



**Get Clarity On Generics**

Cost-Effective CT & MRI Contrast Agents

**FRESENIUS  
KABI**

**WATCH VIDEO**

**AJNR**

This information is current as  
of August 19, 2025.

## **Application of a Computerized Language Lateralization Index from fMRI by a Group of Clinical Neuroradiologists**

S.E. Jones, S.Y. Mahmoud, J. Gonzalez-Martinez, D.S.  
Lockwood, D. Moon, A.S. Smith, T.W. Stultz, A.L. Tievsky  
and M.D. Phillips

*AJNR Am J Neuroradiol* 2013, 34 (3) 564-569

doi: <https://doi.org/10.3174/ajnr.A3271>

<http://www.ajnr.org/content/34/3/564>

# Application of a Computerized Language Lateralization Index from fMRI by a Group of Clinical Neuroradiologists

S.E. Jones, S.Y. Mahmoud, J. Gonzalez-Martinez, D.S. Lockwood, D. Moon, A.S. Smith, T.W. Stultz, A.L. Tievsky, and M.D. Phillips

## ABSTRACT

**BACKGROUND AND PURPOSE:** Deriving accurate language lateralization from fMRI studies in the clinical context can be difficult, with 10%–20% incorrect conclusions. Most interpretations are qualitative, performed by neuroimaging experts. Quantitative lateralization has been widely described but with little implementation in the clinical setting and is disadvantaged by the use of arbitrary threshold techniques. We investigated the application and utility of a nonthreshold CLI, in a clinical setting, as applied by a group of practicing neuroradiologists.

**MATERIALS AND METHODS:** Twenty-two patients with known language lateralization (11 left and 11 nonleft dominant) had their images reviewed by 8 neuroradiologists in 2 settings, all randomized, once by using a CLI and once without using a CLI. For each review, neuroradiologists recorded their impressions of lateralization for each language sequence, the overall lateralization conclusion, their impression of scan quality and noise, and the subjective confidence in their conclusion.

**RESULTS:** The inter-rater  $\kappa$  for lateralization was 0.64, which increased to 0.70 with the use of CLI. The group accuracy of overall lateralization was 78%, which increased to 81% with the use of a CLI. Using a CLI removed 2 instances of significant errors, with a neuroradiologist's impression of left lateralization in a patient with known right lateralization. Using a CLI had no effect on examinations with conclusions formed with either high confidence or no confidence.

**CONCLUSIONS:** Although the overall clinical benefit of a CLI is modest, the most significant impact is to reduce the most harmful misclassification errors, particularly in fMRI examinations that are suboptimal.

**ABBREVIATION:** CLI = computerized lateralization index

fMRI has become a standard presurgical mapping examination for the preoperative assessment of brain tumor and epilepsy resections, with important advantages of noninvasiveness, safety, and ease of performance, particularly compared with the invasive alternative of the Wada test.<sup>1</sup> fMRI is now entering its third decade since its first demonstrations in the early 1990s<sup>2–4</sup> and has moved from a purely research tool toward the clinical arena. Yet within the clinical arena, most fMRI examinations tend to be read by subspecialists within neuroradiology, rather than general neuroradiologists. For example, often only 1 or several neuroradiologists in a group have fMRI experience. While 1 major reason may

be the seemingly complex nature of fMRI, a more important and practical reason is the relatively low caseload compared with common structural MRI. Nevertheless, the annual number of clinical fMRI examinations is slowly growing, particularly as applied to presurgical planning, and current trends suggest the future will require more specialists to have fMRI experience.

Therefore, there is a growing clinical need to expand the capabilities of general neuroradiologists to encompass fMRI. A difficulty of this expansion is the potential for increased variability of interpretation, particularly for practitioners new to the procedure. Thus, a major motivation for this article is to investigate and enhance the uniform application of fMRI by an entire group practice of clinical neuroradiologists. In this regard, the specific fMRI application studied is language lateralization, currently the major clinical application of fMRI used in presurgical planning, particularly in the disease state of intractable epilepsy. The measure used to assess the uniformity of lateralization interpretation is an automated CLI, which was recently published by our group and provides a robust and unbiased preoperative assessment.<sup>5</sup> Not

Received March 15, 2012; accepted after revision June 16.

From the Divisions of Neuroradiology (S.E.J., S.Y.M., D.S.L., D.M., A.S.S., T.W.S., A.L.T., M.D.P.), and Neurosurgery (J.G.-M.), Imaging Institute, Cleveland Clinic, Cleveland Ohio.

Paper previously presented in part at: Annual Meeting of the American Society of Neuroradiology, May 16–21, 2009; Vancouver, British Columbia, Canada.

Please address correspondence to Stephen E. Jones, PhD, MD, 9500 Euclid Ave, U 15, Cleveland, Ohio 44195; e-mail: Jones19@ccf.org

<http://dx.doi.org/10.3174/ajnr.A3271>

only can this measure assess the uniformity of practice, but it can aid in training new practitioners. In summary, we hypothesize that a CLI can aid the fMRI capability of a group practice of neuroradiologists by increasing the uniformity of fMRI interpretations and minimizing errors.

## **MATERIALS AND METHODS**

### **Subject Selection**

Patients were selected from the data base of all fMRI studies conducted at our institution by using 3T Trio MR imaging (software version VB15; Siemens, Erlangen, Germany) between August 23, 2005, and October 14, 2009. This list was cross-referenced to the electronic medical record (Epic Systems; Verona, Wisconsin) to identify a subset of patients who also underwent a corroborating language lateralization examination (either a Wada test, intraoperative mapping, or subdural grid examination). Relevant clinical information from the electronic medical record focused on the final clinical conclusion from a non-fMRI lateralization examination (categorized as left, bilateral, or right), whose lateralization defined the criterion standard for this investigation.

From this set of patients, 2 further subsets were selected, each being equal in number, with the first group comprising solely left-language dominance and the second group comprising solely nonleft dominance (ie, right or bilateral dominance). The patients for each group were selected randomly from the prior set of patients. Finally, the 2 groups of patients were combined and randomly sorted, thereby producing an equal but random admixture of left and non-left-dominant patients. The purpose of this mixture was to minimize any pretest bias about lateralization on the basis of presumed incidence.

The Cleveland Clinic institutional review board approval was obtained for all studies, and Health Insurance Portability and Accountability Act policies were strictly followed.

### **Study Design**

For each patient, the fMRI study was reviewed twice by each neuroradiologist, at different time points, once with and once without the aid of a sheet summarizing results from a CLI (described below). Collectively, the review of all patients was divided into 2 sessions, with each patient reviewed once in each session. The first session reviewed every patient in a randomized order of admixed left and nonleft dominance. A subset of these patients was reviewed in conjunction with a CLI, with the total number and selection of patients being random. The second session was the complement to the first session—that is, CLI sheets were now included for those patients who were reviewed without CLI in the first session and vice versa. Thus, after both sessions, each patient would have been reviewed twice, once with and once without the CLI, with all orderings random.

All 11 neuroradiologists in our department were invited to participate. Each participating neuroradiologist reviewed the 2 sessions of fMRI images on a Leonardo workstation (Siemens), by using the blood oxygen level–dependent task card. All neuroradiologists were blinded to the known lateralization results and were told that there was some mixture of left and nonleft dominance whose fraction did not reflect the expected incidence. The

actual fraction used (50%) was not revealed to minimize any incidence bias.

During review of each patient's study, the neuroradiologist was provided a score sheet evaluating the following factors from the statistical maps of each functional sequence: image quality as measured on a 4-point scale (unreadable, minimal, adequate, and excellent) and the degree of lateralization as measured on a 5-point scale (exclusively left, most left, bilateral, most right, and exclusively right). Final assessments were then provided by quantitatively combining all 3 language sequences, specifically measuring overall study quality (same 4-point scale), overall lateralization (same 5-point scale), and a 4-point scale evaluating the subjective degree of the reader's confidence in the above conclusions (none, marginal, adequate, and very confident). In addition to all language sequences, a bilateral finger-tapping motor sequence was evaluated as a control sequence because bilateral activation was expected.

The CLI method used was developed and reported earlier,<sup>5</sup> and the reader is referred to that reference for a complete description of the method. Briefly, the method uses a hemibrain histogram analysis, specifically computing a histogram from *t*-score maps for all parenchymal voxels in each hemisphere. No region-of-interest analysis within a hemisphere was used. Statistically, functional activation in the brain increases the number of voxels with high *t*-scores, which is manifest as increases in the shape of the tail of the histogram distribution. Functional activation for each hemisphere is derived by quantifying the magnitude of changes in the histogram tail, and a laterality index was computed from the values in both hemispheres. This method is inherently independent of the selection of a threshold, which is commonly used in the literature to compute lateralization indices and whose arbitrary selection can often bias results. This method was applied to a large set of patients with known lateralization, thereby forming a library to compare future CLI measures. The criterion standard for language lateralization was taken to be either the Wada test<sup>1</sup> or direct electrophysiologic lateralization as measured by intraoperative electrode stimulation or postoperative cortical stimulation of retained intracranial electrodes.

The final product of this method, as applied to a new patient with unknown lateralization, is a summary sheet that can accompany the neuroradiologist's review, comparing the new patient's CLI measures with the library of prior patients with known lateralization. The summary sheet includes comparison plots for each of the 4 paradigms (1 motor and 3 language), in addition to combined language. Last, 3 numeric probabilities are presented assessing the chance that a new patient's language could be lateralized left, bilateral, or right, as determined by comparison with the library of patients. This analysis was performed by using in-house software developed by using IDL (Interactive Data Language), Version 6.3 (ITT Visual Information Solutions, Boulder, Colorado).

Results from all participating neuroradiologists were combined into 2 groups, based on the use or nonuse of CLI, thereby permitting group comparisons. Lateralization was categorized in 2 ways: It was trichotomized into left, bilateral, and right dominance; and to increase power, it was dichotomized into left and nonleft dominance, where nonleft was defined as both bilateral

and right. This latter categorization is clinically important because often the neurosurgeon's concern is the possibility of significant language in the right hemisphere. Last, the data were presented as  $2 \times 2$  or  $3 \times 3$  contingency tables, and statistical assessments of significance were performed by using the Fisher exact test with 2-sided *P* values. Comparisons of ordinal data used the Mann-Whitney test. Interobserver reproducibility used Cohen  $\kappa$  calculation.

### MR Imaging Acquisition

The standard clinical fMRI examination at our institution for presurgical planning begins with an anatomic sequence (T1-weighted 3D axial magnetization-prepared rapid acquisition of gradient echo, one hundred twenty 0.94-mm-thick sections, TE/TR/TI/flip angle = 1.7/900/1900 ms/7°,  $128 \times 256$  matrix,  $256 \times 256$  mm FOV, receive bandwidth = 125.44 kHz), followed by 4 functional paradigms by using an EPI sequence (160 volumes of 3.8-mm-thick axial sections, by using a prospective motion-controlled gradient recalled-echo, echo-planar acquisition with TE/TR/flip angle = 29/2000 ms/90°, matrix =  $64 \times 64$ ,  $256 \times 256$  mm FOV, receive bandwidth = 125 kHz). Images were prospectively motion-corrected by using PACE software (VB15; Siemens).

Each functional paradigm was performed by using the same block design comprising 4 blocks, with each block containing 16 volume acquisitions during rest alternating with 16 volume acquisitions during a task. Because the TR was 2 seconds, it represented 32 seconds of rest alternating with 32 seconds of task. Sixteen null blocks were acquired both before and after the cycles of tasks/rests. Thus, the entire acquisition contained 160 volume acquisitions, for a total acquisition time of 5 minutes 20 seconds. Occasionally, other additional clinical sequences were performed—for example fluid-attenuated inversion recovery, post-contrast T1, and diffusion tensor imaging. Sequences were repeated at the discretion of the specialized fMRI technologist for reasons such as excessive motion, failure to adequately perform tasks, and equipment failure. All imaging planes were oriented parallel to the anterior/posterior commissure line.

The behavioral paradigms composing our clinical fMRI were 1 motor and 3 language tasks. The motor task comprised 4 cycles of simultaneous bilateral finger tapping and rest. The 3 language tasks were covert word generation, rhyming decision, and passive listening. During covert word generation, patients viewed either a single letter (activation phase) or a nonsense symbol (control phase). During activation phases, patients were asked to covertly think of any words beginning with the visualized letter, at a comfortably rapid pace. During the control phase, patients were asked to simply view the symbol and to minimize other unrelated mental activity. For the rhyming task, patients were shown word pairs every 4 seconds during the activation phase; then, they were asked to press 1 of 2 buttons, depending on whether the words rhymed. Similarly, during the control phase, symbol pairs of matching or nonmatching stick figures were shown, and patients were asked to press 1 of 2 buttons depending on whether the symbols matched. For the passive listening task, patients listened through headphones to 4 cycles of recorded audio segments read from a familiar story, with each cycle including a segment read forward for 32

**Table 1: Inter-rater agreement of 8 neuroradiologists reviewing 22 language fMRI examinations, each with and without the use of CLI<sup>a</sup>**

	Mean $\kappa$	
	Without CLI	With CLI
Overall lateralization	0.64	0.70
Overall quality	0.11	0.08
Overall noise	0.05	0.07
Overall confidence	0.12	0.03

<sup>a</sup> Measured parameters include overall lateralization, image quality, overall image noise, and subjective confidence of the lateralization conclusion. Lateralization was dichotomized for either left or nonleft.

**Table 2: Comparison of a neuroradiologist's subjective impression of language lateralization from fMRI compared with known lateralization, conducted without (top half) and with (bottom half) the use of CLI<sup>a</sup>**

Known Lateralization	Neuroradiologist's Subjective Impression	
	Left	Nonleft
Without CLI		
Left	77 (0.44)	11 (0.06)
Nonleft	28 (0.16)	60 (0.34)
With CLI		
Left	81 (0.46)	7 (0.04)
Nonleft	27 (0.15)	61 (0.35)

<sup>a</sup> The numbers refer to the total assessments obtained after review of all 8 neuroradiologists with all 22 studies, for a total of 176 comparisons. The number in parentheses is the corresponding fraction. Lateralization was dichotomized into 2 categories of either left or nonleft lateralization, with the latter group combining bilateral and right lateralization (ie, any examination with some component of right-sided activation).

seconds followed by the same segment played backward for 32 seconds. Afterward, patients were asked 4 simple questions from the story to assess their degree of attention.

## RESULTS

A total of 22 fMRI studies with known language lateralization were included in this study, of which 11 were classified as nonleft (right or bilateral) and 11 as left-language dominant. There were 10 males and 12 females, with a mean age of  $28 \pm 15$  years (range, 6–68 years). Regarding disease states, 13 patients had epilepsy and 9 had tumors. A total of 8 neuroradiologists from a division of 11 staff participated, producing a total of 176 fMRI reviews without a CLI and 176 reviews with a CLI. There was a wide range of experience of the neuroradiologists regarding fMRI, ranging from 12 years to a newly trained staff member with no experience.

The  $\kappa$  measure of inter-rater agreement for dichotomized language-lateralization categories (left and nonleft) was 0.64 without the use of a CLI, and it increased to 0.70 with the use of a CLI (Table 1). With trichotomized language lateralization categories (left, bilateral, and right), the inter-rater  $\kappa$  was 0.49 without the use of CLI, and it increased to 0.59 with the use of a CLI. There was poor inter-rater agreement ( $<0.4$ ) for subjective assessments of image quality, image noise, and confidence of lateralization conclusion.

Table 2 shows the accuracy of fMRI language lateralization from a comparison of neuroradiologists' impressions of language lateralization compared with known lateralization, for the case of dichotomized lateralization. Without using a CLI, the accuracy was 78% ( $P < 10^{-6}$ ); the left-lateralization sensitivity was 88% and the nonleft lateralization sensitivity was 68%. When a CLI was used, the accuracy increased to 81% ( $P < 10^{-6}$ ).

**Table 3: Neuroradiologist's subjective impression of language lateralization, with trichotomized lateralization categories of left, bilateral, and right<sup>a</sup>**

Known Lateralization	Neuroradiologist's Subjective Impression		
	Left	Bilateral	Right
Without CLI			
Left	77 (0.44)	11 (0.06)	0 (0.00)
Bilateral	26 (0.15)	12 (0.07)	10 (0.06)
Right	2 (0.01)	21 (0.12)	17 (0.10)
With CLI			
Left	81 (0.46)	7 (0.04)	0 (0.00)
Bilateral	27 (0.15)	11 (0.06)	10 (0.06)
Right	0 (0.00)	25 (0.14)	15 (0.09)

<sup>a</sup> Similar to Table 2. The top half displays results obtained without using CLI, while the bottom half displays results obtained using CLI. The number in parentheses is the corresponding fraction.

Table 3 is similar to Table 2, except for trichotomized lateralization. Without using a CLI, the accuracy was 61% ( $P < 10^{-6}$ ). The left-lateralization sensitivity remained 88%, the bilateral sensitivity was 25%, and the right lateralization sensitivity was 43%. When a CLI was used, the accuracy was unchanged at 61% ( $P < 10^{-6}$ ). Although the overall changes due to a CLI are modest, more important are specific changes; usage of the CLI corrected the 2 cases with the most significant error of incorrectly concluding a right-lateralized patient as left-lateralized. The cause of these significant errors was investigated, without any evidence of a systematic cause. The 2 errors occurred on different patients by different readers, 1 of whom assessed a large value of image noise that was discordant from the remaining readers. Thus, the errors were most likely sporadic perceptual misinterpretations.

Of the 176 language-lateralization assessments made by all reviewers over all studies, 52 conclusions (30%) were different with the use of a CLI. Regarding lateralization of motor function, assumed to be bilateral and serving as a control, 27 conclusions (15%) were changed; this finding was significantly different compared with that of the language studies ( $P = .002$ ).

Use of the CLI increased the confidence of lateralization conclusion. Without the CLI, the fractions of impressions with strong, adequate, marginal, and no confidence were 17%, 45%, 31%, and 6%, respectively. The use of CLI changed these values to 32%, 48%, 18%, and 2%. The radiologist's subjective assessment of image quality and noise was unaffected by the use of a CLI.

A subgroup analysis of the accuracy of lateralization versus subjective confidence found that CLI caused no change when the confidence was either very high or none. Thus the effect of CLI mainly benefitted those studies with adequate or marginal subjective confidence.

There was a marginal variation of results with respect to the reader's experience. Specifically, the newly trained reader with no prior experience had both lower average confidence (1.52) and quality assessment (1.38) compared with the 7 remaining readers with more experience (confidence,  $1.98 \pm 0.33$ ; quality assessment,  $1.73 \pm 0.14$ ). However, the assessment of image noise and lateralization showed no difference. The 7 more experienced readers showed no other reliable variation with experience.

## DISCUSSION

We sought to determine the accuracy and inter-rater reliability of subjective language lateralization from a group of clinical neuro-

radiologists, and we examined how these conclusions changed with the use of a CLI during review of fMRI statistical maps. We also included an analysis of the neuroradiologist's subjective level of confidence in reporting lateralization, in addition to subjective assessment of image quality and noise.

The group inter-rater reproducibility as measured by the  $\kappa$  statistic was moderate, even without the use of CLI; this outcome was somewhat surprising, given the range of experience of the readers. While the most experienced readers had years of fMRI reading, the most junior members had, at most, a few months of training, amounting to approximately 10–20 fMRI examinations. The  $\kappa$  statistic slightly increased with the use of a CLI, indicating improved uniformity of interpretations from the entire group of neuroradiologists.

While there are many studies that assess the variability of fMRI statistical maps,<sup>6–9</sup> there are few studies that assess the variability of visual interpretation of fMRI statistical maps. Our inter-rater reproducibility compares well with other works; for example, Gaillard et al<sup>10,11</sup> used a related Cramer V statistic showing an inter-rater agreement of 0.77–0.82 for a single reading task, which increased to 0.93 for a panel of 3 language tasks. Gutbrod et al<sup>12</sup> showed high inter-rater agreement, depending on the location of the lateralization assessment, 0.90 for the inferior frontal gyrus and 0.97 for the superior temporal gyrus. Our study extends these findings in 3 ways. First, all fMRI examinations in this study were performed on patients, while the other studies incorporated mixtures of patients and healthy subjects. Second, our study incorporated an unknown but equal admixture of known nonleft and left-dominant patients to decrease any pretest incidence bias, whereas the other studies used a normal population incidence with approximately 5%–10% nonleft dominant. Last, this study incorporated 8 clinical neuroradiologists with a wide range of experience, while the other studies used either 3 or 4 specialized neuroimagers.

The sensitivity of fMRI lateralization to the left hemisphere compares well with other studies; for example, a recent meta-analysis<sup>13</sup> of 23 studies showed the sensitivity for detection of left dominance to be 88.1% (95% confidence interval, 87.0–89.2%). However, the authors showed the sensitivity for nonleft dominance to be 83.5% (95% confidence interval, 80.2%–86.7%), which is substantially higher than our value of 69%.

In addition to increased uniformity of lateralization conclusions, the accuracy of lateralization increased modestly with the use of a CLI. In addition to improving accuracy, junior members commented that CLI aided in training them in fMRI interpretation, lending confidence about conclusions. The use of a CLI may benefit smaller departments because it is easier for a larger group practice to train newer members due to the proximity of more experienced members.

Perhaps more important than the mild improvement of CLI on the averaged accuracy and reproducibility of a neuroradiologist's subjective assessment of lateralization is the reduction in the number of outliers—that is, false-negative and false-positive assessments. Despite the moderately high accuracy of subjective lateralization conclusions without CLI, there remains a non-negligible fraction of false-positive and false-negative conclusions from experienced readers. The use of CLI can reduce the fraction



of these errors; for example, the trichotomized analysis showed that using a CLI eliminated the serious errors of a neuroradiologist reading left dominance in patients with known right dominance.

False-negatives and false-positives can directly affect patient care by misinforming neurosurgeons while evaluating and planning surgical procedures. For example, for patients with epilepsy with either mesial temporal sclerosis or cortical dysplasia, presurgical planning for a temporal lobectomy depends more on lateralization than on localization. In either pathology, the epileptogenic focus is generally localized somewhere in the temporal lobe, yet the treatment is a more widespread temporal lobectomy, for which the surgical technique uses a formulaic approach that depends more on language lateralization than on language localization: If the temporal lobe with the lesion is nondominant, a larger resection (in general 6 cm from the temporal pole) is preferable to maximize the chances of completely resecting the epileptogenic zone. If the temporal lobe is dominant, resection will be restricted (in general 4.5 cm from the temporal pole) or tailored, with the need for speech mapping (intraoperatively through an awake craniotomy or extraoperatively with the placement of subdural grids and strips). In distinction, the surgical strategy is different for resection of tumors or in patients with intractable focal epilepsy due to MR imaging-visible lesions, where the detailed relationship of fMRI activated speech areas to the tumoral tissue is of paramount concern; thus, localization of the speech areas assumes an important variable in the resection strategy—that is, because CLI only localizes to the extent of 1 hemisphere, this technique does not have any practical application for surgical guidance for focal lesions, where sublobar accuracy is required.

From a more pragmatic aspect, fMRI results should be interpreted according to the clinical context to avoid misleading surgical actions and serious consequences to the patients. In general, there is little role for speech lateralization tests when the patient is right-handed and the lesion is located in the right hemisphere. If the patient is left-handed and the lesion is located in the right hemisphere and there is no clinical evidence of speech abnormality, the right hemisphere is likely nondominant and a confirmatory fMRI can be helpful for verification. In this case, if the fMRI shows right-sided speech, a high suspicion for a false-positive should be raised and fMRI results should be challenged. A confirmatory test, such as the Wada test, would be necessary to verify these results. If the fMRI confirms the clinical hypothesis, showing speech localized in the contralateral hemisphere, additional tests are unnecessary. However, if the lesion is located in the left hemisphere and there is clinical evidence of speech impairment, most likely this is a dominant hemisphere, regardless of whether the patient is right- or left-handed; the fMRI should be strongly challenged if it shows contralateral speech and a follow-up Wada test is mandatory. From a clinical point of view, this is the worst case scenario because it will mislead the surgeon to perform a surgical intervention in a possibly dominant hemisphere without any additional surgical plan for intraoperative speech localization, bringing catastrophic consequences to the patient.

There are several factors that contribute to variability of a neuroradiologist's conclusion about overall language lateralization. Image quality is important, particularly increased motion-in-

duced activation that is primarily seen along the brain's periphery, skull base, and periventricular regions. The overall visual assessment, including contributions from these regions, will tend to be more bilateral than unilateral. When image quality is high, interpretation is easier and there is less variability among the readers and no significant errors. Thus, a primary use of a CLI is for studies that do not have optimal quality. The choice of a threshold for imaging the overlay of activated regions on the anatomic maps is another source of variability because low values will tend to incorporate more regions of the brain and thereby produce a bilateral appearance. For this reason, we used a CLI method that is independent of threshold. A benefit of such a measure is that it weighs the effects of large volumes of lower activation with smaller volumes of higher activation, which may represent the underlying widespread character of language. Such patterns may be elusive and variable to a qualitative visual assessment, particularly given the requirement of the eye to visually integrate activation maps over many sections and compare one hemisphere with the other.

Another source of variability is in training the neuroradiologist's visual assessment of lateralization, wherein the use of a CLI may provide more uniformity within a group of readers, particularly a group including newly trained readers. Although the most inexperienced reader showed discrepant assessments of image quality and confidence, there was otherwise uniformity within the remaining members, several with only 1–2 years of experience, suggesting a short learning curve. A final source of variability is in the subjective opinion dividing a unilateral from bilateral assessment. For this reason, a numeric score or probabilities are provided from the CLI, rather than categorical conclusions.

## CONCLUSIONS

We have shown substantial inter-rater agreement within a group of neuroradiologists reading fMRI examinations for language lateralization, which can be mildly increased by the use of a CLI. The group of neuroradiologists had a range of experience that might be applicable to many group practices. The use of a CLI had a small but positive effect on the accuracy of a neuroradiologist's impression of language lateralization, particularly when the neuroradiologist had less confidence about his or her impression, either due to reduced image quality or a complicated and apparently ambiguous pattern of activation. Perhaps the greatest utility of incorporating the use of a CLI in clinical practice is to minimize the most egregious error of language lateralization, specifically that of false identification of a nonleft component.

Disclosures: Jorge Gonzalez-Martinez—UNRELATED: Grants/Grants Pending: Citi-  
zens United for Research in Epilepsy.\* \*Money paid to the institution.

## REFERENCES

1. Wada J, Rasmussen T. **Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance.** *J Neurosurg* 1960;17:266–82
2. Kwong KK, Belliveau JW, Chesler DA, et al. **Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation.** *Proc Natl Acad Sci U S A* 1992;89:5675–79
3. Ogawa S, Tank DW, Menon R, et al. **Intrinsic signal changes accompanying sensory stimulation: functional brain mapping**

- with magnetic resonance imaging. *Proc Natl Acad Sci U S A* 1992;89:5951–55
4. Bandettini PA, Wong EC, Hinks RS, et al. Time course EPI of human brain function during task activation. *Magn Reson Med* 1992;25:390–97
  5. Jones SE, Mahmoud SY, Phillips MD. A practical clinical method to quantify language lateralization in fMRI using whole-brain analysis. *Neuroimage* 2011;54:2937–49
  6. Casey BJ, Cohen JD, O'Craven K, et al. Reproducibility of fMRI results across four institutions using a spatial working memory task. *Neuroimage* 1998;8:249–61
  7. Le TH, Hu X. Methods for assessing accuracy and reliability in functional MRI. *NMR Biomed* 1997;10:160–64
  8. Maitra R, Roys SR, Gullapalli RP. Test-retest reliability estimation of functional MRI data. *Magn Reson Med* 2002;48:62–70
  9. McGonigle DJ, Howseman AM, Athwal BS, et al. Variability in fMRI: an examination of intersession differences. *Neuroimage* 2000;11(6 pt 1):708–34
  10. Gaillard WD, Balsamo L, Xu B, et al. Language dominance in partial epilepsy patients identified with an fMRI reading task. *Neurology* 2002;59:256–65
  11. Gaillard WD, Balsamo L, Xu B, et al. fMRI language task panel improves determination of language dominance. *Neurology* 2004;63:1403–08
  12. Gutbrod K, Spring D, Degonda N, et al. Determination of language dominance: Wada test and fMRI compared using a novel sentence task. *J Neuroimaging* 2012;22:266–74
  13. Dym RJ, Burns J, Freeman K, et al. Is functional MR imaging assessment of hemispheric language dominance as good as the Wada test?: a meta-analysis. *Radiology* 2011;261:446–55