



**Get Clarity On Generics**

Cost-Effective CT & MRI Contrast Agents

**FRESENIUS  
KABI**

**WATCH VIDEO**

**AJNR**

This information is current as  
of August 20, 2025.

## **Identification of the Language Network from Resting-State fMRI in Patients with Brain Tumors: How Accurate Are Experts?**

S.K. Gujar, K. Manzoor, J. Wongsripuentet, G. Wang, D.  
Ryan, S. Agarwal, M. Lindquist, B. Caffo, J.J. Pillai and H.I.  
Sair

*AJNR Am J Neuroradiol* published online 23 February 2023  
<http://www.ajnr.org/content/early/2023/02/23/ajnr.A7806>

# Identification of the Language Network from Resting-State fMRI in Patients with Brain Tumors: How Accurate Are Experts?

 S.K. Gujar,  K. Manzoor,  J. Wongsripumtet,  G. Wang,  D. Ryan,  S. Agarwal,  M. Lindquist,  B. Caffo,  J.J. Pillai, and  H.I. Sair



## ABSTRACT

**BACKGROUND AND PURPOSE:** Resting-state fMRI helps identify neural networks in presurgical patients who may be limited in their ability to undergo task-fMRI. The purpose of this study was to determine the accuracy of identifying the language network from resting-state-fMRI independent component analysis (ICA) maps.

**MATERIALS AND METHODS:** Through retrospective analysis, patients who underwent both resting-state-fMRI and task-fMRI were compared by identifying the language network from the resting-state-fMRI data by 3 reviewers. Blinded to task-fMRI maps, these investigators independently reviewed resting-state-fMRI ICA maps to potentially identify the language network. Reviewers ranked up to 3 top choices for the candidate resting-state-fMRI language map. We evaluated associations between the probability of correct identification of the language network and some potential factors.

**RESULTS:** Patients included 29 men and 14 women with a mean age of 41 years. Reviewer 1 (with 17 years' experience) demonstrated the highest overall accuracy with 72%; reviewers 2 and 3 (with 2 and 7 years' experience, respectively) had a similar percentage of correct responses (50% and 55%). The highest accuracy used ICA50 and the top 3 choices (81%, 65%, and 60% for reviewers 1, 2, and 3, respectively). The lowest accuracy used ICA50, limiting each reviewer to the top choice (58%, 35%, and 42%).

**CONCLUSIONS:** We demonstrate variability in the accuracy of blinded identification of resting-state-fMRI language networks across reviewers with different years of experience.

**ABBREVIATIONS:** BOLD = blood oxygen level–dependent; ICA = independent component analysis; rs = resting-state

Resting-state (rs) fMRI has emerged as a novel tool to analyze brain function. In contrast to traditional task-fMRI, no explicit task is required in rs-fMRI while blood oxygen level–dependent (BOLD) images are acquired. Owing to an assortment of naturally occurring fluctuations of BOLD activity in various regions of the brain, a set of intrinsic brain networks can be identified by examining spatially distinct, however temporally synchronous, BOLD signals at rest. The number of discrete brain networks is somewhat

arbitrary and depends on the specific threshold used to define a network; nevertheless, a relatively consistent set of major networks has been reliably demonstrated in many studies.<sup>1</sup>


Although clinical use of rs-fMRI is currently not widespread, potentially partly due to the variability of brain networks when examined at the single-subject level,<sup>2</sup> there was considerable effort to translate this technique into clinical practice. Thus, the most promising use of rs-fMRI currently seems to be in the domain of presurgical brain mapping.<sup>3</sup> While task-fMRI has been successfully used to identify critical brain regions for preoperative planning, task-fMRI requires compliance in performing behavioral paradigms necessary for determining brain activation. This compliance may be limited or absent in many cases, for example, in the pediatric population or in the elderly; in patients with language barriers, visual or hearing impairment, or cognitive/memory impairment that may preclude language task performance; or in those with physical debilitation limiting movement for performing motor tasks. Given the lack of a task, therefore, the use of rs-fMRI is attractive as an alternative technique to assess brain function. Furthermore, rs-fMRI may potentially require less time compared with task-fMRI to obtain comparable data.

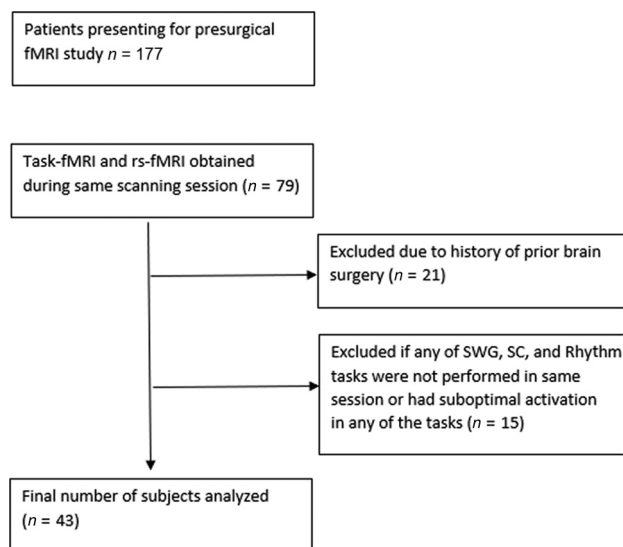
Received October 28, 2021; accepted after revision January 4, 2023.

From the Division of Neuroradiology (S.K.G., K.M., J.W., D.R., S.A., J.J.P., H.I.S.), The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, Maryland; and Department of Biostatistics (G.W., M.L., B.C.), Department of Neurosurgery (J.J.P.), and The Malone Center for Engineering in Healthcare (H.I.S.), The Whiting School of Engineering, Johns Hopkins University, Baltimore, Maryland.

This work was supported by grant RSCH1420 from the grant program Radiological Society of North America (RSNA) Research & Education Foundation Carestream Health/RSNA Research Scholar Grant.

Please address correspondence to Haris I. Sair, MD, Johns Hopkins University School of Medicine, 600 North Wolfe St, Phipps B-112A, Baltimore, MD 21287; e-mail: hsair1@jhmi.edu; @hsairmd

 Indicates article with online supplemental data.  
<http://dx.doi.org/10.3174/ajnr.A7806>



**FIG 1.** Flow diagram of inclusion and exclusion criteria. SWG indicates Silent Word Generation; SC, Sentence Completion.

As 1 example, rs-fMRI-derived motor networks have been shown to be comparable with motor regions activated during task-fMRI as well as motor regions identified during direct cortical stimulation.<sup>4</sup> The concordance between rs-fMRI-derived language networks and task-fMRI-activated language regions is more variable. Earlier reports suggested moderate concordance;<sup>5</sup> however, more recent reports have demonstrated high subject-level variability.<sup>6,7</sup> While motor networks are relatively easily identified from rs-fMRI due to the relative invariance of the anatomic-functional relationship of the motor system across individuals, identification of the language network may be more challenging due to relatively high variance in localization of language areas across individuals<sup>8</sup> and the similarity of elements of the language network to the spatial distribution of other networks such as the ventral attention network.<sup>9</sup>

If rs-fMRI is to be used as an alternative to task-fMRI, the accuracy of identification of the language network solely from rs-fMRI must be assessed when one is presented with multiple network correlation maps, ie, the output generated from commonly used methods of rs-fMRI analysis such as independent component analysis (ICA). While automated methods of network identification are currently being developed, we were interested in the ability of humans to correctly identify the language network. We examined the human accuracy of identifying subject-level rs-fMRI language networks using their task-fMRI language activation maps as the reference standard. We hypothesize that there is low accuracy in human identification of rs-fMRI language networks when language-related task-fMRI activation maps are unavailable.

## MATERIALS AND METHODS

### Participants

The radiology information system was interrogated for any patient who underwent fMRI for presurgical brain mapping between January 1, 2009, and July 1, 2015. Seventy-nine patients with intracranial neoplasms were identified for whom both language task-fMRI and rs-fMRI were available in the same

imaging session. Twenty-one patients had a history of an invasive intracranial procedure (surgery and/or biopsy) and were excluded to minimize confounding the effect of susceptibility. Because individual language paradigms commonly activate only subsets of the global language system, we included only patients who had completed 3 different language tasks (Silent Word Generation, Sentence Completion, and Rhyming) during the same session. Patients who had suboptimal activation on any 1 of the 3 language task-fMRIs, assessed subjectively as is routine in clinical fMRI cases, were excluded (15 patients). This exclusion was determined by lack of the expected localization of activation with excessive spurious activation in nonbrain regions with subject-specific statistical thresholding. After exclusion, data from 43 subjects were available for analysis (Fig 1).

### Handedness

The Edinburgh Handedness Inventory was used to determine patient handedness.<sup>10</sup>

### Imaging

Images were acquired on 3T Tim Trio MR imaging system (Siemens) using a 12-channel head matrix coil. For both task-fMRI and rs-fMRI, T2\*-weighted BOLD images were acquired using 2D gradient-echo echo-planar imaging: TR = 2000 ms, TE = 30 ms, flip angle = 90°, FOV = 24 cm, acquisition matrix = 64 × 64 × 33, section thickness = 4mm, section gap = 1 mm, interleaved acquisition. Instructions for rs-fMRI were the following: Keep your eyes closed, don't move, and don't think of anything in particular. One hundred eighty volumes were acquired for rs-fMRI (6 minutes). We also obtained 3D T1-weighted structural images: TR = 2300 ms, TI = 900 ms, TE = 3.5 ms, flip angle = 9°, FOV = 24 cm, acquisition matrix = 256 × 256 × 176, section thickness = 1 mm.

### Task-fMRI Paradigms

As is routine at our institution, we instructed the patients and performed practice sessions outside the scanner before fMRI to ensure that patients understood the tasks. Real-time fMRI maps were monitored by the neuroradiologist administering the study to assess global data quality. Any task with suboptimal activation assessed subjectively was repeated per our protocol; for final analysis, the single best run of each task was chosen. The Prism Software Suite was used for stimulus presentation (Prism Clinical Imaging). A block design of either 30 (rhyming) or 20 (Silent Word Generation, Sentence Completion) seconds of alternative tasks and control blocks was used for an imaging time of 3 (Rhyming) or 4 minutes (Silent Word Generation, Sentence Completion).

### Image Processing and Analysis

Statistical Parametric Mapping (SPM) Version 8 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm12>) and custom Matlab (MathWorks) scripts were used to process the fMRI.

### Task-fMRI

Task-fMRI underwent slice timing correction followed by motion correction. Images were normalized to a Montreal Neurological Institute-152 template and spatially smoothed using

a 6-mm full width at half maximum Gaussian kernel. A general linear model analysis was performed using a canonical hemodynamic response function convolved with the boxcar function for each task, without using model derivatives or global intensity normalization. A 128-second high-pass filter was used. An autoregressive was used to account for temporal autocorrelations. No confound matrix was used. A contrast design matrix set to detect activation across all 3 tasks compared with rest was used for each subject. SPM T-contrast maps were generated without clustering or multiple comparison correction, which is the approach that we use routinely for clinical language mapping because clustering is usually not necessary at the selected thresholds and correction for multiple comparisons would be too stringent for such single-subject language-activation analysis.

### Rs-fMRI

Rs-fMRI underwent slice timing correction followed by motion correction. The ArtRepair toolbox ([https://www.nitrc.org/projects/art\\_repair/](https://www.nitrc.org/projects/art_repair/))<sup>11</sup> was then used to detect volumes with large shifts in global average signal intensity related to scan-to-scan motion; both the outlier volumes and additional volumes recommended for de-weighting were tagged for subsequent removal from analysis (ie, for “scrubbing”). Rs-fMRI was linearly detrended, and following coregistration of rs-fMRI and T1-weighted images, physiologic nuisance regression of rs-fMRI was performed using component-based noise correction (CompCor method)<sup>12</sup> using signal extracted from eroded white matter and CSF masks. After bandpass filtering from 0.01 to 0.1 Hz, smoothing was performed with a 6-mm full width at half maximum Gaussian kernel. Finally, images tagged by ArtRepair were removed (scrubbed) from the rs-fMRI volumes.

The Group Independent Component Analysis of fMRI Toolbox (GIFT, Medical Image Analysis Lab, <http://mialab.mrn.org/software/gift>) was used to generate ICA maps for each subject using 20 (ICA20) and 50 (ICA50) target components, using the InfoMax algorithm with ICASSO<sup>13</sup> set at 5 repeats. The language network was identified by first sorting the ICA components using multiple regression in GIFT, with the task-fMRI SPM T-maps as the reference template. Then, the component that demonstrated the highest spatial overlap with the task-fMRI maps localized to Broca and Wernicke activation was selected as the rs-fMRI language network. Of note, in all cases, there was 1 ICA component that best represented the primary language network for both ICA orders. This ICA component was labeled the rs-fMRI language network map for each subject.

### Rs-fMRI Language Network Identification

Three participants independently reviewed the rs-fMRI ICA maps to identify the potential language network. All were blinded to the task-fMRI activation maps. The reviewers ranged in fMRI experience: 17 years (neuroradiologist), 7 years (neuroradiologist), and 2 years (neuroimaging postdoctoral researcher). Each of the raters had similar previous exposure to rs-fMRI and had experience in interpreting rs-fMRI network maps. Images were presented to each reviewer using the orthogonal viewer in FSLView (Version 3.1; <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslView/UserGuide>) with each subject's rs-fMRI maps overlaid on their T1-weighted images. Review was performed independently

for ICA20 and ICA50. The reviewers were allowed to modify image contrast and thresholds. Reviewers ranked up to 3 top choices for the candidate rs-fMRI language map, as well as their confidence in their assessment ranging from 1 (highly confident) to 5 (not confident).

### Statistical Tests

We aimed to discover the association between the probability of correct identification of the language network and some potential factors including the reviewers, ICA type, and each selection scenario.

We modeled this study using generalized linear mixed-effects models similar to studies of measurement reproducibility.  $Y_{ijkl}$  denoted the correctness of identification for subject  $i$  by reviewer  $j$  using ICA type  $k$  and top choices  $l$ .  $Y_{ijkl}$  takes values 1 or 0, indicating whether the identification is correct. The cross-sectional structure motivates the generalized mixed-effects model as follows:

$$\text{logit}(P(Y_{ijkl} = 1|u_i)) = \alpha + u_i + \beta_j + x_k + z_l + h_i \quad (1),$$

where the  $u_i$  is mutually independent normally distributed,  $N(0, \sigma^2)$ , random intercepts. The covariate  $\beta_j$  represents the fixed effect of reviewer  $j$ . The effect of ICA20 and ICA50 is denoted by  $x_1$  and  $x_2$ , respectively. The covariates  $z_1, z_2, z_3$  denote the effect of 3 different selection scenarios (top 1 choice, top 2 choices, and top 3 choices). The term  $h_i$  denotes the handedness for subject  $i$ . To ensure the identifiability of the model, we introduced constraints  $\beta_1 = 0, x_1 = 0, z_1 = 0$ .

We performed likelihood ratio tests to investigate interaction terms between each pair of the fixed effects. The resulting tests were nonsignificant. We also found that the effect of ICA type and handedness does not help to improve the model fit, so the associated terms were excluded. In other words, there was no significant effect for the ICA type and handedness, given the other variables. We also performed the test for necessity of random intercept effect ( $H_0 : \sigma^2 = 0$  versus  $H_1 : \sigma^2 > 0$ ) by parametric bootstrapping. We found improvement in model fit (bootstrapped  $P$  value estimated to be 0) by fitting the model, accounting for the subject-specific random effects.

To further explore the association between the probability of correct identification and the confidence rating or the location of lesions, adjusted for the effect of each reviewer, ICA types, and selection scenarios, we fit a similar generalized mixed-effects model, accounting for these effects. Tumor location was recorded for involvement of the frontal lobe, parietal lobe, temporal lobe, occipital lobe, or deep (subcortical) areas, with additional specific tags for involvement of the left inferior frontal gyrus and right inferior frontal gyrus (either of them potentially involving the Broca areas) and of the left posterior temporal lobe and right posterior temporal lobe (either of them potentially involving the Wernicke areas).

The final model we fit is the following:

$$\text{logit}(P(Y_{ijkl} = 1|u_i)) = \alpha + u_i + \beta_j + z_l \quad (2),$$

where  $u_i$  follows the normal distribution,  $N(0, \sigma^2)$ , independently. This is the simplified version of model 1 after a sequence of

**Table 1: Correctness percentage by reviewer and ICA target component**

Reviewer	ICA Component	Top Choices	No. (% Correctness)	95% CIs
1	Overall		72.09%	66.12%–77.39%
1	20	Top 1	28 (65.12%)	49.01%–78.55%
1	20	Top 2	32 (74.42%)	58.53%–85.96%
1	20	Top 3	33 (76.74%)	61.00%–87.72%
1	50	Top 1	25 (58.14%)	42.21%–72.63%
1	50	Top 2	33 (76.74%)	61.00%–87.72%
1	50	Top 3	35 (81.40%)	66.08%–91.08%
2	Overall		50.39%	44.13%–56.63%
2	20	Top 1	16 (37.21%)	23.39%–53.28%
2	20	Top 2	23 (53.49%)	37.83%–68.53%
2	20	Top 3	24 (55.81%)	40.01%–70.59%
2	50	Top 1	15 (34.88%)	21.45%–50.99%
2	50	Top 2	24 (55.81%)	40.01%–70.59%
2	50	Top 3	28 (65.12%)	49.01%–78.55%
3	Overall		55.04%	48.74%–61.18%
3	20	Top 1	22 (51.16%)	35.68%–66.44%
3	20	Top 2	26 (60.47%)	44.45%–74.63%
3	20	Top 3	26 (60.47%)	44.45%–74.63%
3	50	Top 1	18 (41.86%)	27.37%–57.79%
3	50	Top 2	24 (55.81%)	40.01%–70.59%
3	50	Top 3	26 (60.47%)	44.45%–74.63%

**Table 2: Difference in correctness percentages comparing the top choice with the top 2 and 3 choices**

Reviewer	ICA	Pair Comparison	Number (Percentage Differences)	P Value <sup>a</sup>
1	20	Top 1 vs top 2	4 (8.30%)	.063
	20	Top 1 vs top 3	5 (11.63%)	.031 <sup>a</sup>
	20	Top 2 vs top 3	1 (2.33%)	.500
1	50	Top 1 vs top 2	8 (18.60%)	.004 <sup>a</sup>
	50	Top 1 vs top 3	10 (23.26%)	<.001 <sup>a</sup>
	50	Top 2 vs top 3	2 (4.65%)	.250
2	20	Top 1 vs top 2	7 (16.28%)	.008 <sup>a</sup>
	20	Top 1 vs top 3	8 (18.60%)	.004 <sup>a</sup>
	20	Top 2 vs top 3	1 (2.32%)	.500
2	50	Top 1 vs top 2	9 (20.93%)	.002 <sup>a</sup>
	50	Top 1 vs top 3	13 (30.23%)	<.001 <sup>a</sup>
	50	Top 2 vs top 3	4 (9.30%)	.063
3	20	Top 1 vs top 2	4 (9.30%)	.063
	20	Top 1 vs top 3	4 (9.30%)	.063
	20	Top 2 vs top 3	0 (0.00%)	1.0
3	50	Top 1 vs top 2	6 (13.95%)	.016 <sup>a</sup>
	50	Top 1 vs top 3	8 (18.60%)	.004 <sup>a</sup>
	50	Top 2 vs top 3	2 (4.65%)	.250

<sup>a</sup> P value from the 1-sided sign test.

tests, as described above and will be used for inference and interpretation. From model 2, we can see that the probability of correct identification is associated with the subject-specific effect, reviewer-specific effect, and the selection scenarios.

A conditional logistic regression model given the subject-specific effect was constructed to explore the association between the probability of correct identification and the confidence rating adjusted for the effect of each reviewer and ICA type. No significant association resulted. To measure the concordance of rs-fMRI language network map identification among and within the 3 reviewers, we calculated intrarater and interrater  $\kappa$  statistics.

All statistical analyses were performed in R statistical and computing software, Version 3.6.2 (<http://www.r-project.org/>).

## RESULTS

Patients included 29 men and 14 women with a mean age of 41 years (minimum 18 and maximum 69 years). Thirty-three patients are right-handed, 8 were left-handed, and 2 were ambidextrous.

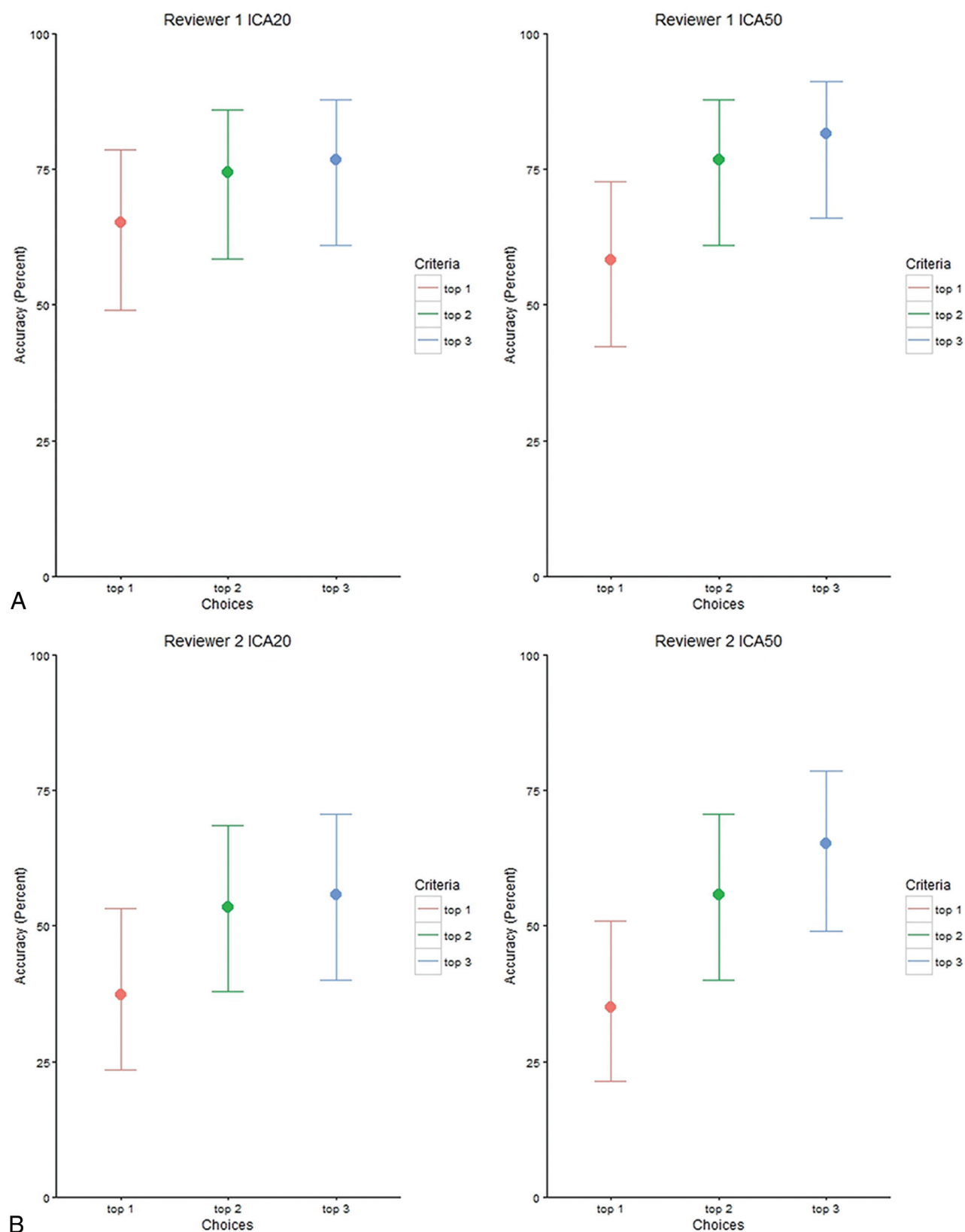
Tables 1 and 2 illustrate the marginal association between correctness percentage and the reviewers, ICA type, and selection scenarios. Table 1 and the Online Supplemental Data demonstrate the percentage correct for each reviewer for ICA20 and ICA50, assessing their top choice (1), the top 2 choices (1 + 2), or the top 3 choices (1 + 2 + 3), with Fig 2 graphically depicting the overall comparisons. Reviewer 1, with 17 years of fMRI experience, demonstrated the highest overall accuracy with 72% (95% CI, 66.1%–77.4%) correct responses across all conditions. Reviewers 2 and 3, with 2 and 7 years of experience respectively, had overall similar percentages of correct responses (50% [95% CI, 44.1%–56.6%] and 55% [95% CI, 48.7%–61.2%]). For each reviewer, the highest accuracy was obtained using ICA50 and top 3 choices (81%, 65%, and 60% for reviewers 1, 2, and 3, respectively). Conversely, the lowest accuracy was also obtained using ICA50, however limiting each reviewer to the top choice (58%, 35%, and 42%).

Significant differences in accuracy were seen between selecting only the top choice versus selecting the top 3 choices (Table 2), with the exception of ICA20 for reviewer 3. In both ICA conditions and across all raters, no significant difference in accuracy was seen between the selection of the top 2 choices versus the top 3 choices.

When we compared the single top choice with the top 2 choices, mixed findings were noted between ICA orders and raters. Across the raters, inclusion of 3 choices compared with 1 choice improved the accuracy by 13% for ICA and 20% and 24% for ICA50. When examining the converse problem, whether ICA order increases accuracy, no significant difference was found for any of the reviewers in any of the 3 choice conditions (Online Supplemental Data). Figure 3 demonstrates, averaged across all subjects, the spatial distribution of task-fMRI language activation, correctly selected rs-fMRI language component, and incorrectly selected rs-fMRI.

$\kappa$  statistics (Table 3) demonstrated an overall fair concordance between reviewer 1 and reviewers 2 and 3 ( $\kappa = 0.35$ –0.40).





**FIG 2.** Correctness percentage by reviewer and ICA target components with their corresponding confidence intervals.

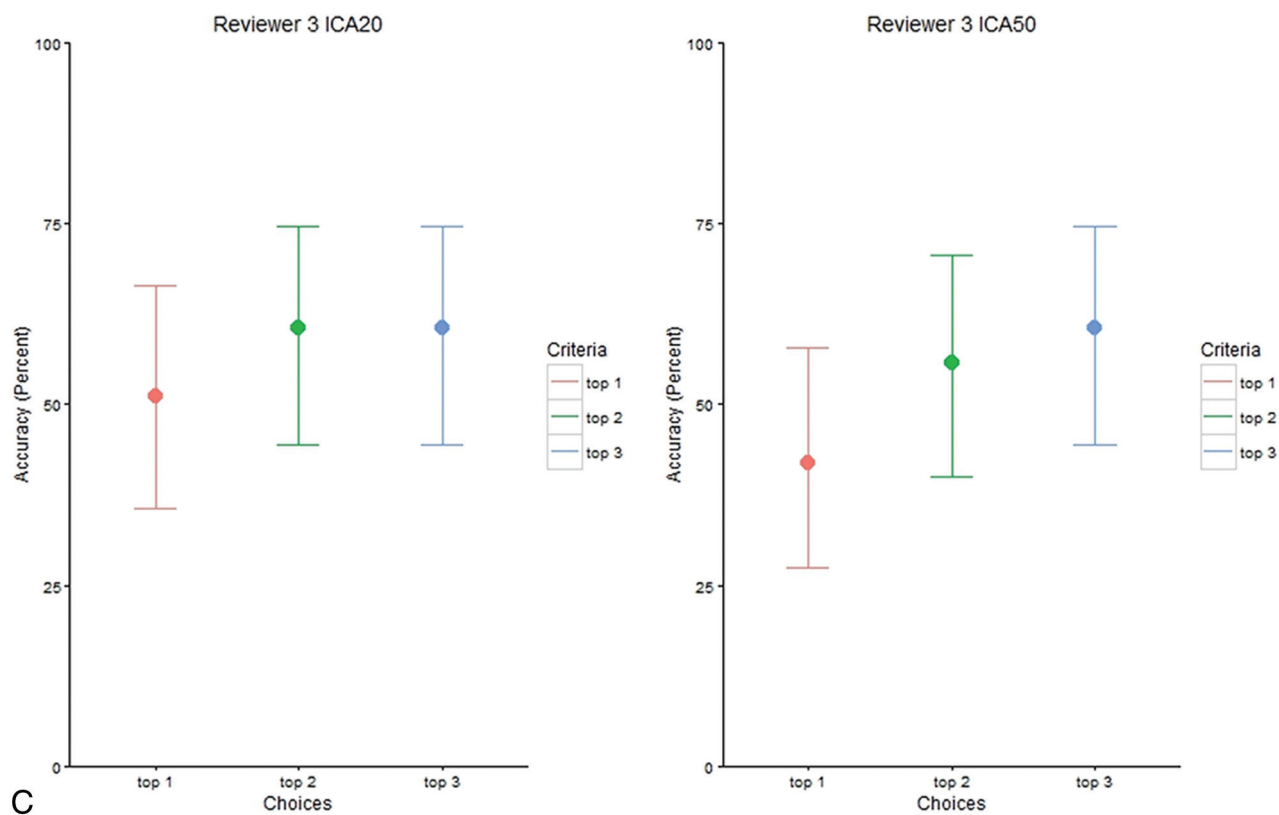


FIG 2. Continued.

Poor concordance was observed between reviewers 2 and 3 ( $\kappa = 0.16$ ). When we limited analysis to the top choice, again reviewer 1 demonstrated fair concordance with reviewers 2 and 3 ( $\kappa = 0.21$ – $0.47$ ), and there was poor concordance between reviewers 2 and 3 ( $\kappa = 0.028$ ). Similar findings were seen when analyzing the concordance of any of the top 3 choices (reviewer 1 versus 2 and 3 [ $\kappa = 0.31$ – $0.42$ ], and reviewer 2 versus 3 [ $\kappa = 0.12$ ]).

On the basis of model 2, conditional on subject, the odds of correct identification for reviewer 2 were 0.26 (95% CI, 0.17–0.41) times the odds of reviewer 1, given the same selection scenario. Similarly, the OR of reviewer 3 against reviewer 1 was 0.34 (95% CI, 0.22–0.53), and the OR of reviewer 3 against reviewer 2 was 1.31 (95% CI, 0.86–1.99). Note that there was no significant difference between odds of reviewer 2 and reviewer 3, given the same selection scenario conditional on the subject.

From the same model, the odds of correct identification for selecting the top 2 choices were 2.39 (95% CI, 1.56–3.69) times the odds of the top 1 choice, given the same reviewer and conditional on subject. The OR of selecting the top 3 choices against the top 1 choice was 3.05 (95% CI, 1.98–4.76), and the OR of including the top 3 choices against the top 2 was 1.28 (95% CI, 0.83–1.98). The difference between the odds of selecting the top 3 and top 2 choices is not statistically significant.

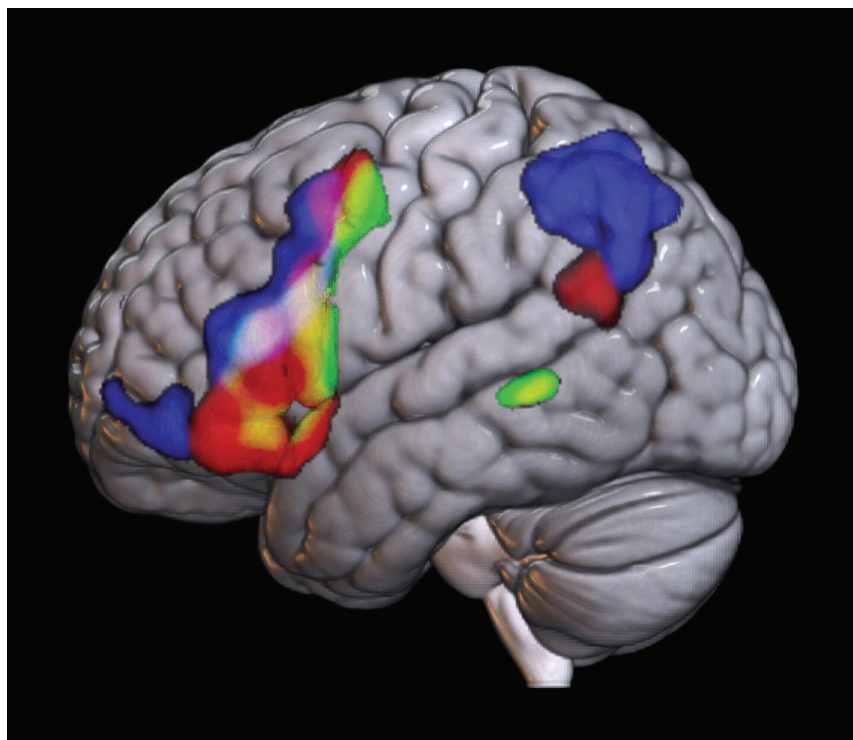
Regarding the association between the probability of correct identification and the confidence rating or the location of lesions, the likelihood ratio test informs us that there is no

significant effect by the confidence rating and lesion locations ( $P$  value = .365).

## DISCUSSION

Rs-fMRI is being increasingly used in the setting of presurgical brain mapping.<sup>3,14,15</sup> Despite the potential subject level variability of data accuracy,<sup>6,7</sup> nevertheless in select cases, rs-fMRI may be considered a viable option for presurgical brain mapping; indeed, at least 1 institution has included rs-fMRI in their presurgical brain mapping paradigm without obtaining task-fMRI.<sup>3,16</sup> For this purpose, obtaining highly accurate intrinsic brain network data is paramount to avoid adverse outcomes following surgery.

Two widely used methods of rs-fMRI analysis are associated with unique limitations. Seed-based analysis necessitates placing ROIs in selected areas of the brain, depending on the network to be defined. The advantage of this method is that anatomy can guide placement of ROIs to target specific networks if clear landmarks for ROI placement are available, such as for the motor network. However, in the setting of presurgical language mapping, several confounds appear. First, primary language areas are more widely distributed anatomically across subjects;<sup>8</sup> thus, placement of an ROI in the inferior portion of the pars opercularis, for example, may work for 1 patient but not another. Second, often there are anatomic distortions in expected language regions in patients with large brain lesions, making it difficult to determine the correct



**FIG 3.** Comparison of task- versus rs-fMRI averaged across subjects. Green denotes average task-fMRI activation of the language system across all subjects. Red denotes the average spatial map of the rs-fMRI component correctly selected across subjects by all reviewers. Blue represents the average spatial map of the incorrectly selected rs-fMRI component across subjects by all reviewers. Overlays are additive: Yellow denotes overlapping voxels between green (task-fMRI) and red (correct rs-fMRI component). The white area in the inferior frontal gyrus along the inferior frontal sulcus denotes overlap across task-fMRI, incorrect rs-fMRI, and correct rs-fMRI. While some overlap is noted here, which may be a function of spatial smoothing, there is a clear distinction between the correct rs-fMRI versus incorrect rs-fMRI distribution.

**Table 3: Interrater (between reviewers)  $\kappa$  statistic by top choice groups**

Reviewer	1	2	3
Overall			
1	1.0	0.35	0.40
2	0.35	1.0	0.16
3	0.40	0.16	1.0
Top choice			
1	1.0	0.47	0.21
2	0.47	1.0	0.028
3	0.21	0.028	1.0
Top 2 choices			
1	1.0	0.37	0.37
2	0.37	1.0	0.27
3	0.37	0.27	1.0
Top 3 choices			
1	1.0	0.31	0.42
2	0.31	1.0	0.12
3	0.42	0.12	1.0

anatomic landmarks. Damage to primary language regions due to tumor infiltration or prior surgery may also cause reorganization of language networks,<sup>17,18</sup> further limiting accurate ROI placement. Finally, neurovascular uncoupling may adversely

affect seed placement due to lack of expected BOLD fluctuations in network subregions.<sup>19</sup>

The alternative is to use a data-driven approach such as ICA. Here, rs-fMRI time-series typically at the voxel level is separated into maximally independent components. The resultant component maps each may represent one of the various intrinsic brain networks, or nuisance. To determine relevant maps, one may use automated methods such as template-matching; however as is the case with seed-based analysis in patients with large lesions, anatomic distortion or network reorganization may impede its accuracy. More commonly, relevant network maps are selected by visual inspection, taking into consideration potential changes in network topology.

Aside from (however related to) the problem of choosing the ideal number of target components in ICA (ie, ICA order), for which there is no clear paradigm,<sup>20</sup> the issue of categorizing networks derived from ICA can be especially challenging due to several factors. First, network maps do not necessarily break evenly across components, and using low ICA orders may cause merging of one or more networks (or even networks and noise), and using high ICA orders may result

in fragmentation of networks into subnetworks. For our study, using task-fMRI maps as the target, we found that a single component for ICA20 and ICA50 best matched the target, with no fragmentation at these levels of ICA orders. The general range of optimum ICA order is affected by several technical factors, including hardware, length of the scan, and postprocessing. The rs-fMRI output maps, therefore, may need to be evaluated in each unique experimental condition to optimize the general range of target ICA orders. At our institution for example, an ICA order of 50 appears to be ideal for use when rs-fMRI is included as part of the presurgical brain mapping protocol in addition to task-fMRI, the scanning of which is limited to 1 scanner. However, empirically, a higher target order has been necessary when the rs-fMRI protocol is used on a different scanner, at least partially due to differences in contrast to noise.

Despite having a relatively long clinical and research fMRI career of 17 years, the first rater's accurate identification of rs-fMRI language maps peaked at 72% across any condition; this fact should be noted by those who are using this method for research or clinical reasons. Limiting the choice to the best guess decreased accuracy to about 65%. With fewer years of



experience, the accuracy was overall not better than chance (50%). Caution, therefore, is advised when identification of language maps is a critical component of management, for example, to decide on an operative approach or the potential extent of tumor resection. Confirmation of the selected language component with additional methods (eg, task-fMRI or intraoperative mapping) would be critical.

Automated methods of network identification may, therefore, be necessary when task-fMRI is not available. Using a multilayer perceptron (a neural network), Mitchell et al<sup>21</sup> demonstrated that reliable intrinsic brain networks could be characterized in 13 patients with distorted brain anatomy with electrocortical stimulation as the criterion standard. Further validation in larger sample sizes would be necessary to ensure that methods such as this have high reliability and reproducibility. Finally, a measure of inherent reliability of the rs-fMRI maps would be important to develop.

There are several limitations in our study. We used task-fMRI as the reference standard for ground truth, due to a lack of availability of consistent intraoperative mapping. While intraoperative mapping with direct cortical stimulation itself has specific limitations such as a potential lack of specificity, generally, it is considered superior to task-fMRI for functional localization.

We limited ICA orders to 20 and 50 empirically, on the basis of the optimum number of target components based on prior analysis of this data set. In this specific case, the language component was seen in a single component for both ICA20 and ICA50 as noted above. However, especially in lower ICA orders, rs-fMRI language networks may have been mixed with other networks, confounding the labeling of the ICA component to the correct network. In the setting of using data-driven approaches to ICA estimation (eg, as is implemented in the MELODIC tool in FSL, <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MELODIC>) or setting the ICA order to an arbitrary higher number, mixing (due to low order) or fragmentation (due to higher order) may result in greater challenges in detecting the language network.

Finally, while potential distortion of language networks was addressed in this study, the contribution of technical and physiologic factors that result in further deviation of the networks from the expected appearance could introduce further difficulty in identification of the network. For example, the presence of hemorrhage or prior surgery may result in susceptibility artifacts masking certain components of the network. Similar loss of detection of network subcomponents may occur in the presence of neurovascular uncoupling.

## CONCLUSIONS

We demonstrate the variability of the accuracy of blinded identification of rs-fMRI language networks across raters and different years of fMRI experience, using task-fMRI language activation maps as the reference. In addition to rs-fMRI-based brain mapping, additional confirmatory methods to supplement this information such as through task-fMRI or intraoperative stimulation may be necessary. Further work is needed to

determine the diagnostic utility of rs-fMRI using a more reliable criterion standard.

**Disclosure forms** provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

## REFERENCES

1. Barkhof F, Haller S, Rombouts SA. **Resting-state functional MR imaging: a new window to the brain.** *Radiology* 2014;272:29–49 [CrossRef Medline](#)
2. Airan RD, Vogelstein JT, Pillai JJ, et al. **Factors affecting characterization and localization of interindividual differences in functional connectivity using MRI.** *Hum Brain Mapp* 2016;37:1986–97 [CrossRef Medline](#)
3. Lee MH, Miller-Thomas MM, Benzinger TL, et al. **Clinical resting-state fMRI in the preoperative setting: are we ready for prime time?** *Top Magn Reson Imaging* 2016;25:11–18 [CrossRef Medline](#)
4. Qiu T, Yan C, Tang W, et al. **Localizing hand motor area using resting-state fMRI: validated with direct cortical stimulation.** *Acta Neurochir (Wien)* 2014;156:2295–302 [CrossRef Medline](#)
5. Tie Y, Rigolo L, Norton IH, et al. **Defining language networks from resting state fMRI for surgical planning: a feasibility study.** *Hum Brain Mapp* 2014;35:1018–30 [CrossRef Medline](#)
6. Sair HI, Yahyavi-Firouz-Abadi N, Calhoun VD, et al. **Presurgical brain mapping of the language network in patients with brain tumors using resting-state fMRI: comparison with task fMRI.** *Hum Brain Mapp* 2016;37:913–23 [CrossRef Medline](#)
7. Cochereau J, Deverduin J, Herbet G, et al. **Comparison between resting state fMRI networks and responsive cortical stimulations in glioma patients.** *Hum Brain Mapp* 2016;37:3721–32 [CrossRef Medline](#)
8. Sanai N, Mirzadeh Z, Berger MS. **Functional outcome after language mapping for glioma resection.** *N Engl J Med* 2008;358:18–27 [CrossRef Medline](#)
9. Corbetta M, Patel G, Shulman GL. **The reorienting system of the human brain: from environment to theory of mind.** *Neuron* 2008;58:306–24 [CrossRef Medline](#)
10. Oldfield RC. **The assessment and analysis of handedness: the Edinburgh Inventory.** *Neuropsychologia* 1971;9:97–113 [CrossRef Medline](#)
11. Mazaika P, Hoeft F, Glover G, et al. **Methods and software for fMRI analysis of clinical subjects.** *Neuroimage* 2009;47:S58 [CrossRef](#)
12. Behzadi Y, Restom K, Liao J, et al. **A component based noise correction method (CompCor) for BOLD and perfusion based fMRI.** *Neuroimage* 2007;37:90–101 [CrossRef Medline](#)
13. Himberg J, Hyvärinen A, Esposito F, et al. **Validating the independent components of neuroimaging time series via clustering and visualization.** *Neuroimage* 2004;22:1214–22 [CrossRef Medline](#)
14. Hsu A, Chen HS, Hou P, et al. **Presurgical resting-state functional MRI language mapping with seed selection guided by regional homogeneity.** *Magn Reson Med* 2020;84:375–83 [CrossRef Medline](#)
15. Leuthardt EC, Allen M, Kamran M, et al. **Resting-state blood oxygen level-dependent functional MRI: a paradigm shift in preoperative brain mapping.** *Stereotact Funct Neurosurg* 2015;93:427–39 [CrossRef Medline](#)
16. Leuthardt EC, Guzman G, Bandt SK, et al. **Integration of resting state functional MRI into clinical practice: a large single institution experience.** *PLoS One* 2018;13:e0198349 [CrossRef Medline](#)
17. Southwell DG, Hervey-Jumper SL, Perry DW, et al. **Intraoperative mapping during repeat awake craniotomy reveals the functional plasticity of adult cortex.** *J Neurosurg* 2016;124:1460–69 [CrossRef Medline](#)

18. Wang N, Zeng W, Chen L. **SACICA: a sparse approximation coefficient-based ICA model for functional magnetic resonance imaging data analysis.** *J Neurosci Methods* 2013;216:49–61 [CrossRef](#) [Medline](#)
19. Agarwal S, Sair HI, Airan R, et al. **Demonstration of brain tumor-induced neurovascular uncoupling in resting-state fMRI at ultrahigh field.** *Brain Connect* 2016;6:267–72 [CrossRef](#) [Medline](#)
20. Hui M, Li J, Wen X, et al. **An empirical comparison of information-theoretic criteria in estimating the number of independent components of fMRI data.** *PLoS One* 2011;6:e29274 [CrossRef](#) [Medline](#)
21. Mitchell TJ, Hacker CD, Breshears JD, et al. **A novel data-driven approach to preoperative mapping of functional cortex using resting-state functional magnetic resonance imaging.** *Neurosurgery* 2013;73:969–82; discussion 982–83 [CrossRef](#) [Medline](#)