WILEY

RESEARCH ARTICLE

Repeatability of language fMRI lateralization and localization metrics in brain tumor patients

Shruti Agarwal[1] | Jun Hua[2,3] | Haris I. Sair[1] | Sachin Gujar[1] | Chetan Bettegowda[4] |
Hanzhang Lu[2,3] | Jay J. Pillai[1,4]

[1]Division of Neuroradiology, Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, Maryland

[2]Division of MR Research, Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, Maryland

[3]F. M. Kirby Research Center For Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, Maryland

[4]Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, Maryland

Correspondence
Jay J. Pillai, MD, Division of Neuroradiology, Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins Hospital, Phipps B-100, 1800 Orleans Street, Baltimore, MD 21287.
Email: jpillai1@jhmi.edu

Abstract

To assess the within-subject intra-scan session repeatability of language functional MRI (fMRI) activation maps in patients with brain tumors who were undergoing presurgical fMRI as part of their preoperative clinical workup. Sentence completion (SC) and silent word generation (SWG) tasks were used for language localization and hemispheric lateralization for identifying the primary language cortex. Within-subject repeatability for each of these paradigms was assessed in right-handed patients—37 for SC and 78 for SWG. Repeatability of activation maps between consecutive runs of the same task within the same scan session was evaluated by comparing lateralization indexes in holohemispheric and regional language areas. Displacement of center of activation between consecutive runs was also used to assess the repeatability of activation maps. Holohemispheric and regional language lateralization results demonstrated high intra-subject intra-scan repeatability when lateralization indices were calculated using threshold-dependent and threshold-independent approaches. The high repeatability is demonstrated both when centers of mass of activation are considered within key eloquent regions of the brain, such as Broca's area and Wernicke's area, as well as in larger more inclusive expressive and receptive language regions. We examined two well-known and widely accepted language tasks that are known to activate eloquent language cortex. We have demonstrated very high degree of repeatability at a single-subject level within single scan sessions of language mapping in a large cohort of brain tumor patients undergoing presurgical fMRI across several years at our institution.

KEYWORDS

language fMRI, lateralization, presurgical mapping, repeatability, reproducibility

1 | INTRODUCTION

Quantitative claims that are central to blood oxygen level dependent (BOLD) functional MRI (fMRI) mapping, such as localizing centers of activations, measuring the spatial extent of active cortex, and the overall spatial distribution and relative contribution of different activated regions to the entire language network (i.e., hemispheric dominance) have been critical to mapping language areas for presurgical assessment in patients with focal brain lesions. Especially when the brain lesions infiltrate or displace eloquent language areas, anatomic landmarks alone cannot accurately delineate language areas. Brain plasticity which may occur to compensate for possible language deficits can also alter language lateralization, which makes proper assessment of hemispheric language dominance important. Language function is known to typically display unilateral hemispheric dominance (Broca, 1861). The intracarotid amobarbital procedure (Wada & Rasmussen, 1960) and electrocortical stimulation (Ojemann, 1979) have largely being used to determine hemispheric language dominance

and localize language functional cortex, respectively. 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 the Wada test and Ojemann stimulator during awake craniotomy (Binder et al., 1996; FitzGerald et al., 1997; Gartus, Foki, Geissler, & Beisteiner, 2009; Rutten, van Rijen, van Veelen, & Ramsey, 1999). In a recent white paper, the American Society of Functional Neuroradiology (ASFNR) provided guidelines on fMRI paradigm algorithms for presurgical language assessment (Black et al., 2017). In the ASFNR guidelines, the sentence completion (SC) and silent word generation (SWG) tasks were recommended as primary tasks to be used for effective language cortical localization and hemispheric lateralization/dominance determination. SC is a semantic language paradigm that is effective in activating the superior and middle temporal gyri including Wernicke's area (WA; Zaca, Nickerson, Deib, & Pillai, 2012). SC can also activate Broca's area (BA) in the dominant hemisphere because performance of this task requires both receptive and expressive language processing. SWG task activates mainly frontal lobe language and cognitive support areas but are less consistent activators of temporal language regions (Pillai & Zaca, 2011; Zaca et al., 2012; Zaca, Jarso, & Pillai, 2013). These tasks are often repeated to confirm the reproducibility of activations (Carp, 2013; Fernández et al., 2003; Harrington, Buonocore, & Tomaszewski Farias, 2006; Maïza et al., 2011; Morrison et al., 2016; Poline, Strother, Dehaene-Lambertz, Egan, & Lancaster, 2006; Rutten, Ramsey, van Rijen, & van Veelen, 2002; Voyvodic, 2012). Previous validation studies evaluated across-subject reproducibility; however, within-subject repeatability in the same scan session for reproducibility of language lateralization and localization has not yet been comprehensively validated. Within-subject repeatability of language fMRI is an important measurement in the assessment of its clinical usefulness. In general, reliability metrics are used to quantify test–retest variability of activation clusters across sessions and across subjects (Gorgolewski, Storkey, Bastin, Whittle, & Pernet, 2013; Otzenberger, Gounot, Marrer, Namer, & Metz-Lutz, 2005; Raemaekers, Du Plessis, Ramsey, Weusten, & Vink, 2012). Previous studies reported varied results due to the varied statistical methods used to quantify test–retest reliability (Chen & Small, 2007; Fesl et al., 2010; Morrison et al., 2016). The majority of studies used correlation coefficients as a prominent measure of reliability which informs about the consistency of activation between subjects (Caceres, Hall, Zelaya, Williams, & Mehta, 2009; Maldjian, Laurienti, Driskill, & Burdette, 2002; Stevens, Clarke, Stroink, Beyea, & D'Arcy, 2015). In general, within-subject variance (Zandbelt et al., 2008) is used to assess the repeatability of observations across repeated measurements.

In this study, we sought to explore repeatability of language fMRI activation in patients with brain tumors who underwent presurgical fMRI as part of their preoperative clinical workup. The objective of this study was to assess the repeatability of measurements of laterality index and center of mass across consecutive runs of the same two language tasks in brain tumor patients in a single scan session via within-subject variation metrics. To the best of our knowledge, no prior investigation has evaluated language fMRI "within-subject intra-scan session" repeatability in a large clinically representative cohort of brain tumor patients. Furthermore, no such studies have specifically evaluated these two most important clinical language tasks that have been recommended by the ASFNR white paper.

2 | MATERIALS AND METHODS

2.1 | Imaging protocol

Scanning was performed using our standard clinical sequences for fMRI studies on a 3.0 Tesla (T) Siemens Trio MRI system (Siemens Medical Solutions, Erlangen, Germany) equipped with a 12-channel head matrix coil. Imaging protocol included a three-dimensional (3D) T1-weighted magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence (TR [repetition time] = 2,300 ms, TI [inversion time] = 900 ms, TE (echo time) = 3.5 ms, flip angle = 9°, field of view = 24 cm, acquisition matrix = $256 \times 256 \times 176$, slice thickness = 1 mm) as well as a two-dimensional (2D) T2 fluid attenuated inversion recovery (FLAIR) sequence (TR = 9,310 ms, TI = 2,500 ms, TE = 116 ms, flip angle = 141°, field of view = 17.2 cm × 23 cm, acquisition matrix = $320 \times 240 \times 50$, slice thickness = 3 mm) for structural imaging and multiple 2D gradient echo-echo planar imaging (GE-EPI) T2*-weighted BOLD sequences for task-based functional imaging (TR = 2000 ms, TE = 30 ms, flip angle = 90°, field of view = 24 cm, acquisition matrix = $64 \times 64 \times 33$, slice thickness = 4 mm, slice gap = 1 mm, interleaved acquisition).

2.2 | fMRI paradigms

Prism Acquire (Prism Clinical Imaging, Elm Grove, WI) was used for fMRI paradigms. Different combinations of clinical fMRI paradigms were performed during the scan session for comprehensive clinical fMRI presurgical mapping as per our institutional clinical protocol based on individual patient needs for lesion characterization and for effective clinical presurgical mapping. For the purpose of our current study, only two language tasks—SC and SWG—were evaluated.

The SC task was used to map both expressive and receptive language areas, and the SWG task was used to map expressive language areas. Each of these covert block design paradigms involved alternating control and active blocks lasting 20 s each for a total task duration of 4 min (Zaca et al., 2012, 2013).

In SC, patients were asked to scan through five consecutive samples of scrambled letters arranged to resemble words in a sentence during the control block. During the active block of the SC paradigm, patients were asked to silently read five consecutive real sentences with the last word missing and silently generate a word to complete each sentence.

In the SWG task, patients were asked to visually fixate on two consecutive nonsense drawings, each for 10 s during control block. During the active block of the SWG paradigm, patients were asked to do covert generation of words for two consecutively presented letters, each for 10 s.

A comprehensive prescan training session outside the MRI scanner ensured full patient understanding of task instructions and confirmed each patient's ability to adequately perform the tasks. Each patient's

task performance was monitored during the scan via real-time fMRI for assessment of activation, bulk head motion, and physiologic noise.

2.3 | Patient data

From an overall study pool of 170 patients undergoing clinical presurgical fMRI language mapping from year 2011 to year 2015, a total of 60 cases (37 right-handed and 23 left-handed/ambidextrous) out of 153 patients who performed the SC task and 98 cases (78 right-handed and 20 left-handed/ambidextrous) out of 149 patients who performed the SWG task include two or more consecutive runs of the task within the same scan session with qualifying quality control (QC) metrics based on head motion parameters (Table 1). The maximum spatial displacement from the volume taken as reference during motion correction calculation was <2.0 mm, and <2° of rotation (less than one voxel size in each direction) was noted in all cases. To avoid the additional confound of handedness differences, we have limited our investigation to only right-handed patients (n = 37 for SC and n = 78 for SWG).

2.4 | fMRI data processing

SPM12 (Wellcome Trust Center for Neuroimaging, London, UK: http://www.fil.ion.ucl.ac.uk/spm/) software implemented in MATLAB R2014b (The MathWorks, Natick, MA) was used for processing of language fMRI data.

Preprocessing steps are as follows: (a) slice timing correction of the interleaved EPI scans, (b) realignment to correct for motion of the subject during the functional scans utilizing rigid body translation and rotation transformation, (c) normalization onto the Montreal Neurologic Institute (MNI) atlas at 2 mm voxel resolution based on the first EPI scan with the default values for nonlinear corrections utilizing the Sinc interpolation algorithm, (d) spatially smoothing of the normalized images with a Gaussian kernel, using a full width at half-maximum (FWHM) of 6 mm.

Regression analysis was then performed by fitting the observed fMRI time course of each voxel to a theoretical expected time course generated by convoluting condition box-car time course with a standard hemodynamic response function available in SPM software (gamma function). A condition box-car time course was defined by setting values to 1 at time points at which the modeled condition is defined (active) and 0 at all other time points. T-contrast maps were obtained to analyze the contrast between the language activation and baseline conditions.

Z-score maps for both SC and SWG language tasks were obtained from T-contrast maps. These Z-score maps were thresholded using an automatic internal normalization method, known as activation mapping as a percentage of local excitation (AMPLE) (Voyvodic, 2012). In each case, the maximum Z-score in the region of interest (ROI) was determined, and the Z-score threshold for all voxels was set to 50% of this local maximum (Figure 1).

In our methods, we define "activation" as either "threshold-dependent" or "threshold-independent" based on whether or not we apply a 50% AMPLE normalization threshold. In both cases, we evaluate both amplitude and spatial extent of task-related activation and we consider the Z-score-weighted distribution of all positively correlated voxels in each hemispheric ROI.

In the subsequent text of this article, the term "Threshold-Independent" (ThI) refers to un-thresholded language activation Z-score maps and "threshold-dependent" (ThD) refers to Z-score maps that display suprathreshold activation following 50% AMPLE normalization.

2.5 | ROI selection

An automated anatomical labeling (AAL) template was used for the automated atlas-based parcellation of regions of interest (ROI) (Smith, 2002; Tzourio-Mazoyer et al., 2002). For holohemispheric ROIs (Holo), separate masks of left and right hemispheres excluding cerebellum were created. Each expressive language ROI (Exp) contained frontal 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. Each receptive language ROI (Rec) contained temporal and parietal regions including parietal lobe (superior and inferior parietal lobules), including supramarginal gyrus, angular gyrus, temporal lobes (superior temporal gyrus, middle temporal gyrus, inferior temporal gyrus, temporal poles-superior, poles-mid), and fusiform gyrus. The Broca's area (BA) ROI contained only the inferior frontal gyrus (pars opercularis and pars triangularis) and the Wernicke's area (WA) ROI contained only temporal lobe gyri (specifically, superior, and middle temporal gyri; Figure 2).

2.6 | Lateralization index

The overall distribution and relative contributions of activated language regions (i.e., hemispheric dominance) were assessed using a

TABLE 1 Number of patients investigated in the current study

Language task	SC		SWG	
Number of patients who performed one or two runs of the task	N_{SC} = 153		N_{SWG} = 149	
Number of patients who performed two runs of the task	N_{2SC} = 63		N_{2SWG} = 99	
Number of patients who met quality control (QC) standards (net head displacement <2 mm for both runs)	n_{SC} = 60		n_{SWG} = 98	
	Right-handed	Left-handed/ambidextrous	Right-handed	Left-handed/ambidextrous
	$^r n_{SC}$ = 37	$^l n_{SC}$ = 23	$^r n_{SWG}$ = 78	$^l n_{SWG}$ = 20

N = 170 (108 males and 62 females, average age 42 years ranging from 15 to 81 years). N = number of patients who underwent clinical presurgical language fMRI mapping from years 2011 to 2015 at our institution.

lateralization index (LI). LI is computed as (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 (Gaillard et al., 2002).

The LI was computed using two different approaches. First, a "Threshold-Independent LI" (ThI LI) was determined by comparing the integrated Z-score weighted distributions of all positively task-correlated voxels between the left and right hemispheric homologous ROIs. "Threshold-dependent LI" (ThD LI) was computed using the integrated weighted distribution of 50% AMPLE suprathreshold voxels in language activation Z-score maps (Pillai & Zaca, 2011).

ThI LI and ThD LI values from both language (i.e., SC and SWG) activation maps for right-handed patients are provided in Figure 3.

2.7 | Lateralization index variability (LI$_{VAR}$)

Within-subject variability of holohemispheric and regional language lateralization from run 1 to run 2 for each task within a scan session was calculated as follows:

$$LI_{VAR} = \left| \frac{LI1 - LI2}{LI1 + LI2} \right|$$

where LI1 and LI2 are referred to lateralization index obtained from run 1 and run 2 of a particular language task, respectively. LI$_{VAR}$ > 1 implies that the laterality from run 1 to run 2 changes across

hemispheres. LI$_{VAR}$ ≤ 1 implies that the laterality from run 1 to run 2 remains within the same hemisphere.

2.8 | Center of mass

For localizing centers of activations, the center of mass (COM) of regional language activation areas in each hemisphere was determined as follows:

$$X_c = \frac{x_1 m_1 + x_2 m_2 + \cdots x_n m_n}{m_1 + m_2 + \cdots m_n}; \quad Y_c = \frac{y_1 m_1 + y_2 m_2 + \cdots y_n m_n}{m_1 + m_2 + \cdots m_n}; \quad Z_c = \frac{z_1 m_1 + z_2 m_2 + \cdots z_n m_n}{m_1 + m_2 + \cdots m_n}.$$

where X_c, Y_c, Z_c are the coordinates of the COM, x_i, y_i, z_i are the coordinates of the ith voxel in the ROI and m_i is the Z-score of the ith voxel. The COM for each ROI in the left and right hemisphere was determined separately. ThI COM is the center of mass of all activated (i.e., positively task-correlated) voxels within the ROI whereas ThD COM is the center of mass of 50% AMPLE suprathreshold activated voxels within the ROI.

The ThD COM for each ROI from both language (i.e., SC and SWG) activation maps for right-handed patients are displayed in Figure 4.

2.9 | Center of mass variability (COM$_{VAR}$)

Within-subject variability of centers of localization from run 1 to run 2 in regional language activation ROIs is calculated as follows:

$$COM_{VAR} = ||COM1 - COM2||$$

where COM1 and COM2 are referred to center of mass obtained from run 1 and run 2 of a particular language task within the same
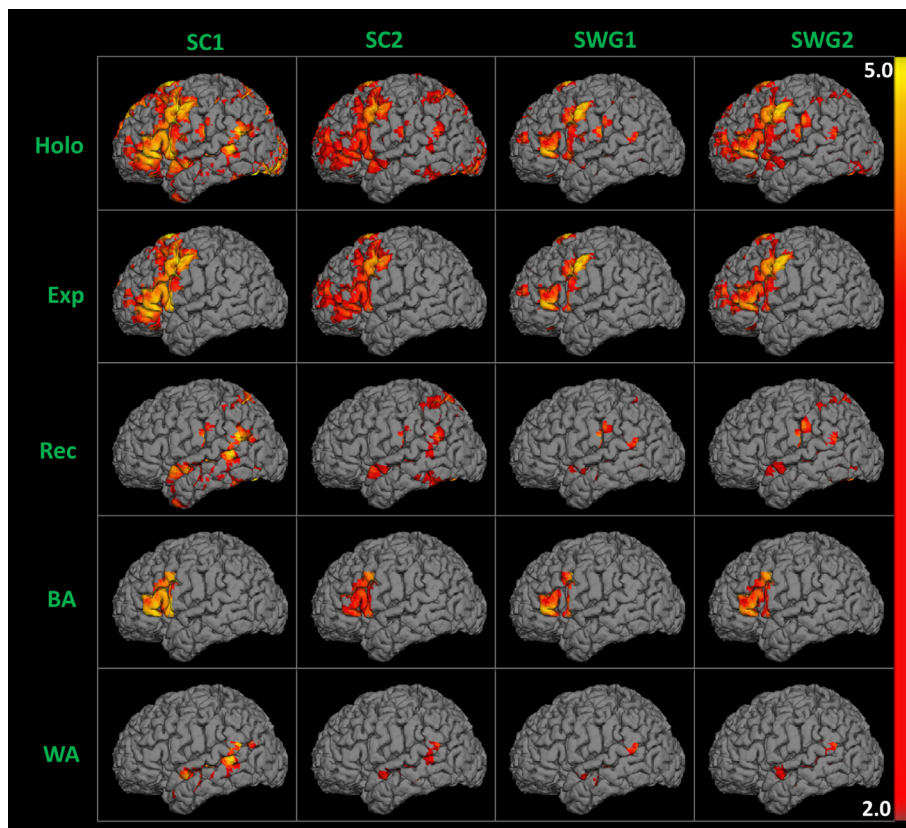
FIGURE 1 Suprathreshold 50% AMPLE (activation mapping as a percentage of local excitation) language activation maps overlaid on the 3D surface of anatomical T1 MPRAGE obtained from two consecutive runs of sentence completion (SC) and silent word generation (SWG) fMRI tasks performed in the same scan session by a right-handed patient with a right frontal mass [Color figure can be viewed at wileyonlinelibrary.com]

scan session, respectively. COM$_{VAR}$ is referred to as Euclidean distance between COM1 and COM2 and is defined as the following:

$$COM_{VAR} = \sqrt{(X_{c1} - X_{c2})^2 + (Y_{c1} - Y_{c2})^2 + (Z_{c1} - Z_{c2})^2}$$

where (X_{c1}, Y_{c1}, Z_{c1}) and (X_{c2}, Y_{c2}, Z_{c2}) are the coordinates of COM1 and COM2, respectively.

3 | RESULTS

Table 2 includes correlation coefficient (r) between LI values from within-session runs 1 and 2 across right-handed subjects. The highest correlation in LI values between two consecutive runs of the SC task was found in the receptive ROI (r = 0.69) whereas the highest correlation in LI values between two consecutive runs for the SWG task was found in Broca's area (r = 0.86). Median of LI values from both runs of each language task across the right-handed subjects was also listed.

Figure 5 displays within-subject LI$_{VAR}$ which indicates the variability in LI from run 1 (LI1) to run 2 (LI2) in holohemispheric and local regions of interest. The box plots of LI$_{VAR}$ obtained across the right-handed subject population is displayed. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Those patients with LI$_{VAR}$ beyond 75th percentile are plotted individually using the "+" symbol. Note that cases for which LI$_{VAR}$ > 1 were those where laterality changed from one hemisphere dominance to the opposite hemisphere dominance from run 1 to run 2, that is, LI1 and LI2 were of opposite signs. The median values of LI$_{VAR}$ (i.e., variation in laterality between run 1 and run 2) across subjects within holohemispheric and each ROI show that the absolute value of LI in the second run was less than double that of the first run(i.e., [LI1/LI2 < 2] or [LI2/LI1 < 2]) as LI$_{VAR}$ is less than 0.33 (i.e., LI$_{VAR}$ < 0.33).

Figure 6 displays within-subject COM$_{VAR}$ which indicates the variability in COM location from run 1 (i.e., COM1) to run 2 (i.e., COM2) of each language task within the same scan session. COM$_{VAR}$ is the measurement of relocation of center of activation from run 1 to run 2 in a particular ROI. The box plots of COM$_{VAR}$ obtained from COM1 and COM2 of left hemispheres across all right-handed subjects is displayed. The median values of COM$_{VAR}$ (i.e., distance between COM1 and COM2 in millimeters) across subjects within each ROI show that distance between COM from run 1 to run 2 is less than 5 mm in BA and WA for each of the SC and SWG tasks.

4 | DISCUSSION

This study was undertaken on two well-known and widely accepted language tasks, that is, SC and SWG that are known to activate eloquent language cortex. We examined the variability metrics in a large cohort of patients with brain tumors. In this cohort of patients, we have demonstrated very high degree of repeatability at a single-subject level within single scan sessions. The laterality index and center of mass of holohemispheric and local language areas were investigated in this study for quantification of within-subject variability of activated language clusters.

In a recent review article, Bradshaw, Bishop, and Woodhead (2017), evaluated various methods used in fMRI studies since year 2000 for quantifying laterality. LI (Springer et al., 1999) as a means to evaluate relative extents of activation in the left and right hemispheres has been used widely to assess language hemispheric dominance (Adcock, Wise, Oxbury, Oxbury, & Matthews, 2003; Branco et al., 2006; Pillai & Zaca, 2011; Suarez et al., 2009). The review article by Bradshaw et al. (2017) suggested that threshold-independent LI methods are most beneficial for assessing heterogeneity of language laterality across multiple regions of interests and tasks. The majority of studies on laterality of language fMRI paradigms have been done exclusively on right-handed individuals. Pillai and Zaca (2011) compared ThD versus ThI 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 ThD and ThI analyses and receptive tasks were less effective for language lateralization. SWG, which is an expressive task, provided effective language lateralization even in the subgroup of patients with lesions located in the left hemisphere and in the frontal or parietal lobes. SC, which is both a receptive task and invokes expressive processing (Niskanen et al., 2012; Vigneau et al., 2011), has been demonstrated to less effectively lateralize BA compared to SWG because of involvement of the homologous right hemisphere in speech comprehension tasks. In this study, we also noticed similar results. The highest correlation (which is a measure of reliability) in LI values between consecutive runs of the SC task was in
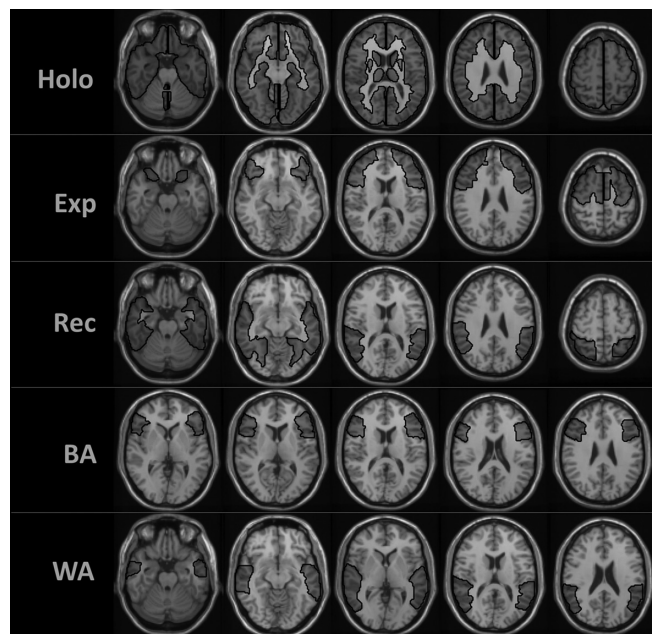
FIGURE 2 Region of interest (ROI) selection: holohemispheric (holo)—left and right hemispheres excluding cerebellum; expressive (exp)—frontal regions; receptive (rec)—temporal and parietal regions; Broca's area (BA)—inferior frontal gyrus; Wernicke's area (WA)— temporal lobe gyri. An automated anatomical labeling (AAL) template was used for the automated atlas-based parcellation of ROIs. Delineated ROIs are being shown overlaid on the MNI 2 mm template
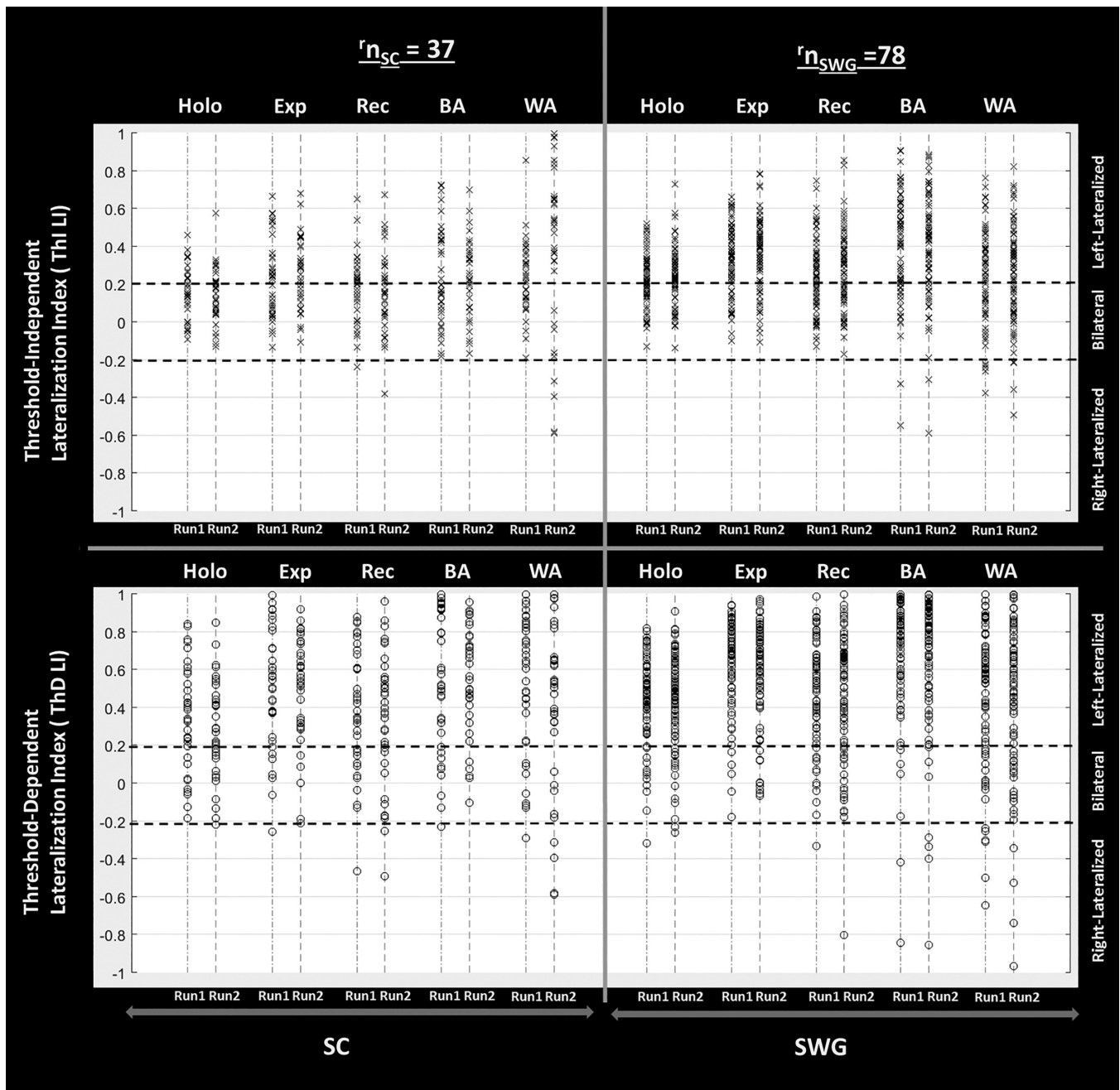
FIGURE 3 Lateralization index LI = (L − H)/(L + H) where L refers to activation in the left hemisphere and R refers to the activation in the right hemisphere. Threshold-independent LI (ThI LI) is calculated from the integrated Z-score weighted distributions of all active voxels, and threshold-dependent LI (ThD LI) is obtained using the integrated weighted distribution of 50% AMPLE (activation mapping as a percentage of local excitation) suprathreshold voxels in Z-score activation maps. Hemispheric dominance for language function is determined by the LI values—left hemispheric dominance indicates LI ≥ 0.2, right hemispheric dominance indicates LI ≤ −0.2 and bilateral/mixed dominance indicates that −0.2 < LI < 0.2. ThI LI and ThD LI values from both sessions of sentence completion (SC) and silent word generation (SWG) language activation maps for holohemispheric and local regions of interest (ROIs) that is, expressive (Exp), receptive (Rec), Broca's area (BA), and Wernicke's area (WA) across the entire right-handed patient population included in this study are displayed below. Each subject ThI LI or ThD LI value is depicted as an individual "x" or "o" symbol respectively in this plot

the receptive ROI (r = 0.69) whereas the highest correlation in LI values in two consecutive runs of the SWG task was in BA (r = 0.86), which indicates the consistency of repeatability across subjects.

In previous studies, it was demonstrated that region-based lateralization indices may provide an alternative method of assessment of language dominance than simply holohemispheric assessments (Zaca et al., 2012, 2013) and in these articles, the SC paradigm was

specifically evaluated in this context. As both expressive and receptive lateralizations are important clinically, regional analysis would complement holohemispheric analysis and therefore, in our current study, we evaluated the within-subject variability both at holohemispheric and regional levels. In addition, use of such smaller ROIs is essential for assessments of center of mass of clinically important language eloquent cortex, such as BA and WA.

For the within-subject variability of holohemispheric and regional language lateralization, LI_{VAR} was calculated across the right-handed patient population. LI_{VAR} is an absolute number which is defined in a manner that it takes into account the signs of lateralization indices (i.e., LI1 and LI2) of two consecutive runs within the same scan session. When LI1 and LI2 are of opposite signs, LI_{VAR} is always greater than 1 ($LI_{VAR} > 1$). Cases where LI1 and LI2 are of same sign (i.e., in same hemisphere) and cases where LI1 and LI2 are of opposite sign (i.e., lateralization to the opposite hemisphere), can have similar differences in LI values (i.e., LI1–LI2 can be similar); therefore, LI_{VAR} was defined as (LI1 – LI2)/(LI1 + LI2) instead of simple subtraction of L2 from L1. Due to the denominator term (LI1 + LI2), cases where LI's are of opposite signs, have larger variability in LI. Our current study found that LI_{VAR} has a task-dependence, similar to the findings by Nadkarni et al. (2014) who explored laterality within expressive versus receptive language tasks. Similarly, our study suggests that LI variability has a threshold-dependence which is similar to the findings by Ruff et al. (2008). Furthermore, LI_{VAR} was found to be lower for the SWG task in comparison to the SC task, which again confirms that the SWG task is a better determinant of language lateralization. The influence of choice of statistical threshold used to map brain activity can be seen on LI_{VAR}. Overall, the median values of LI_{VAR} across all right-handed subjects in each holohemispheric and local ROI indicate that the absolute value of LI in the second run was less than double that of the first run, that is, $LI_{VAR} < 0.33$, indicating

very similar LI values for both runs. The presence of a few outliers (beyond 75th percentile) in which laterality changed from one hemisphere dominance to the other, as shown in Figure 5, may be accounted for by several factors. First of all, in cases where task performance was suboptimal in the first run despite adequate QC metrics reflecting head motion, and performance improved in the second run, one may expect to see differences in lateralization. Secondly, if both runs demonstrated adequate task performance, but the second run was influenced by habituation effects, resulting in less robust activation and more unilateral suprathreshold activation clusters, one would expect to see major changes in lateralization.

Center of mass of regional language activation areas was obtained to localize the centers of activations of each language task. The displacement of COM across two runs was given as measurement of COM_{VAR} in local regions. The displacement of COM coordinates was found to be repeatable within 5 mm in the present patient group. Our finding of higher repeatability of COM in local language ROIs compared to holohemispheric ROIs is similar to the findings in previous studies (Wurnig et al., 2013). Although repeatability of COM of a given cluster is high, it does not necessarily mean that it is representative of the true spatial coordinates of neural activity (Fesl et al., 2008).

As subjects received extensive training prior to scanning, and showed no significant changes in performance over time, the decreases in extent of activation in some cases is likely due to the habituation effect. Similarly, small variations in noise levels may result in large differences in activation extent in thresholded activation maps (Cohen & DuBois, 1999). The heterogeneity of the tumors in this patient cohort, differences in tumor grade, location, size, mass effect, may have contributed to differences in lateralization; however, since our current study was only meant to assess within-subject repeatability between two consecutive runs of same task within the same scan session, these factors should equally affect both runs since no major changes in physiologic variables is expected within the same scan session. Another limitation is the presence of neurovascular uncoupling (NVU) which may cause regional reductions in BOLD signal in the vicinity of tumors and other lesions, thus resulting in incorrect lateralization. However, since assessment of inter-subject variability of lateralization or localization of eloquent language cortex was not the focus of this study, but rather assessment of intra-subject repeatability, we would expect identical effects of NVU in both consecutive runs of a language fMRI task for any given individual subject. Thus, our assessment of within-session inter-run intra-subject variability, that is, repeatability, would not be adversely affected by NVU or other effects, such as susceptibility, regional mass effect, gyral compression, or functional cortical displacement that would compromise detectability of the BOLD signal.

In conclusion, we examined two well-known and widely accepted language tasks that are known to activate eloquent language cortex, based on multiple prior publications and national efforts that are currently underway to standardize language fMRI. We have demonstrated very high degree of repeatability at a single-subject level within single scan sessions of language mapping in a large cohort of patients undergoing presurgical fMRI across several years at our institution. Holohemispheric and regional language lateralization indicated
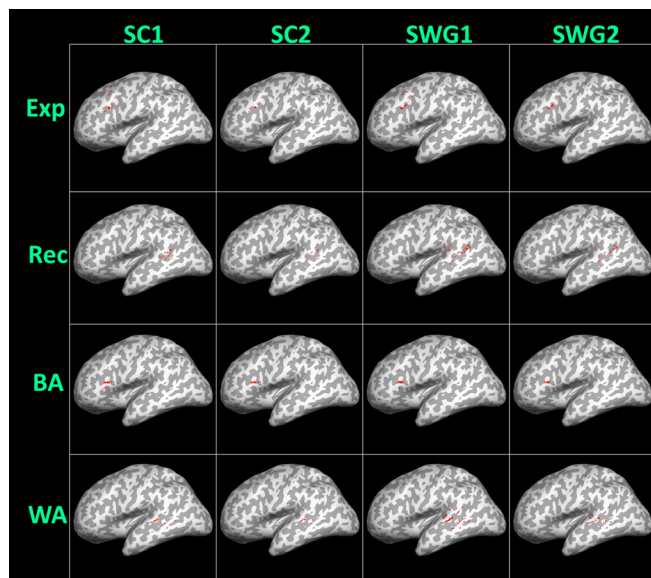
FIGURE 4 Left hemispheric threshold-dependent center of mass (ThD COM), representing the center of mass of 50% AMPLE (activation mapping as a percentage of local excitation) suprathreshold activated voxels from both consecutive runs of sentence completion (SC) and silent word generation (SWG) language tasks, respectively, within the same scan session, is displayed for all right-handed subjects. Individual subject ThD COM values for expressive (Exp), receptive (Rec), Broca's area (BA), and Wernicke's area (WA) regions of interest are depicted as dots on the 3D inflated sagittal surface of the MNI 2 mm anatomical template. The underlying gyral anatomy confirms the localization of activation areas [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Correlation coefficient (*r*) of lateralization index (LI) values between intra-session consecutive runs of sentence completion (SC) and silent word generation (SWG) tasks in each of holohemispheric and local regions of interest (ROIs), that is, expressive (Exp), receptive (Rec), Broca's area (BA), and Wernicke's area (WA) obtained across the right-handed subject population is tabulated

		$^r n_{SC} = 37$		$^r n_{SWG} = 78$	
		Median of LI	Correlation coefficient (*r*)	Median of LI	Correlation coefficient (*r*)
Holohemispheric	ThI LI	0.15	0.55	0.23	0.76
	ThD LI	0.33	0.63	0.46	0.79
Expressive	ThI LI	0.24	0.49	0.36	0.73
	ThD LI	0.52	0.52	0.66	0.81
Receptive	ThI LI	0.18	0.69	0.26	0.73
	ThD LI	0.39	0.63	0.51	0.64
Broca's area	ThI LI	0.22	0.59	0.44	0.79
	ThD LI	0.53	0.67	0.78	0.86
Wernicke's area	ThI LI	0.32	0.57	0.24	0.53
	ThD LI	0.54	0.63	0.51	0.72

The median of LI values from both runs of each language task is also provided. The highest correlation in LI values between two consecutive runs of the SWG task was in Broca's area (*r* = 0.86) and the highest correlation in LI values for the SC task was in the receptive region (*r* = 0.69).

the high intra-subject repeatability when lateralization indices are considered obtained using ThD and ThI approaches. The repeatability 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.
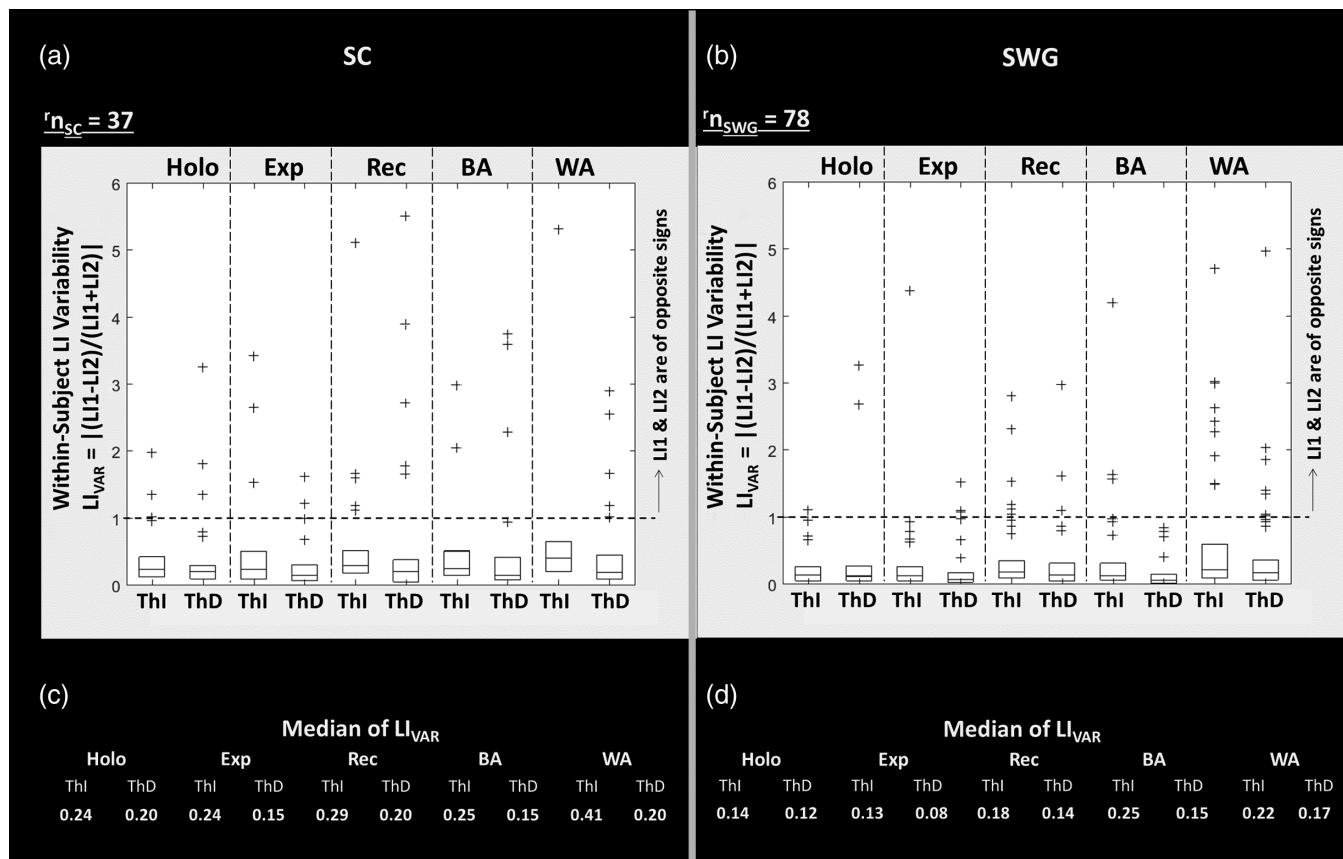
FIGURE 5 Within-subject lateralization index variability $LI_{VAR} = |(LI1 − LI2)/(LI1 + LI2)|$ represents the variability in lateralization index (LI) from run 1 to run 2 in holohemispheric and local regions of interest (ROIs), that is, expressive (Exp), receptive (Rec), Broca's area (BA), and Wernicke's area (WA). The box plots of LI_{VAR} obtained across all right-handed subjects is displayed. In each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Those patients with LI_{VAR} beyond 75th percentile are plotted individually using the "+" symbol. Note that cases for which $LI_{VAR} > 1$ were those where hemispheric dominance changed from one hemisphere to the other, that is, LI1 and LI2 are of opposite signs. The median values of LI_{VAR} across all right-handed subjects in each holohemispheric and local ROIs are listed in c and d parts of the figure for the sentence completion (SC) and silent word generation (SWG) tasks, respectively
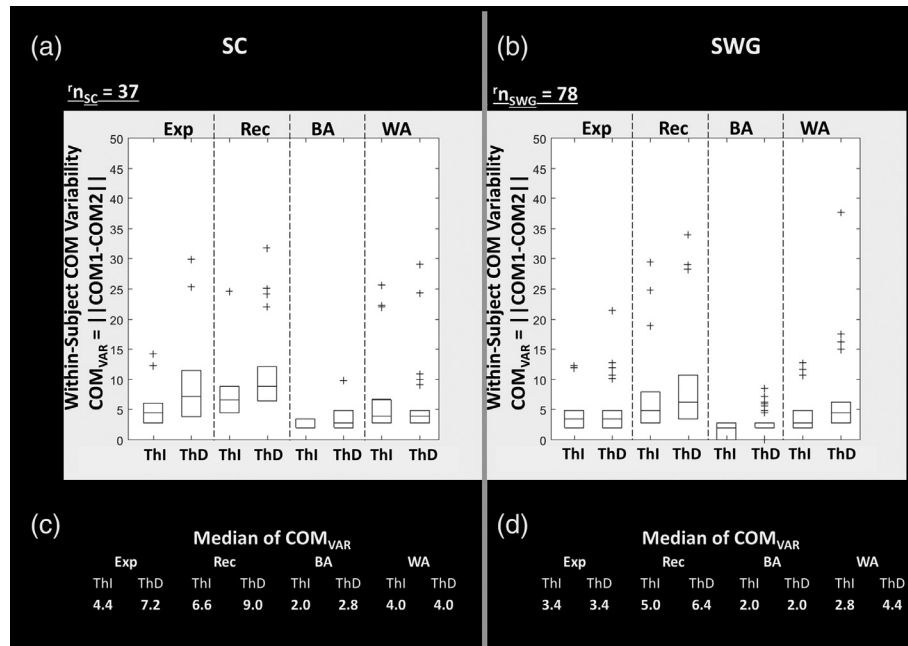
FIGURE 6 Within-subject center of mass (COM) variability $COM_{VAR} = \|COM1 - COM2\|$ represents the variability in COM location from run 1 (COM1) to run 2 (COM2) in local regions of interest (ROIs), that is, expressive (Exp), receptive (Rec), Broca's area (BA), and Wernicke's area (WA). COM_{VAR} is referred to as the Euclidean distance (in mm) between COM1 and COM2. The box plots of COM_{VAR} obtained from COM1 and COM2 of left hemispheres across all right-handed subjects is displayed. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles of distance between COM1 and COM2 across subjects. Those patients with COM_{VAR} beyond 75th percentile are plotted individually using the "+" symbol. The median values of COM_{VAR} (i.e., distance between COM1 and COM2 in mm) across all subjects in each ROI are listed in c and d parts of the figure for sentence completion (SC) and silent word generation (SWG) tasks, respectively; note that these values are <5 mm for both Broca's area (BA) and Wernicke's area (WA)

CONFLICT OF INTEREST

The authors declared that they have no conflict of interest with the contents of this article.

ORCID

Shruti Agarwal 🔟 https://orcid.org/0000-0003-0902-5354
Jay J. Pillai 🔟 https://orcid.org/0000-0002-8306-7192

REFERENCES

Adcock, J. E., Wise, R. G., Oxbury, J. M., Oxbury, S. M., & Matthews, P. M. (2003). Quantitative fMRI assessment of the differences in lateralization of language-related brain activation in patients with temporal lobe epilepsy. *NeuroImage*, *18*, 423–438.

Binder, J. R., Swanson, S. J., Hammeke, T. A., Morris, G. L., Mueller, W. M., Fischer, M., … Haughton, V. M. (1996). Determination of language dominance using functional MRI: A comparison with the Wada test. *Neurology*, *46*, 978–984.

Black, D. F., Vachha, B., Mian, A., Faro, S. H., Maheshwari, M., Sair, H. I., … Welker, K. (2017). American Society of Functional Neuroradiology-Recommended fMRI paradigm algorithms for presurgical language assessment. *AJNR. American Journal of Neuroradiology*, *38*(10), E65–E73. https://doi.org/10.3174/ajnr.A5345

Bradshaw, A. R., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Methodological considerations in assessment of language lateralisation with fMRI: A systematic review. *PeerJ*, *5*, e3557. https://doi.org/10.7717/peerj.3557

Branco, D. M., Suarez, R. O., Whalen, S., O'Shea, J. P., Nelson, A. P., & da Costa, J. C. (2006). Functional MRI of memory in the hippocampus: Laterality indices may be more meaningful if calculated from whole voxel distributions. *NeuroImage*, *32*(2), 592–602.

Broca, P. (1861). Remarques Sur le siège de la faculté du langage articulé; suivies d'une observation d'aphemie. *Bulletins de la Société Anatomique de Paris*, *6*, 330–357.

Caceres, A., Hall, D. L., Zelaya, F. O., Williams, S. C. R., & Mehta, M. A. (2009). Measuring fMRI reliability with the intra-class correlation coefficient. *NeuroImage*, *45*(3), 758–768. https://doi.org/10.1016/j.neuroimage.2008

Carp, J. (2013). Better living through transparency: Improving the reproducibility of fMRI results through comprehensive methods reporting. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(3), 660–666. https://doi.org/10.3758/s13415-013-0188-0

Chen, E. E., & Small, S. L. (2007). Test-retest reliability in fMRI of language: Group and task effects. *Brain and Language*, *102*(2), 176–185.

Cohen, M. S., & DuBois, R. M. (1999). Stability, repeatability, and the expression of signal magnitude in functional magnetic resonance imaging. *Journal of Magnetic Resonance Imaging*, *10*, 33–40.

Fernández, G., Specht, K., Weis, S., Tendolkar, I., Reuber, M., Fell, J., … Elger, C. E. (2003). Intrasubject reproducibility of presurgical language lateralization and mapping using fMRI. *Neurology*, *60*(6), 969–975.

Fesl, G., Braun, B., Rau, S., Wiesmann, M., Ruge, M., Bruhns, P., … Brückmann, H. (2008). Is the center of mass (COM) a reliable parameter for the localization of brain function in fMRI? *European Radiology*, *18*(5), 1031–1037. https://doi.org/10.1007/s00330-008-0850-z

Fesl, G., Bruhns, P., Rau, S., Wiesmann, M., Ilmberger, J., Kegel, G., & Brückmann, H. (2010). Sensitivity and reliability of language laterality

assessment with a free reversed association task—A fMRI study. *European Radiology*, 20(3), 683–695. https://doi.org/10.1007/s00330-009-1602-4

FitzGerald, D. B., Cosgrove, G. R., Ronner, S., Jiang, H., Buchbinder, B. R., Belliveau, J. W., ... Benson, R. R. (1997). Location of language in the cortex: A comparison between functional MR imaging and electrocortical stimulation. *AJNR. American Journal of Neuroradiology*, 18, 1529–1539.

Gaillard, W. D., Balsamo, L., Xu, B., Grandin, C. B., Braniecki, S. H., Papero, P. H., ... Theodore, W. H. (2002). Language dominance in partial epilepsy patients identified with an fMRI reading task. *Neurology*, 59(2), 256–265.

Gartus, A., Foki, T., Geissler, A., & Beisteiner, R. (2009). Improvement of clinical language localization with an overt semantic and syntactic language functional MR imaging paradigm. *AJNR. American Journal of Neuroradiology*, 30, 1977–1985.

Gorgolewski, K. J., Storkey, A. J., Bastin, M. E., Whittle, I., & Pernet, C. (2013). Single subject fMRI test-retest reliability metrics and confounding factors. *NeuroImage*, 69, 231–243. https://doi.org/10.1016/j.neuroimage.2012

Harrington, G. S., Buonocore, M. H., & Tomaszewski Farias, S. (2006). Intrasubject reproducibility of functional MR imaging activation in language tasks. *American Journal of Neuroradiology*, 27(4), 938–944.

Maïza, O., Mazoyer, B., Hervé, P. Y., Razafimandimby, A., Dollfus, S., & Tzourio-Mazoyer, N. (2011). Reproducibility of fMRI activations during a story listening task in patients with schizophrenia. *Schizophrenia Research*, 128(1–3), 98–101. https://doi.org/10.1016/j.schres.2011.01.025

Maldjian, J. A., Laurienti, P. J., Driskill, L., & Burdette, J. H. (2002). Multiple reproducibility indices for evaluation of cognitive functional MR imaging paradigms. *American Journal of Neuroradiology*, 23(6), 1030–1037.

Morrison, M. A., Churchill, N. W., Cusimano, M. D., Schweizer, T. A., Das, S., & Graham, S. J. (2016). Reliability of task-based fMRI for preoperative planning: A test-retest study in brain tumor patients and healthy controls. *PLoS One*, 11(2), e0149547. https://doi.org/10.1371/journal.pone.0149547

Nadkarni, T. N., Andreoli, M. J., Nair, V. A., Yin, P., Young, B. M., Kundu, B., ... Prabhakaran, V. (2014). Usage of fMRI for pre-surgical planning in brain tumor and vascular lesion patients: Task and statistical threshold effects on language lateralization. *NeuroImage: Clinical*, 7, 415–423. https://doi.org/10.1016/j.nicl.2014

Niskanen, E., Könönen, M., Villberg, V., Nissi, M., Ranta-Aho, P., Säisänen, L., ... Vanninen, R. (2012). The effect of fMRI task combinations on determining the hemispheric dominance of language functions. *Neuroradiology*, 54, 393–405. https://doi.org/10.1007/s00234-011-0959-7

Ojemann, G. A. (1979). Individual variability in cortical localization of language. *Journal of Neurosurgery*, 50, 164–169.

Otzenberger, H., Gounot, D., Marrer, C., Namer, I. J., & Metz-Lutz, M. N. (2005). Reliability of individual functional MRI brain mapping of language. *Neuropsychology*, 19(4), 484–493.

Pillai, J. J., & 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. https://doi.org/10.1016/j.neuroimage.2010.03.082.

Poline, J. B., Strother, S. C., Dehaene-Lambertz, G., Egan, G. F., & Lancaster, J. L. (2006). Motivation and synthesis of the FIAC experiment: Reproducibility of fMRI results across expert analyses. *Human Brain Mapping*, 27(5), 351–359.

Raemaekers, M., Du Plessis, S., Ramsey, N. F., Weusten, J. M. H., & Vink, M. (2012). Test-retest variability underlying fMRI measurements. *NeuroImage*, 60, 717–727. https://doi.org/10.1016/j.neuroimage.2011

Ruff, I. M., Petrovich Brennan, N. M., Peck, K. K., Hou, B. L., Tabar, V., Brennan, C. W., & Holodny, A. I. (2008). Assessment of the language laterality index in patients with brain tumor using functional MR imaging: Effects of thresholding, task selection, and prior surgery. *American Journal of Neuroradiology*, 29(3), 528–535. https://doi.org/10.3174/ajnr.A0841

Rutten, G. J., Ramsey, N. F., van Rijen, P. C., & van Veelen, C. W. (2002). Reproducibility of fMRI-determined language lateralization in individual subjects. *Brain and Language*, 80(3), 421–437.

Rutten, G. J., van Rijen, P. C., van Veelen, C. W., & Ramsey, N. F. (1999). Language area localization with three-dimensional functional magnetic resonance imaging matches intrasulcal electrostimulation in Broca's area. *Annals of Neurology*, 46, 405–408.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3), 143–155. https://doi.org/10.1002/hbm.10062

Springer, J. A., Binder, J. R., Hammeke, T. A., Swanson, S. J., Frost, J. A., & Bellgowan, P. S. (1999). Language dominance in neurologically normal and epilepsy subjects: A functional MRI study. *Brain*, 122(Pt 11), 2033–2046.

Stevens, M. T. R., Clarke, D. B., Stroink, G., Beyea, S. D., & D'Arcy, R. C. (2015). Improving fMRI reliability in presurgical mapping for brain tumours. *Journal of Neurology, Neurosurgery, and Psychiatry*, 87(3), 1–8.

Suarez, R. O., Whalen, S., Nelson, A. P., Tie, Y., Meadows, M. E., Radmanesh, A., & Golby, A. J. (2009). Threshold-independent functional MRI determination of language dominance: A validation study against clinical gold standards. *Epilepsy & Behavior*, 16(2), 288–297. https://doi.org/10.1016/j.yebeh.2009.07.034

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., ... Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, 54(1), 577–593. https://doi.org/10.1016/j.neuroimage.2010.07.036

Voyvodic, J. T. (2012). Reproducibility of single-subject fMRI language mapping with AMPLE normalization. *Journal of Magnetic Resonance Imaging*, 36(3), 569–580. https://doi.org/10.1002/jmri.23686

Wada, J. A., & Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance. *Journal of Neurosurgery*, 17, 266–282.

Wurnig, M. C., Rath, J., Klinger, N., Höllinger, I., Geissler, A., Fischmeister, F. P., ... Beisteiner, R. (2013). Variability of clinical functional MR imaging results: A multicenter study. *Radiology*, 268, 521–531.

Zaca, D., Jarso, S., & Pillai, J. J. (2013). Role of semantic paradigms for optimization of language mapping in clinical FMRI studies. *AJNR. American Journal of Neuroradiology*, 34(10), 1966–1971. https://doi.org/10.3174/ajnr.A3628

Zaca, D., Nickerson, J. P., Deib, G., & Pillai, J. J. (2012). Effectiveness of four different clinical fMRI paradigms for preoperative regional determination of language lateralization in patients with brain tumors. *Neuroradiology*, 54(9), 1015–1025. https://doi.org/10.1007/s00234-012-1056-2

Zandbelt, B. B., Gladwin, T. E., Raemaekers, M., van Buuren, M., Neggers, S. F., Kahn, R. S., ... Vink, M. (2008). Within-subject variation in BOLD-fMRI signal changes across repeated measurements: Quantification and implications for sample size. *NeuroImage*, 42(1), 196–206. https://doi.org/10.1016/j.neuroimage.2008.04.183

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.