2-button finger switch held in the dominant hand for a YES/NO answer. (5) Silent Verb Generation Task, in which the patient is asked to generate verbs following presentation of a noun; baseline is visual fixation. (6) Word Listening Task for which the patient hears words (nouns), during which the patient is supposed to repeat the words silently.

 Table 1 provides a comprehensive list of fMRI paradigms, which are currently used to activate language brain areas at the Johns Hopkins Hospital (JHH). Figure 2 provides samples of control block and active block of language fMRI paradigm designs which are currently used at the JHH.

 Monitoring and Processing

 A comprehensive pre-scan training session outside the MRI scanner should be performed to ensure full patient understanding of task instructions and confirmation of each patient’s ability to adequately perform the tasks. Each patient’s task performance should be monitored during the scan via real-time fMRI for assessment of activation, bulk head motion, and physiologic noise.

 Once the data are obtained, raw data are corrected for slice timing due to interleaved EPI scans, followed by realignment to correct for motion of the subject during the functional scans utilizing rigid body translation and rotation transformation. Thereafter, as a pre-processing step, spatial smoothing of the images with a Gaussian kernel should be performed on raw data.

 For generating activation maps from the pre-processed data, regression analysis is performed by fitting the observed fMRI signal time course of each voxel to a theoretical expected time course generated by a standard hemodynamic response function (HRF). The theoretical HRF is a mathematical expression of the mechanism behind BOLD fMRI. Due to an increase in neuronal activity, fMRI signal initially decreases because the active neurons consume oxygen.

 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.
thereby increasing the relative level of deoxyhemoglobin in the blood.\textsuperscript{52,53} This decrease, however, is tiny and is not always found.\textsuperscript{54} After an initial undershoot, fMRI signal reaches its maximum in approximately 6 seconds.\textsuperscript{52–55} This increase is due to oversupply of oxygen-rich blood, which causes a large decrease in the relative level of deoxyhemoglobin, which in turn results in increased BOLD signal. Finally, the level of deoxyhemoglobin slowly returns to normal after an initial undershoot in approximately 24 seconds\textsuperscript{52} and the fMRI signal decays until it has reached its original baseline level. This HRF is then convoluted with a condition box-car function having a value of 1 corresponding to active blocks of the fMRI paradigm and 0 for control blocks. T-contrast maps are then obtained to analyze the contrast between the language activation and baseline conditions. Figure 3 display suprathreshold language activation maps overlaid on anatomical T1 MPRAGE obtained from SC and SWG tasks performed in the same scan session by a right-handed patient with left medial temporal lobe lesion.

**Lateralization and Localization**

Language function is known to display unilateral hemispheric dominance.\textsuperscript{58} Presurgical fMRI mapping for the assessment of language areas in patients with focal brain lesions has been used to determine hemispheric language dominance, thus replacing traditional invasive approach of intracarotid amobarbital procedure\textsuperscript{57} also known as the “Wada Test,” which involves arterial catheterization. Many studies have clinically validated hemispheric language lateralization obtained from language fMRI mapping through comparison with hemispheric dominance obtained from Wada testing; concordance between lateralization indices (LIs) from these 2 techniques has generally been in the range of 90% to 100%.\textsuperscript{58–63}

In a recent review article, Bradshaw et al\textsuperscript{64} evaluated various methods used in fMRI studies since the year 2000 for quantifying laterality. Laterality Index\textsuperscript{65} as a means to evaluate relative extents of activation in the left and right hemispheres has been used widely to assess language hemispheric dominance.\textsuperscript{42,66–68} LI is computed as

**FIGURE 1.** General regions of the brain where language related functional areas may be located in the Frontal, Temporal, and Parietal lobes.
TABLE 1. A Comprehensive List of fMRI Paradigms That Are Currently Used to Activate Language Brain Areas at the Johns Hopkins Hospital (JHH)

<table>
<thead>
<tr>
<th>Category</th>
<th>Paradigm Nomenclature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptive/Semantic</td>
<td>Sentence Completion (SC)</td>
<td>Control Block. Visual fixation on consecutive samples of scrambled letters arranged to resemble words in a sentence.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Active Block. Covert reading of consecutive real sentences with the last word missing and covert generation of a word to complete each sentence.</td>
</tr>
<tr>
<td>Expressive</td>
<td>Silent Word Generation (SWG)</td>
<td>Control Block. Visual fixation on consecutive samples of nonsense drawings.</td>
</tr>
<tr>
<td>Expressive</td>
<td>Rhyming (R)</td>
<td>Control Block. Visual fixation on consecutive samples of nonsense drawings.</td>
</tr>
<tr>
<td>Expressive</td>
<td>Object Naming (ON)</td>
<td>Control Block. Visual presentation of consecutive pairs of words. Keypad button press is required if the words rhyme.</td>
</tr>
<tr>
<td>Receptive</td>
<td>Passive Story Listening (PL)</td>
<td>Active Block. Listening to garbled (backward) speech without button presses.</td>
</tr>
<tr>
<td>Receptive</td>
<td>Listening Comprehension (LC)</td>
<td>Active Block. Aural presentation of sentences. Keypad button press is required if sentence represents a true statement.</td>
</tr>
<tr>
<td>Receptive</td>
<td>Reading Comprehension (RC)</td>
<td>Active Block. Visual fixation on consecutive samples of nonsense symbols.</td>
</tr>
</tbody>
</table>

(\(L-R)/(L+R)\), where \(L\) refers to activations in the left hemisphere and \(R\) refers to activations in the right hemisphere. This formula yields values between \(-1\) and \(+1\), which are positive for left hemispheric dominance \((LI > 0.2)\), negative \((LI < -0.2)\) for right hemispheric dominance, and bilateral when \((-0.2 < LI < 0.2)\). \(LI\) can be determined by comparing all positively task correlated voxels between the left and right hemisphere, referred as “Threshold-Independent LI”. When \(LI\) is computed only between supra-threshold voxels in language activation map, it is referred to as “Threshold-dependent LI”. In their review article, Bradshaw et al suggested threshold-independent LI methods for assessing heterogeneity of language laterality across multiple regions of interests and various language tasks. Pillai et al compared threshold-dependent versus threshold-independent techniques in a large series of brain tumor patients undergoing presurgical language mapping. Their findings suggested that expressive tasks provided the best hemispheric language lateralization based on concordant threshold-dependent and threshold-independent analyses and receptive tasks were less effective for language lateralization.

Language presurgical fMRI mapping not only localizes critical areas for language processing but also identifies other less critical areas of the language network. Various studies have been performed to validate language fMRI by comparing the locations of activation found during fMRI with the language functional mapping obtained from Ojemann Stimulator during awake craniotomy. As not all patients are candidates for awake ICS mapping, preoperative fMRI often becomes essential for surgical planning. Many studies have also demonstrated the ability of fMRI to make a substantial clinical impact, particularly with respect to surgical planning and preoperative risk assessment.

Reproducibility and Repeatability

Reproducibility and repeatability of language fMRI is an important measurement in the assessment of its clinical usefulness. fMRI tasks are often repeated to confirm both across-subject reproducibility and within-subject repeatability of language laterализation and localization. Many studies have evaluated across-subject reproducibility of language activations to assess the reliability of fMRI paradigms by calculating the correlation coefficients as similarity metric between the activations obtained from the same tasks performed in 2 or multiple sessions by same participants. Another prominent measure of similarity metric is the Dice coefficient. Dice coefficient and other reliability metrics are used to quantify test-retest variability of activation clusters across sessions and across subjects. Several reliability studies of language mapping paradigms inform about the consistency of activation between subjects.

There are national efforts currently underway to standardize language fMRI. In a recent white paper, the ASFNR provided guidelines on fMRI paradigm for presurgical language assessment. In the ASFNR guidelines, the sentence completion and silent word generation tasks were recommended as primary tasks to be used for effective language cortical localization and hemispheric lateralization/dominance determination. The ASFNR conducted a poll among its entire membership on the number of language-assessment fMRIs per month that were performed at their various institutions to determine the most commonly used language paradigms. Results of the ASFNR poll produced 3 tiers of commonly used paradigms across the nation. Silent Word Generation and Sentence Completion stood out as the most frequently employed tasks followed by Verb Generation, Object Naming, Rhyming, and Reading Comprehension.
in the second-tier group. On the basis of the poll results and literature review, the ASFNR task force recommended language paradigm algorithms for both adults and pediatric subjects. As language can be represented across phonologic, semantic, receptive, expressive, and so on domains and one task cannot simultaneously activate all of these aspects, multiple tasks are recommended to provide a more sensitive and specific map of language function that will aid in surgical planning. The default adult algorithm includes SC, SWG, and Rhyming. The radiologist may choose to drop Rhyming and repeat the SC (most appropriate task to repeat) or SWG task. Radiologist may choose either the Object Naming (which is an expressive task) or Passive Story Listening (which is a receptive

<table>
<thead>
<tr>
<th>Control Block</th>
<th>Active Block</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentence Completion (SC)</strong></td>
<td>Jhz RknrHp pgllp koup kg ___</td>
</tr>
<tr>
<td><strong>Silent Word Generation (SWG)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Rhyming (R)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Object Naming (ON)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Reading Comprehension (RC)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Noun Verb Association (NVA)</strong></td>
<td></td>
</tr>
</tbody>
</table>

FIGURE 2. Samples of control block and active block of language fMRI paradigm designs.
task) task when patient has difficulty performing SC/SWG tasks. The default pediatric algorithm includes SC, Rhyming, and Antonym Generation (AG).

In a recent paper, Agarwal et al\textsuperscript{96} demonstrated a very high degree of repeatability at the single-subject level within single scan sessions of language mapping with sentence completion and silent word generation tasks in a large cohort of patients (37 right-handed patients for SC task and 78 right-handed patients for SWG task) undergoing presurgical fMRI. They investigated the laterality index and center of mass (COM) of holohemispheric (ie, individual left and right hemispheres excluding cerebellum) and region-specific language areas (BA and WA, as well as larger more inclusive expressive and receptive language regions) for quantification of within-subject variability of activated language clusters. In that paper, Within-Subject Lateralization Index Variability (LI\textsubscript{VAR}) was calculated across the right-handed patient population to quantify the within-subject variability of holohemispheric and regional language lateralization. LI\textsubscript{VAR} was defined as (LI1-LI2)/(LI1+LI2) where LI1 and LI2 are the LIs of two consecutive runs within the same scan session. High intrasubject language lateralization repeatability was found using both threshold-dependent and threshold-independent approaches for both holohemispheric and regional language lateralization. Except for a few outliers, within-subject variability of LI from run 1 to run 2 in holohemispheric as well as in local regions for each subject was below 1.0. Variability in LI was lower for the SWG task in comparison to the SC task, which confirmed that the SWG task is a better determinent of language lateralization.

In the same paper, Agarwal et al\textsuperscript{96} obtained the COM of regional language activation areas to localize the centers of activations of the SWG and SC language tasks. Within-subject Center of Mass Variability (COM\textsubscript{VAR}) was calculated to quantify the variability in COM location from run 1 (ie, COM1) to run 2 (ie, COM2) of each language task within the same scan session using the Euclidean distance between COM1 and COM2. The displacement of COM coordinates from run 1 to run 2 was found to be within 5 mm. The repeatability in localization centers is demonstrated both when centers of mass of activation are considered within key eloquent regions of the brain, such as BA and WA, as well as in larger more inclusive expressive and receptive language regions. Note that COM of regional language activation areas is usually obtained to localize the centers of activations of each language task; however, it does not necessarily mean that it is representative of the true spatial coordinates of neural activity.\textsuperscript{97}

CONCLUSION

In this review, current standards and reproducibility of presurgical language mapping with task-based fMRI were discussed. Details of methods of image acquisition, monitoring, and analysis are provided. There are national efforts currently underway to standardize language fMRI for clinical applications, and more work needs to be done to accomplish this fully. Recently, the ASFNR provided guidelines on fMRI paradigm algorithms for presurgical language assessment for language lateralization and localization. In the near future, with the wide range of ongoing research, new windows into brain language function and connectivity will open that will result in improved clinical care.

REFERENCES


