Neuroradiology (2005) 47: 835–844
DOI 10.1007/s00234-005-1431-3

M. V. Baciu
J. M. Watson
L. Maccotta
K. B. McDermott
R. L. Buckner
F. G. Gilliam
J. G. Ojemann

Evaluating functional MRI procedures for assessing hemispheric language dominance in neurosurgical patients

Received: 16 November 2004
Accepted: 20 May 2005
Published online: 2 September 2005

M. V. Baciu (✉)
Laboratory of Psychology and Neurocognition, UMR CNRS 5105, Pierre Mendes-France University, BP 47, 38040 Grenoble Cedex 09, France
E-mail: mbaciu@upmf-grenoble.fr

J. M. Watson · L. Maccotta
K. B. McDermott · R. L. Buckner
Department of Psychology, Washington University, St. Louis, USA

R. L. Buckner
Howard Hughes Medical Institute at Washington University, St. Louis, USA

F. G. Gilliam
Department of Neurology, Washington University School of Medicine, St. Louis, USA

J. G. Ojemann
Department of Neurological Surgery, Washington University School of Medicine, St. Louis, USA

Abstract Two methods of quantifying hemispheric language dominance (HLD) in neurosurgical patients are compared: (1) an average magnitudes (AM) method, which is a calculation of the average signal intensity variation in regions of interest for each patient that were predefined in a group analysis for each task, and (2) a lateralization indices (LI) method, which is based on the number of activated pixels in regions of interest predefined in each individual patient. Four language tasks [a living/nonliving (LNL) judgment, word stem completion (WSC), semantic associate (SA) and a phonological associate (PA) task] were compared with "gold standard" measures such as the Wada test or electrocortical stimulation. Results showed that the LI method was more accurate (73% agreement with gold standard methods) than the AM method (only 40% agreement) across tasks and subjects. Furthermore, by varying the threshold used for determining laterality, the ability of functional magnetic resonance imaging (fMRI) to predict HLD was influenced for the AM method, whereas the LI method was relatively unaffected by changing the threshold. Using the LI method, the SA task was the most accurate for quantifying HLD (100% agreement with gold standard methods) with respect to the other three language tasks (80% accuracy for WSC, 65% for the LNL and 63% for phonological task). Depending on the method and the task, fMRI may be a promising tool for assessing HLD in neurosurgical patients.

Keywords Functional MRI · Hemispheric language dominance · lateralization · Epilepsy · Wada test · Electrocortical stimulation

Introduction

The pre-surgical evaluation of hemispheric language dominance (HLD) in epileptic patients is classically performed with "gold standard" methods such as the Wada test [24] or electrocortical stimulation (ES) [15]. However, the ability of these gold standard methods to predict patient outcome is variable, in part because methodological differences may exist between testing centres [10]. Many neuroimaging studies have evaluated the ability of functional magnetic resonance imaging (fMRI) as an alternative tool to determine HLD [2, 3, 5, 6, 17, 18, 20, 21, 27]. In general, when assessing HLD, high agreement (e.g., a range of 91–95%) has been observed between Wada and fMRI results [19, 25]. For example, by using a combined analysis of four language tasks, Rutten et al. [18] observed good agreement (91%) between fMRI results and the Wada test when the left hemisphere was dominant for language. In this same study, the concordance between fMRI and Wada decreased to 67% when the right hemisphere was dominant for language and decreased to 75% when there was a bilateral representation of language (see [1] for similar findings).

High agreement (e.g., a range of 81–100%) for HLD assessment has also been observed between ES and fMRI [7, 16, 26]. Language dominance is usually quantified in fMRI studies by calculating lateralization indices (LI). Their computation is based on the number of activated pixels detected within the two hemispheres. However, there may be inconsistencies between lateralization estimates provided by gold standard methods and those provided by fMRI methods. One potential source of inconsistency could be related to LI computation. Specifically, LI are typically calculated by counting the pixels within activated regions without knowing whether such regions are indeed critical for language. Hence, one drawback of fMRI is that it does not distinguish between essential and nonessential regions activated during task performance. If pixels belonging to nonessential language regions were included in the LI computation, one might reach an inappropriate conclusion about the hemispheric dominance for language. One solution might be to first identify regions that might be critical for language and to compute LI based only on pixels belonging to these regions. Consistent with this idea, it has been suggested that regions active during fMRI across subjects may be the regions essential for language [17, 18]. The language tasks used in epileptic patients typically include semantic decision [5, 6, 21], phonological decision [2, 4], word generation [4, 9, 11, 23, 27] or sentence reading [8, 17].

The present study had two main aims. The first aim was to compare two methods of quantifying the HLD when using fMRI in neurosurgical patients. With the average magnitudes (AM) method, an initial analysis was performed at the group level to identify regions that might be critical for language for each task. This was followed by the calculation of the average signal intensity variation in these predefined critical regions at the level of single patients within each task. The comparison of average signal intensity levels in homologous regions in the two hemispheres provided an estimate of the HLD. The second method involved the calculation of LI based on the number of activated pixels in regions similar to those identified by the AM method. Lateralization results from each method, AM versus LI, were then compared with results of the Wada test, ES or post-surgical outcome (PS) for each patient. The second aim of the study was to determine which among the four tasks—living/nonliving (LNL), word stem completion (WSC), semantic associate (SA) or phonological associate (PA)—is the most consistent with estimates of HLD provided by gold standard methods such as the Wada test, ES or PS.

Methods

Patients

A total of 52 fMRI examinations were performed on 35 patients including 26 scan sessions for a LNL judgment, ten scan sessions for a WSC task, eight scan sessions for a SA task and eight scan sessions for a PA task (i.e., some patients completed multiple tasks). All 35 participants in the present study were epileptic patients and the fMRI examinations were performed within the framework of their pre-surgical evaluation. Table 1 summarizes some clinical data (e.g., location of lesion, age and handedness) for each group of patients. All tumors presented in these patients were gliomas.

Table 1 Clinical data for epileptic patients who performed a LNL, WSC, SA, or PA task, respectively (*EZ* epileptogenic zone)

Tasks/number of patients	Lesion regional location	Lesion hemispheric location	Average age	Handedness
LNL/26 patients	Frontal tumors (8) Fronto-temporal tumor (1) Temporal tumors (5) Temporal EZ (12)	Left hemisphere (24) Right hemisphere (2)	32.7 years	Right handed (22) Left handed (3) Ambidextrous (1)
WSC/10 patients	Frontal tumors (5) Temporal tumors (5)	Left hemisphere (10)	34.3 years	Right handed (9) Ambidextrous (1)
SA and PA/8 patients	Frontal tumors (1) Temporal tumors (4) Frontal seizures (1) Parietal tumor (1) Fronto-temporal seizures (1)	Left hemisphere (8)	33.5 years	Right handed (6) Left handed (2)

The type and the location of the lesion, the age, and the handedness are indicated for each group of patients

Stimuli

In all four tasks, the stimuli were English words presented through a Power Macintosh G3 computer (Apple Computer) and projected with a Sharp LCD projector onto a mirror placed over the head coil. All words were presented in the center of the field of view.

Tasks

LNL task

At the start of each scan, subjects maintained visual fixation on a centrally placed crosshair. After an initial fixation period, subjects were presented with blocks of low frequency words (12 words/block) alternating with blocks of fixation crosshair. On each trial, the stimulus was presented for 2 s, followed by 500 ms of fixation crosshair. Subjects were instructed to decide if each word shown was either a living or a nonliving object and to make a key press response with their right index finger if the word represented a living object. Three 12-word blocks were presented in each functional scan (3 min/scan) for a total of 108 words presented over three functional scans.

WSC task

At the start of each scan, subjects maintained visual fixation on a centrally placed crosshair. After an initial fixation period, subjects were presented with blocks of three letter strings (12 strings/block) alternating with blocks of fixation crosshair. In each trial, the stimulus was presented for 2 s, followed by 500 ms of fixation crosshair. Subjects were instructed to complete silently each word stem without speaking aloud (with emphasis on avoiding movement of the head or jaw). Three blocks of stem completion were performed in each functional scan (3 min/scan) for a total of 108 stems presented over three functional scans.

SA and PA tasks

As described by McDermott et al. [12] in five, six or eight scans, subjects studied lists of associates (16 words/list, 12 lists/scan, 5 min/scan). A blocked design was used such that each subject studied semantic and phonological lists (randomly-ordered) within each scan. At the beginning of each list of words, a cue was displayed (i.e., ‘‘meaning’’ or ‘‘rhyme’’) to inform subjects of the type of list that they were about to see. Subjects were instructed to use the cue to help them focus on the meaning or sound relations among the upcoming words. Specifically, in the semantic condition, they were told to think about how each word (e.g., *tiger*, *circus*, *jungle*) could be meaningfully connected to the other words in the list, whereas in the rhyme condition they were told to think about how each word (e.g., *skill*, *fill*, *hill*) sounded like the other words in the list and to say each word silently to themselves. Words were displayed quickly so that each 16-word list was displayed in about 10 s. The individual words within a list were displayed one at a time for approximately 560 ms each with a 50 ms inter-stimulus interval. Following each block of words, there was a brief period (12.5 s) in which subjects were shown a crosshair and asked to fixate on it and await the next list of associates.

Data acquisition

Scans were obtained on a 1.5 Tesla Siemens Vision System (Erlangen, Germany) with a standard circularly polarized head coil. Structural images were acquired using a high-resolution T1-weighted sagittal sequence. Functional images were collected with an echoplanar asymmetric spin-echo sequence [3, 12]. In each functional scan, 76 (during LNL and WSC tasks) or 128 (during SA and PA tasks) sets of 16 contiguous, 8-mm-thick axial images (TR = 2500 ms, TE = 50 ms, 3.75×3.75 mm in-plane resolution) were acquired parallel to the anterior–posterior commissure plane.

Data processing and analysis

Analyses of data collected were performed using the general linear model (GLM) implemented with in-house software [13].

Pre-processing of raw data

Several steps were followed. The first step was a slice-by-slice normalization to correct changes in signal intensity introduced by the acquisition of interleaved slices. In the second step, the images were realigned within and across scans by using rigid-body translations and rotations. The rigid-body rotations and translations provided data on the adjustments required for realignment of each image at each MR frame. The functional images were spatially normalized in 2×2×2 mm^3 Talairach space [22].

Statistical analysis

To detect task-correlated activation, functional data were analysed using an implementation of the GLM. Comparisons between task and control periods were performed for each task. When the LI method was used for evaluating HLD, individual analyses were performed using the frontal and temporal activity for each patient in each task ($P < 0.05$, uncorrected). When the AM method was used for determining HLD, a random group analysis was performed to determine the commonly activated regions across subjects for each task. In this group analysis, the contrasts between conditions were corrected for multiple comparisons ($P < 0.05$, corrected). Subsequently, the AM of the MR signal variation within these commonly activated, multiple comparison corrected, predefined frontal and temporoparietal regions were calculated for each patient who received a particular task.

fMRI assessment of hemispheric language dominance

Two methods were used to assess HLD:

(1) Lateralization based on regions predefined by group level analysis (or AM method). As noted above, with the AM method, a group analysis was performed first, followed by a patient-by-patient analysis. The group analysis was conducted on data from all of the patients who performed a particular task. Specifically, the group analysis consisted of a voxelwise *t*-test ($P < 0.05$, corrected) that isolated brain regions significantly active during task performance relative to the control period (fixation) in all patients who performed the same language task (i.e., the LNL, WSC, SA or PA task). After obtaining a set of regions of interest (ROIs) for each task, further analysis was restricted to frontal (near Brodmann Area 44, 45, 46, and 47) and temporoparietal regions (near Brodmann Area 37, 21, 22, and 40) because, a priori, one might argue that in general these anterior and posterior networks are indeed critical for language (roughly corresponding to Broca's and Wernicke's areas, respectively). Interestingly, to preview, as shown in Fig. 1, the group analysis revealed that temporoparietal regions were active only during the LNL and SA tasks, whereas frontal regions were active during all four tasks (including the WSC and PA tasks). Therefore, masks were created representing the frontal or temporoparietal regions of interest for a particular language task, and these masks were applied to each patient's functional data on that task. In this way, the average signal level for all activated pixels within the frontal and temporoparietal ROIs were computed for each patient that performed the LNL and SA tasks, and within the frontal ROI for each patient that performed the WSC and PA tasks. The criterion used for assessing HLD in each patient was that the dominant hemisphere must have a signal level of at least twice (i.e., an AM ratio of 2) that of its non-dominant homologue. However, this criterion is somewhat arbitrary. To assess the influence of threshold variation on estimates of HLD, a slightly lower (an AM ratio of 1.5) and a slightly higher (an AM ratio of 2.5) threshold than the "standard" criterion of 2 were also used.

(2) Lateralization based on LI (or LI method). As noted above, with the LI method, individual analyses (*t*-tests) were performed for each patient in each task ($P < 0.05$, uncorrected). More conservative thresholds were used if a patient's data were too noisy at a threshold of $P < 0.05$. LI were calculated based on the number of activated pixels in regions of interest predefined in each patient. Specifically, each index was expressed as a ratio of the difference in the number of active pixels in left frontal and temporoparietal regions and the number of active pixels in right frontal and temporoparietal regions divided by the total number of pixels active in these left and right hemisphere regions. As suggested by Binder et al. [5], an $LI > +0.20$ indicates a left hemisphere dominance for language, an $LI < -0.20$ indicates a right hemisphere dominance for language, and an LI between -0.20 and $+0.20$ indicates a bilateral representation for language. Although these LI values have been adopted as the standard by the majority of the fMRI studies using the LI method to assess HLD, to determine the influence of threshold variation on estimates of HLD, a slightly lower (0.15) and a slightly higher (0.25) LI threshold than the standard value of 0.20 were also used (enabling an additional comparison of what influence, if any, threshold manipulations have on estimates of HLD provided by the LI method).

Gold standard assessments of hemispheric language dominance

Language dominance as determined with fMRI methods was compared with assessment of "true" dominance as determined by other methods, including the Wada test, ES and PS. Of the 35 patients included in this study, HLD was determined with Wada in 13 patients, ES in 19 patients, and PS in 3 patients. In some patients, lateralization was determined by the presence or absence of clinical symptoms after surgery in or directly around language related cortex. For example, a patient who underwent an extensive right temporal lobectomy without any speech disturbance post-operatively would be assumed to be left dominant for language. The authors were blind to the gold standard assessment of HLD for each patient when using the fMRI data to determine HLD.

Results

Figure 1 represents the commonly activated regions during fMRI for each of the four tasks used in the

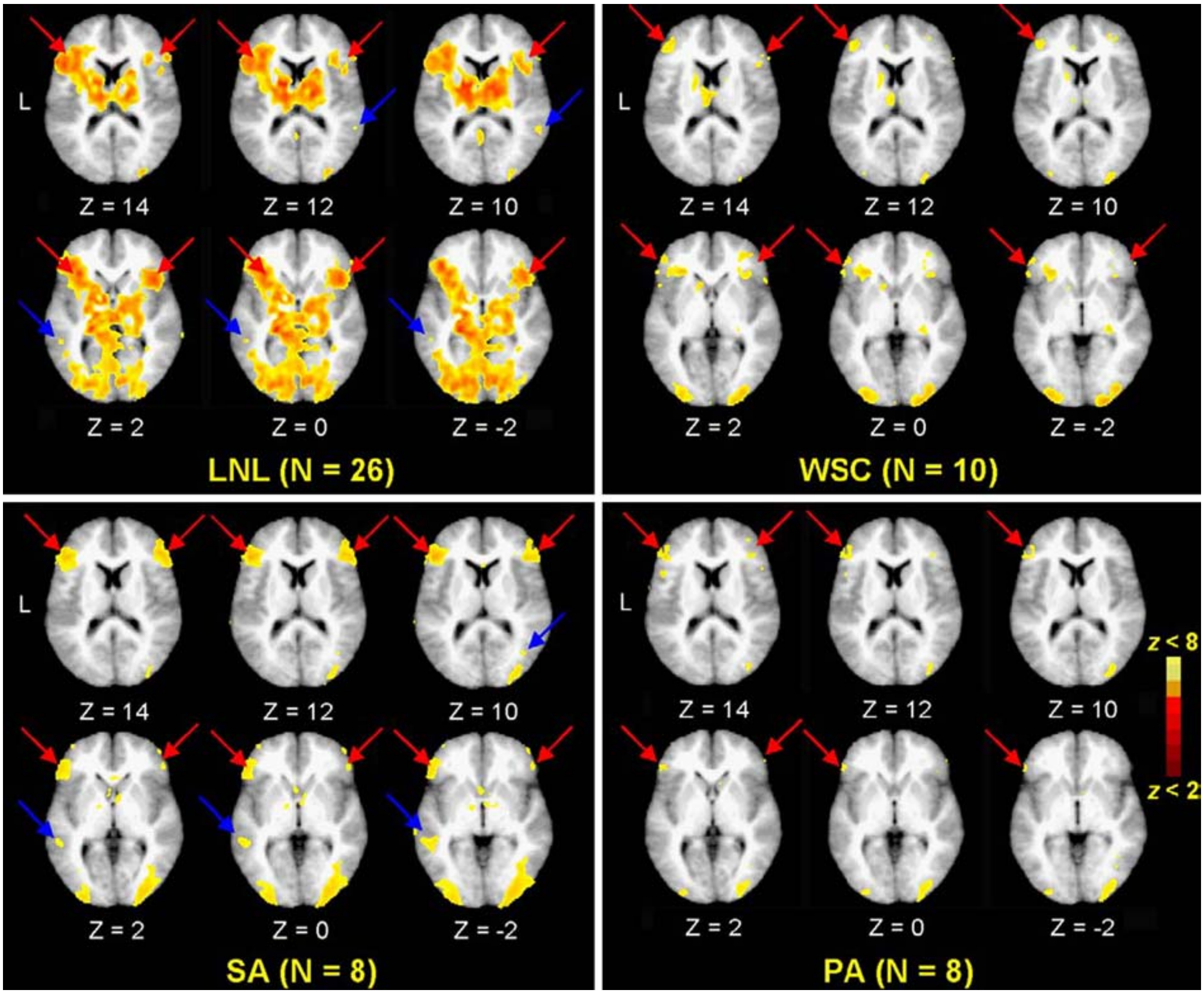

Fig. 1 Representative functional maps obtained during group analyses of the living/nonliving (*LNL, upper left corner*), word stem completion (*WSC, upper right corner*), semantic associate (*SA, lower left corner*), and phonological associate tasks (*PA, lower right corner*). Activity in frontal regions of interest is indicated by *red arrows*, and activity in temporal regions of interest is indicated by *blue arrows*. All of the images are multiple comparison corrected and presented in neurological convention (left hemisphere is on the *left*). In all cases, task activity is relative to fixation baseline

Table 2 The percentage of hemispheric language dominance predictions concordant with gold standard measures of dominance by using either LI (threshold = 0.20) or AM (threshold = 2) as a function of task

Task/number of patients	AM concordance	LI concordance
LNL 26	13/2650%	17/2665%
WSC 10	5/1050%	8/1080%
SA 8	1/813%	8/8100%
PA 8	2/825%	5/863%
Total patients = 52	21/5240%	38/5273%

present study. Among them, only frontal (indicated by *red arrows*) and temporoparietal (indicated by *blue arrows*) regions were used to predefine regions of interest. As indicated in Table 2, by using the AM method with an arbitrary threshold of two, applied across all four tasks, only 21 of the 52 fMRI examinations performed (40%) were in agreement with gold standard measures of HLD (i.e., the Wada test, ES or PS). By comparison, when a more liberal threshold of 1.5 was used, 23/52 (44%) patients were in agreement. However, when a more conservative threshold of 2.5 was used, only 11/52 (21%) patients were in agreement. Turning to the LI computation method with a standard threshold of 0.20, applied across all four tasks, 38 of the 52 fMRI examinations (73%) performed were in agreement with gold standard measures of HLD. Indeed, as shown in the right-hand column of Table 2, the LI method was always better in evaluating HLD than the AM method (even at the level of the individual tasks). Specifically, using an LI method and the standard threshold, the SA task was the best predictor of HLD as determined by gold standard measures like Wada or ES (i.e., a con-

Table 3 Hemispheric language dominance as determined by a LNL task [*AM LH* average signal intensity in the left hemisphere, *AM RH* average signal intensity in the right hemisphere, *AM ratio* ratio between the higher AM hemispheric value and the lower AM hemispheric value; *HLD fMRI AM* hemispheric language dominance as assessed by the AM method of quantifying fMRI results, *HLD fMRI LI* hemispheric language dominance as assessed by the lateralization indices method; *"True" HLD* hemispheric language dominance as determined by gold standard measures including Wada test, ES or PS; *LH* left dominant, *RH* right dominant, *BH* bilateral representation of language; *Y* concordant fMRI assessment of HLD (listed separately for the AM and the LI methods) with gold standard measures of HLD, *N* discordant assessment of HLD]

Patient number	AM LH	AM RH	AM ratio	LI	HLD fMRI AM	HLD fMRI LI	"True" HLD	Concordance
1	0.07	0.30	4.29	0	RH	BH	ES, LH	N, N
2	0.09	0.07	1.29	0.35	BH	LH	PS, LH	N, Y
3	0.37	0.55	1.49	−0.40	BH	RH	PS, LH	N, N
4	0.33	0.44	1.33	−0.62	BH	RH	ES, RH	N, Y
5	0.08	0.02	4	0.24	LH	LH	ES, LH	Y, Y
6	0.48	0.05	9.6	0.93	LH	LH	ES, LH	Y, Y
7	0.52	0.56	1.08	0.23	BH	LH	ES, LH	N, Y
8	0.44	0.20	2.2	0.50	LH	LH	ES, LH	Y, Y
9	0.36	0.17	2.12	0.22	LH	LH	ES, LH	Y, Y
10	0.32	0.28	1.14	0.35	BH	LH	ES, LH	N, Y
11	0.22	0.15	1.47	0.11	BH	BH	ES, LH	N, N
12	0.20	0.19	1.05	0.04	BH	BH	ES, LH	N, N
13	0.60	0.40	1.5	0.40	BH	LH	ES, LH	N, Y
14	0.40	0.20	2	0.55	LH	LH	ES, LH	Y, Y
15	0.44	0.25	1.76	0.23	BH	LH	Wada, BH	Y, N
16	0.02	−0.14	7	0.51	LH	LH	Wada, LH	Y, Y
17	0.45	0.21	2.14	0.35	LH	LH	ES, LH	Y, Y
18	0.72	0.29	2.48	0.83	LH	LH	Wada, LH	Y, Y
19	0.37	0.34	1.09	0.16	BH	BH	Wada, LH	N, N
20	1.13	0.50	2.26	0.40	LH	LH	Wada, LH	Y, Y
21	0.51	0.58	1.14	0.31	BH	LH	Wada, BH	Y, N
22	0.43	1.05	2.44	−0.42	RH	RH	Wada, LH	N, N
23	0.47	0.23	2.04	0.79	LH	LH	PS, LH	Y, Y
24	0.54	0.26	2.08	0.59	LH	LH	Wada, LH	Y, Y
25	0.45	0.30	1.5	0.56	BH	LH	Wada, LH	N, Y
26	0.28	0.23	1.22	−0.15	BH	BH	Wada, LH	N, N

cordant prediction in eight out of eight patients, 100%), followed by the WSC task (eight of ten patients, 80%), the LNL task (17 of 26 patients, 65%), and the PA task (five of eight patients, 63%). Moreover, in contrast to the influence of thresholds that we observed on percent agreement when using the AM method (particularly for more conservative thresholds), percent agreement was not influenced by using a lower (i.e., 0.15, 37 of 52 patients or 71% agreement) or higher threshold (i.e., 0.25, 36 of 52 or 69% agreement). Tables 3, 4, 5 summarize the lateralization results obtained in each patient with the two different statistical methods (AM and LI) for

Table 4 Hemispheric language dominance as determined by a WSC task [*AM LH* average signal intensity in the left hemisphere, *AM RH* average signal intensity in the right hemisphere, *AM ratio* ratio between the higher AM hemispheric value and the lower AM hemispheric value; *HLD fMRI AM* hemispheric language dominance as assessed by the AM method of quantifying fMRI results, *HLD fMRI LI* hemispheric language dominance as assessed by the lateralization indices method; *"True" HLD* hemispheric language dominance as determined by gold standard measures including Wada test, ES or PS; *LH* left dominant, *RH* right dominant, *BH* bilateral representation of language; *Y* concordant fMRI assessment of HLD (listed separately for the AM and the LI methods) with gold standard measures of HLD, *N* discordant assessment of HLD]

Patient number	AM LH	AM RH	AM ratio	LI	HLD fMRI AM	HLD fMRI LI	"True" HLD	Concordance
2	0.32	0.09	3.56	0.97	LH	LH	PS, LH	Y, Y
3	0.14	0.30	2.14	0.33	RH	LH	PS, LH	N, Y
4	0.16	0.55	3.44	−0.86	RH	RH	ES, RH	Y, Y
5	0.37	0.43	1.16	0.15	BH	BH	ES, LH	N, N
6	0.46	0.08	5.75	1	LH	LH	ES, LH	Y, Y
9	0.13	0.22	1.69	0.41	BH	LH	ES, LH	N, Y
10	0.26	0.11	2.36	0.65	LH	LH	ES, LH	Y, Y
11	0.37	0.30	1.23	0.39	BH	LH	ES, LH	N, Y
17	0.55	0.23	2.39	0.31	LH	LH	ES, LH	Y, Y
27	0.15	0.68	4.53	0	RH	BH	ES, LH	N, N

Table 5 Hemispheric language dominance as determined by SA and PA task [*AM LH* average signal intensity in the left hemisphere, *AM RH* average signal intensity in the right hemisphere, *AM ratio* ratio between the higher AM hemispheric value and the lower AM hemispheric value; *HLD fMRI AM* hemispheric language dominance as assessed by the AM method of quantifying fMRI results, *HLD fMRI LI* hemispheric language dominance as assessed by the lateralization indices method; *"True" HLD* hemispheric language dominance as determined by gold standard measures including Wada test, ES or PS; *LH* left dominant, *RH* right dominant, *BH* bilateral representation of language; *Y* concordant fMRI assessment of HLD (listed separately for the AM and the LI methods) with gold standard measures of HLD, *N* discordant assessment of HLD]

Patient number	Task	AM LH	AM RH	AM ratio	LI	HLD fMRI AM	HLD fMRI LI	"True" HLD	Concordance
28	PA	0.42	0.49	1.17	0.43	BH	LH	ES, LH	N, Y
29	PA	0.41	0.39	1.05	0.05	BH	BH	ES, LH	N, N
30	PA	0.13	0.03	4.33	0.11	LH	BH	Wada, RH	N, N
31	PA	0.06	0.02	3	0.18	LH	BH	Wada, BH	N, Y
32	PA	0.04	0.03	1.33	0.27	BH	LH	Wada, LH	N, Y
33	PA	0.21	0.23	1.09	0.68	BH	LH	ES, LH	N, Y
34	PA	0.25	0.06	4.17	0.48	LH	LH	ES, LH	Y, Y
35	PA	0.14	0.05	2.8	0.02	LH	BH	ES, LH	Y, N
28	SA	0.09	0.09	1	0.38	BH	LH	ES, LH	N, Y
29	SA	0.50	0.71	1.42	0.39	BH	LH	ES, LH	N, Y
30	SA	0.23	0.38	1.65	−0.33	BH	RH	Wada, RH	N, Y
31	SA	0.06	0.01	6	0.17	LH	BH	Wada, BH	N, Y
32	SA	0.51	0.78	1.53	0.50	BH	LH	Wada, LH	N, Y
33	SA	0.46	0.61	1.33	0.31	BH	LH	ES, LH	N, Y
34	SA	0.40	0.09	4.44	0.30	LH	LH	ES, LH	Y, Y
35	SA	0.61	0.54	1.13	0.49	BH	LH	ES, LH	N, Y

quantifying the fMRI data collected using the LNL task, the WSC task, and the SA and PA tasks, respectively. For comparison, Tables 3, 4 and 5 also contain the true lateralization results obtained from each patient based on gold standard measures of HLD, including Wada, ES, and PS. The functional maps shown in Fig. 2 illustrate both a concordant case (Fig. 2a, patient 4) and a discordant case (Fig. 2b, patient 22). Specifically, for patient 4, using the LI method with the standard threshold of 0.20, HLD was right as revealed by both the LNL and WSC tasks (Tables 3 and 4, respectively). Moreover, these fMRI results were confirmed by the ES result. In contrast, patient 22 represents a case where the LNL task yielded an fMRI result that was not in agreement with the results of the Wada test (using either the AM or the LI method of assessing HLD as shown in Table 3 with thresholds of 2 and 0.20, respectively). In this patient, the fMRI indicated a right dominance for language, whereas the Wada test showed the opposite pattern and indicated a left dominance.

Discussion

The present study confirms that concordant fMRI predictions of HDL were roughly twice as frequent when using LI (73%) compared with AM (40%) (aim 1). With respect to the second aim, the SA task [12] appeared to be the most compatible with established methods of determining HLD. The SA task correctly predicted HLD in 100% of the epileptic patients who performed this task.

Using the AM method, the idea is that commonly activated regions could be those critical for performing that task [17], including frontal and temporoparietal regions, as indicated by the *red and blue arrows* in Fig. 1, respectively. As shown in Table 2, using standard thresholds (i.e., 0.20 for LI and 2 for AM), whether one collapses across all four tasks used in the present study or considers each task separately, the LI method of quantifying fMRI results predicted the HLD better than the AM method. There are at least three possible explanations for the lack of concordance between HLD as determined by using the AM method gold standard measures. First, the average signal intensity of activated pixels within the predefined regions of interest may not have been robust enough to reveal differences between homologous regions across the two hemispheres. Second, even though the patients' fMRI data were used to predefine the regions of interest for each task, there still may have been a spatial mismatch between the commonly activated critical regions for language and the precise location of the underlying regions in the individual patients that contributed to this group average. Third, the criterion used for establishing HLD was that the dominant hemisphere must have a signal level of at least twice that of its non-dominant homologue. As noted earlier, this criterion is somewhat arbitrary and hence, it is possible that a smaller difference in signal levels could still indicate a lateralization of language. However, this third explanation was not confirmed by decreasing the AM threshold from 2 to 1.5 because percent agreement only increased by 4% when using the more liberal threshold (i.e., percent agreement increased

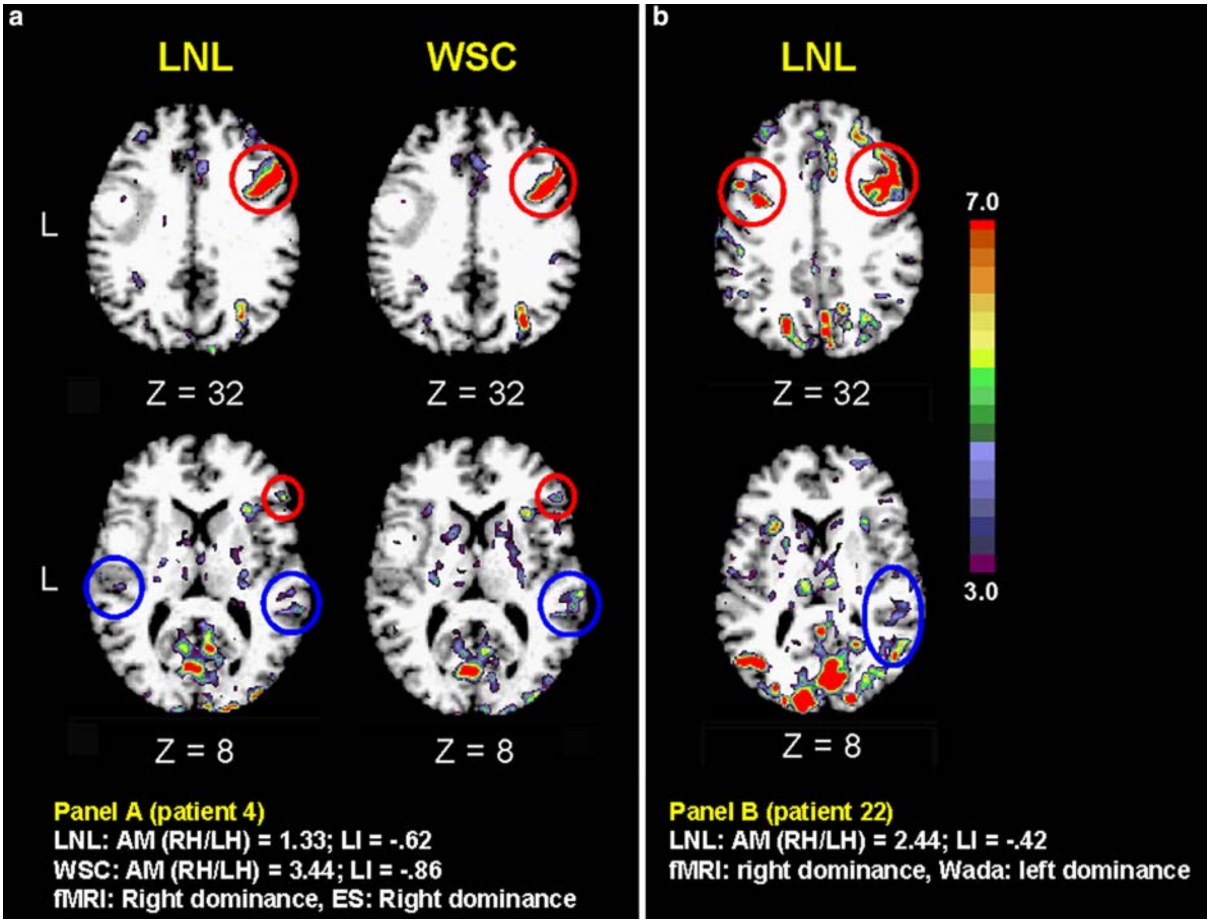

Fig. 2 **a** Representative functional maps obtained in patient 4 during a LNL task and a WSC task relative to fixation. Language was right lateralized according to these tasks (i.e., LI− b Representative functional maps obtained in patient 22 during a LNL task relative to fixation. Language was right lateralized according to this task (i.e., LI−left), and in all cases, task activity is relative to fixation baseline

from 40 to 44% for the 52 fMRI examinations conducted in the present study). Even so, it is noteworthy that percent agreement decreased dramatically to 21% when a more conservative threshold of 2.5 was used to quantify HLD with the AM method. In contrast, concordance with gold standard measures of HLD were not influenced by decreasing (0.15, 71% agreement) or increasing (0.25, 69% agreement) the standard threshold (0.20, 73% agreement) with the LI method.

One obvious difference between the LI and AM methods is that the language dominance quantification is based on the specific number of activated pixels in each patient in predefined regions of interest in frontal and temporoparietal cortex rather than a group average in similar regions. Intuitively, it makes sense that a single patient's fMRI data does a better job of predicting his/her own hemispheric organisation of language as measured by Wada, ES or PS because there is a *one-to-one* unit of analysis. Both the LI and the AM method do not directly compare fMRI results across the two hemispheres to reveal if the observed inter-hemispheric differences are statistically significant. Some studies statistically compared the variation of the signal intensity variation between right and left regions and found that their method provided better agreement with gold standard methods of assessing HLD than the LI method (see [9, 14] for additional methods of quantifying language dominance using fMRI data). Although there are several methods used for determining the dominant hemisphere for language, currently, there is no clear consensus about which method would be the best for quantifying the lateralization expressed by fMRI results. The LI method continues to be widely used by the scientific community in order to determine the HLD in patients. We are aware that a statistical inter-hemispheric comparison of the activation would be more robust and safe. However, for the purposes of surgical planning, neurosurgeons may not need a value that indicates whether the hemispheric dominance of language expressed in a single patient's fMRI data is statistically significant. Rather, consistent with images that can be derived with the LI method for quantifying fMRI data as outlined here, neurosurgeons may prefer to consider fMRI results that qualitatively capture the constellation of neural regions that are robustly active during a language task on a patient-by-patient basis. In this light, although fMRI cannot currently replace gold standard measures of language assessment, fMRI may still be a very useful clinical tool because it is less invasive, relatively inexpensive, and depending on the language task used, provides converging evidence with Wada or ES regarding the underlying organisation of regions correlated with language in the brain [3].

The SA task was the most effective fMRI procedure for assessing HLD because it was concordant with the gold standard in 100% of the epileptic patients that performed this task. Specifically, eight out of eight patients were concordant when using the SA task in conjunction with the LI method, although it is noteworthy that the concordance was much lower when the results of the SA task for these same eight patients were quantified using the AM method (i.e., only one of eight patients or 13% were concordant). This particular SA task was first used by McDermott et al. [12] in healthy young adults and has been used more recently in combination with the PA task by Baciu et al. [3] to preoperatively identify language regions successfully in a patient with frontal lobe epilepsy (see patient 31, Table 5, in the present study, who had a bilateral representation of language according to Wada and fMRI results). As noted by McDermott et al. [12], two important advantages may be gained by using their SA and PA tasks in conjunction with fMRI to assess language function. First, it is possible to elicit statistically significant activation at the level of individual subjects as well as at the group level with these tasks (e.g., see the group averages for the SA and PA tasks in Fig. 1 of the present study and in Figs. 1 and 2 of McDermott et al. [12]). Second, these patterns of activation hold for underlying language networks in both frontal and temporoparietal regions (e.g., see McDermott et al., Fig. 3, for robust functional maps that can be derived by directly comparing semantic and phonological processing using the SA and PA tasks for individual subjects, respectively). As shown in Table 2, using the LI method, the LNL task was only concordant with established measures of HLD for 65% of the patients even though the LNL task also elicited activity in frontal and temporal cortex (Fig. 1).

Group level analyses indicated that the PA and the WSC tasks activated frontal language regions, but not the temporoparietal regions also activated by the SA and LNL tasks (Fig. 1). Few studies have used phonological tasks in epileptic patients [2, 4]. In general, these studies showed that a phonological task has the potential advantage of activating the peri-sylvian regions (i.e., Broca's and Wernicke's area) in individual patients where the majority of essential regions for language are situated as confirmed by ES [15]. These studies also showed that the activation was strongly lateralized within the dominant hemisphere. In the PA task used in the present study, the lateralization was not as strong as for other phonological tasks like rhyme detection. This may explain in part why the concordance of the PA task with gold standard measures of assessing HLD was only 63% using the LI method, even though the HLD of the same eight patients was correctly predicted using the SA task. In other words, despite the small sample size ($n=8$) we used to quantify HLD with the SA and PA tasks, the fMRI data for this same set of eight patients is evidently quite stable because we observed an effect of task on percent agreement using the LI method (with 100% agreement with gold standard measures for the SA task exceeding reasonable expectations of at least 90% agreement). The word generation task is widely used in epileptic patients for assessing language dominance [4, 9, 11, 23, 27]. In the majority of these studies, this task robustly activates the inferior frontal cortex (Broca's area, BA 44, 45, 47) but not the temporoparietal regions. In our study, the agreement between fMRI using WSC (based on the LI method) and gold standard measures of HLD was 80%. Although this result was most likely driven by the frontal activation elicited by this task, it is important to note that temporoparietal regions were active in some individual patients.

Conclusions

Within the scope of the present study, it is intriguing that a relatively straightforward, easy to implement SA task can provide (1) convergent evidence with gold standard measures when assessing HLD in individual epileptic patients with LI, and (2) supplemental information about the relative location of potential critical language regions or networks in the brain (e.g., frontal vs temporal). The preliminary results from the present study suggest that the lexical processing task may yield robust functional maps that can be used qualitatively to identify language regions on a patient-by-patient basis in individual neurosurgical patients. Although the sample size was small ($n=8$), the semantic task was much more lateralising than the other tasks, while it provided 100% agreement with the gold standard methods. Future research may be able to offer additional constraints above and beyond the descriptive statistical comparisons we have provided for the two different methods (AM vs LI) and four different language tasks (LNL vs WSC vs SA vs PA) we used to qualitatively assess laterality in epileptic patients. Hence, the results of the present study should be considered preliminary but promising with regard to the ability of fMRI non-invasively to determine hemispheric language dominance in neurosurgical patients.

Acknowledgements We thank Steve Petersen and Brad Schlaggar for helpful discussion throughout this project, Kristina Visscher and Christine Kang for help with the analyses and for comments on an earlier version of this manuscript, and Kate O'Brien and Margaret Sheridan for help with data collection. We also thank Abraham Snyder and Mark McAvoy for help with data processing and analysis. Support for this research project was provided by the McDonnell Center for Higher Brain Function and NIH/NINDS Grant NS41272.

References

1. Adcock JE, Wise RG, Oxbury JM et al (2003) Quantitative fMRI assessment of the differences in lateralization of language-related brain activation in patients with temporal lobe epilepsy. Neuroimage 18:423–438
2. Baciu M, Kahane P, Minotti L et al (2001) Functional MRI assessment of the hemispheric predominance for language in epileptic patients using a simple rhyme detection task. Epileptic Disord 3:117–124
3. Baciu MV, Watson JM, McDermott KB et al (2003) Functional MRI reveals an interhemispheric dissociation of frontal and temporal language regions in a patient with focal epilepsy. Epilepsy Behav 4:776–780
4. Bahn MM, Lin W, Silbergeld DL et al (1997) Localization of language cortices by functional MR imaging compared with intracarotid amobarbital hemispheric sedation. AJR Am J Roentgenol 169:575–579
5. Binder JR, Swanson SJ, Hammeke TA et al (1996) Determination of language dominance using functional MRI: a comparison with the Wada test. Neurology 46:978–984
6. Desmond JE, Sum JM, Wagner AD et al (1995) Functional MRI measurement of language lateralization in Wada-tested patients. Brain 118:1411–1419
7. FitzGerald DB, Cosgrove GR, Ronner S et al (1997) Location of language in the cortex: a comparison between functional MR imaging and electrocortical stimulation. Am J Neuroradiol 18:1529–1539
8. Gaillard WD, Balsamo L, Xu B et al (2002) Language dominance in partial epilepsy patients identified with an fMRI reading task. Neurology 59:256–265
9. Hertz-Pannier L, Gaillard WD, Mott SH et al (1997) Noninvasive assessment of language dominance in children and adolescents with functional MRI: a preliminary study. Neurology 48:1003–1012
10. Lee GP, Park YD, Hempel A et al (2002) Effect of Wada methodology in predicting lateralized memory impairment in pediatric epilepsy surgery candidates. Epilepsy Behav 3:439–447
11. Lehericy S, Cohen L, Bazin B et al (2000) Functional MR evaluation of temporal and frontal language dominance compared with the Wada test. Neurology 54:1625–1633
12. McDermott KB, Petersen SE, Watson JM et al (2003) A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. Neuropsychologia 41:293–303
13. Miezin FM, Maccotta L, Ollinger JM et al (2000) Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage 11:735–759
14. Nagata SI, Uchimura K, Hirakawa I et al (2001) Method for quantitatively evaluating the lateralization of linguistic function using functional MR imaging. AJNR Am J Neuroradiol 22:985–991
15. Ojemann G, Ojemann J, Lettich E et al (1989) Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. J Neurosurg 71:316–326
16. Ruge MI, Victor J, Hosain S et al (1999) Concordance between functional magnetic resonance imaging and intraoperative language mapping. Stereotact Funct Neurosurg 72:95–102
17. Rutten GJ, Ramsey NF, van Rijen PC et al (2002) Development of a functional magnetic resonance imaging protocol for intraoperative localization of critical temporoparietal language areas. Ann Neurol 51:350–360
18. Rutten GJ, Ramsey NF, van Rijen PC et al (2002) Reproducibility of fMRI-determined language lateralization in individual subjects. Brain Lang 80:421–437
19. Sabbah P, Chassoux F, Leveque C et al (2003) Functional MR imaging in assessment of language dominance in epileptic patients. Neuroimage 18:460–467
20. Spreer J, Arnold S, Quiske A et al (2002) Determination of hemisphere dominance for language: comparison of frontal and temporal fMRI activation with intracarotid amytal testing. Neuroradiology 44:467–474
21. Springer JA, Binder JR, Hammeke TA et al (1999) Language dominance in neurologically normal and epilepsy subjects: a functional MRI study. Brain 122:2033–2046
22. Talairach J and Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Thieme, New York
23. van der Kallen BF, Morris GL, Yetkin FZ et al (1998) Hemispheric language dominance studied with functional MR: preliminary study in healthy volunteers and patients with epilepsy. AJNR Am J Neuroradiol 19:73–77
24. Wada J, Rasmussen T (1958) Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance. J Neurosurg 17:266–282
25. Woermann FG, Jokeit H, Luerding R et al (2003) Language lateralization by Wada test and fMRI in 100 patients with epilepsy. Neurology 61:699–701
26. Yetkin FZ, Mueller WM, Morris GL et al (1997) Functional MR activation correlated with intraoperative cortical mapping. AJNR Am J Neuroradiol 18:1311–1315
27. Yetkin FZ, Swanson S, Fischer M et al (1998) Functional MR of frontal lobe activation: comparison with Wada language results. AJNR Am J Neuroradiol 19:1095–1098