

Multiple Reproducibility Indices for Evaluation of Cognitive Functional MR Imaging Paradigms

Joseph A. Maldjian, Paul J. Laurienti, Lance Driskill, Jonathan H. Burdette

BACKGROUND AND PURPOSE: The purpose of this study was to evaluate a variety of reproducibility indices for cognitive functional MR imaging (fMRI) paradigms that account for both overlapping and extraneous regions of activation.

METHODS: Eight right-handed volunteers were imaged with fMRI by using a word-generation paradigm and a forward-backward text-listening paradigm. The paradigms were performed twice in the session and repeated 1 week later. Reproducibility indices for the four repeated studies were determined on the basis of pair-wise computation of the ratio of the probability-weighted intersection volume divided by the union volume of surviving activation clusters. The intersection volume was determined by using several iterations of the morphologic dilatation operator with additional voxels accrued in the intersection weighted by an exponential function. Computed indices included global reproducibility, language area reproducibility, extraneous activation reproducibility, and laterality.

RESULTS: The word-generation paradigm had reproducibility values that were significantly greater than those of the forward-backward text-listening paradigm (global reproducibility, 0.75 vs 0.5, $P < .005$; language area reproducibility, 0.85 vs 0.6, $P < .008$; mean extraneous activation reproducibility, 0.68 vs 0.41, $P < .002$). The forward-backward text-listening paradigm demonstrated more focal activations, whereas the word-generation demonstrated larger activations outside the dominant language areas that were highly reproducible.

CONCLUSION: For clinically relevant language paradigms, multiple reproducibility indices should be taken into account in selecting an appropriate paradigm. Compared with a forward-backward text-listening task, a word-generation task has higher reproducibility indices at the expense of localizing ability. The forward-backward paradigm demonstrates more focal activations with less extraneous activation.

Blood oxygen level–dependent (BOLD) functional MR imaging (fMRI) is a noninvasive technique for localizing regional brain signal changes in response to task performance (1, 2). The main clinical application of this technique has been the preoperative identification of eloquent cortex in patients with intracranial lesions. In general, motor and language paradigms have been used for this determination. Several studies (3–7) have been performed to validate the accuracy of fMRI for simple motor tasks in comparison with direct intraoperative stimulation. The validity of fMRI in the assessment of cognitive function, and more specifically language function, remains to be determined.

A wide variety of paradigms are used for the identification of language function with fMRI. However, the activation patterns produced by these paradigms can be distinct. The clinically used paradigms must demonstrate good reproducibility as well as good localizing power. These can be competing requirements for a paradigm. For example, a paradigm that consistently activates the entire brain demonstrates excellent reproducibility, but it would have no value in localizing function in a presurgical setting. Similarly, a paradigm that demonstrates reproducible activations outside eloquent language areas (extraneous activations) would be of limited value for clinical use. A simple approach to the determination of reproducibility would be to compute indices on the basis of intersection and union of activation clusters in serial fMRI studies. Such an approach fails to take into account the problem of localizing power, the excessive size of activation clusters, the proximity of non-overlapping clusters, and the extraneous activation areas. For example, a paradigm that demonstrates small focal activations may have excellent localization properties; how-

Received December 20, 2001; accepted after revision March 2, 2002.

From the Department of Radiology, Wake Forest University School of Medicine, Winston-Salem, NC.

Address reprint requests to Joseph A. Maldjian, MD, Department of Radiology, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC.

ever, if the clusters are non-overlapping in serial studies, the reproducibility would be poor.

No consensus exists in the literature in terms of how the reproducibility of fMRI paradigms should be evaluated or what constitutes adequate reproducibility. Prior studies of fMRI reproducibility have measured only a single index, related to some measure of overlap, and the values and measures have varied considerably. The purpose of this study was to define a set of indices for use in characterizing clinically relevant cognitive paradigms. These indices take into account reproducibility, localization power, proximity of activation clusters, and extraneous activation. We assessed the use of these measurements for the characterization of two commonly used language paradigms: a word-generation paradigm and a text-listening paradigm. Although these paradigms are known to activate different regions of the brain, the reproducibility indices are not meant to determine which paradigm is better. Rather, these indices can be used to compare the reproducibility of a particular paradigm in the activation of regions of the brain that it is expected to activate.

Methods

The subject population consisted of eight healthy right-handed volunteers (age range, 24–40 years; mean age, 28 years; five men, three women). Handedness was assessed by using the Edinburgh handedness inventory (8). All subjects were enrolled after they provided written informed consent, as approved by the institutional review board at our institution. The language paradigms included a word-generation paradigm, in which subjects were asked to silently generate words beginning with a specific letter, and a forward-backward text-listening paradigm (FBL), in which subjects listened to digitally recorded passages of text read aloud and the same text played backwards. The paradigms were delivered as blocked tasks, with alternating 30-second on and 30-second off periods for a total of five cycles (5 minutes of imaging). Each paradigm was performed twice in the session, and the subjects returned 1 week later for a second fMRI session in which the paradigms were again performed twice. The digitally recorded passages were different for each experiment, as were the letters used for word-generation to minimize habituation and memory effects. Subjects were asked to respond to a series of questions about the text passages after each FBL experiment to ensure that they were paying attention to the task. fMRI was performed as previously described (9), and image analysis was performed by using SPM99 (10, 11).

Imaging was performed with a 1.5-T human bore magnet (GE Medical Systems, Milwaukee, WI) equipped with echo-speed gradients for echo-planar imaging, and LX 8.3 software. Foam cushions were used to immobilize the patient's head within the coil to minimize motion degradation. Imaging consisted of spoiled GRASS (SPGR; GE Medical Systems) sagittal localization, followed by a T1-weighted acquisition of the entire brain performed in the axial plane (24 cm FOV, 256×256 matrix, 3-mm section thickness). This sequence was used for both anatomic overlays of the functional data and spatial normalization of the data sets to a standard atlas. Functional imaging was performed in the axial plane by using multisection gradient-echo echo-planar imaging (24×15 -cm FOV, 64×40 matrix, 5-mm thickness, no skip, 28 sections, 2000/40 TR/TE).

The T1-weighted images were normalized to a standard template in Montreal Neurological Institute (MNI) coordinate space in SPM99. The functional data sets were motion cor-

rected (intrarun realignment) in SPM99 by using the first image as the reference. The functional data sets were normalized to MNI space by using image header information to determine the 16-parameter affine transform between the functional data sets and the T1-weighted images (5), in combination with the transform computed in SPM99 for the T1-weighted anatomic images to MNI space. The normalized data sets were resampled to $4 \times 4 \times 4$ mm in MNI space by using sinc interpolation. A second realignment step (interrun realignment) was then performed (in SPM99) between successive normalized runs for each subject, by using the initial normalized run as the target. This step was performed to eliminate motion between the successive runs within each subject. The data sets were smoothed by using a $8 \times 8 \times 10$ -mm full width at half maximum gaussian smoothing kernel, and SPMs were generated by using the general linear model in SPM99. A 6-second time-shifted boxcar waveform was used as the reference paradigm, and the analysis of covariance (ANCOVA) model with global activity as a confound was used for the statistical analysis. Temporal smoothing, detrending and high-pass filtering were performed as part of the SPM analysis. Thresholds for individual SPM{t} values were set at $P < .001$.

A second-level analysis was performed to generate group SPMs by using a random-effects model in SPM99 with the individual contrast maps (12). Thresholds for the resulting maps were set at a corrected $P < .001$, by using a stringent Bonferroni alpha correction ($T = 7.22$, corresponding to an uncorrected $P < 2 \times 10^{-8}$), as implemented in SPM99. Anatomic regions were automatically defined by using an anatomic MR imaging atlas (13) that we previously normalized to the same MNI SPM template for use with our fMRI data. MNI coordinates were converted into the Talairach coordinate system (14) by using a nonlinear transform (15). Brodmann areas (BA) were determined for activation maxima by using the Talairach Daemon (16, 17). Additional anatomic and lobar atlases (ie, temporal lobe, frontal lobe) were defined by using the labels and coordinates provided by the Talairach Daemon, which we have previously interrogated to generate complete volume masks normalized to the MNI SPM template.

Reproducibility indices for the four repeated studies (for each paradigm) were determined on the basis of pair-wise computation of the ratio of the probability-weighted intersection volume (I) divided by the union volume (U) of surviving activation clusters ($P < .001$). The union was computed as the sum of all surviving voxels for each pair-wise comparison (a value of 1 was assigned to any activated pixels between the studies). To account for differences in cluster morphology, a probability-weighted intersection volume was computed for the pair-wise comparisons by using an exponential weighting function emanating from the original intersection map obtained by using the morphologic dilatation operator with a unary 3×3 opening filter (18). Each iteration of the morphologic dilatation operator effectively grows the surviving activation clusters by 1 voxel in both dimensions in-plane. This procedure takes into account both differences in cluster morphology and the problem of localized but non-overlapping clusters. Non-overlapping clusters would normally return values of 0 for reproducibility despite their proximity. Application of the dilatation operator provides a means for taking into account incomplete or non-overlapping clusters in close proximity. The probability weighting function was defined as $(1/2)^n$, where n represents each iteration of the dilatation operation. With each iteration of the dilatation operator, the total number of additional voxels surviving the intersection of clusters were multiplied by the weighting function and added to the running total for intersection volume. Five iterations of this procedure were performed (as the probability function decreases to $< .001$ at $n = 5$).

This iterative computational process is depicted in Figures 1 and 2. In this manner, map-wise reproducibility indices were computed as the mean of the pair-wise I/U ratios for the four

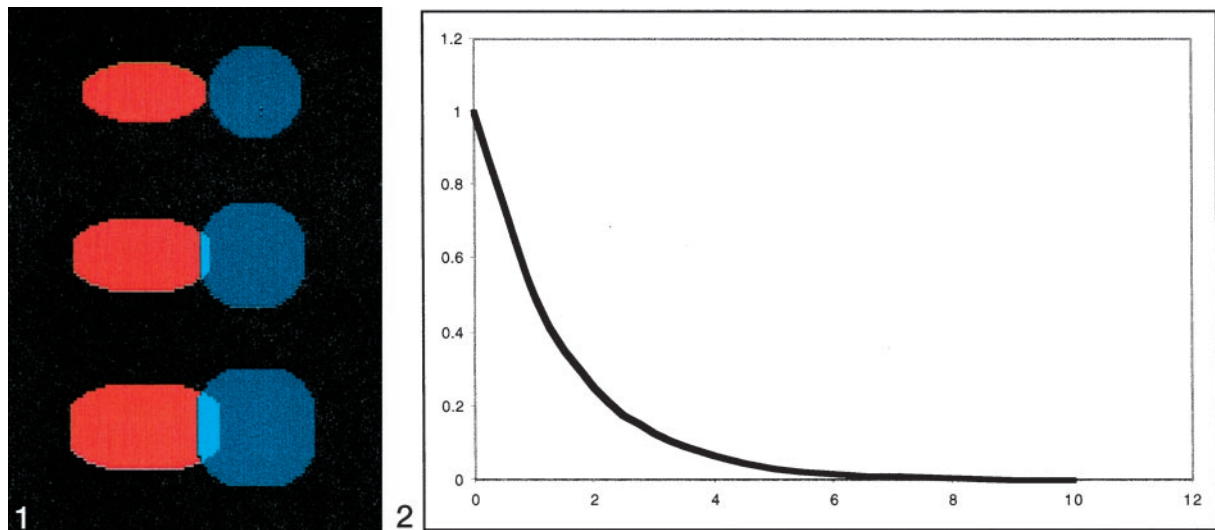


FIG 1. Successive iterations of morphologic dilatation operator on two simulated clusters.

Row 1 (from the top) demonstrates two non-overlapping clusters of different shapes.

Row 2 demonstrates growth of each cluster by using a morphologic dilatation operator that created an overlap region while maintaining the native shape of each cluster.

Row 3 demonstrates increasing overlap with repeated iterations of dilatation operator. During each iteration, additional areas of overlap are weighted by using an exponential weighting function and added to the total overlap volume.

FIG 2. Exponential weighting function. Additional overlapping clusters from dilatation operation are weighted on the basis of the iteration.

runs. Although this process provides indices for overall activation reproducibility, it does not take into account the degree of confounding or extraneous activation. For example, a paradigm that activates large areas has excellent reproducibility indices, but it provides limited information for presurgical localization. The extreme example of this is a paradigm that activates all the voxels in the brain, providing excellent reproducibility, but this paradigm would not be of any value in the localization of function. To refine the characterization of the paradigms, language-area activation reproducibility indices were computed by masking the data for the left frontal and temporal lobes, and extraneous activation reproducibility indices were computed on the basis of the remaining area outside the masked left frontal and temporal lobes. The masked regions were defined in an automated fashion by using Talairach Daemon-based anatomic atlases normalized to the MNI SPM template.

For each paradigm, reproducibility indices were generated for global reproducibility, language-area reproducibility, and extraneous activation reproducibility. Each reproducibility index was defined as the mean of the pair-wise intersection-union ratios for the four runs. In addition, mean laterality indices and extraneous activation indices were computed. The extraneous activation index was computed as EA/TA , where EA represents the total number of activated voxels in nonlanguage areas, and TA represents the total number of activated voxels in the map. To compute the laterality index for each subject, the four sessions of each paradigm were combined into a fixed-effects analysis in SPM99. Thresholds for the resulting maps were set at an alpha-corrected $P < .001$ and separately masked for the frontal lobes and temporal lobes. The laterality index was defined as $(LH - RH)/(RH + LH)$, where RH represents activated voxels in the right hemisphere, and LH represents activated voxels in the left hemisphere. The mean and standard deviation of these values was computed across all sessions for each subject. A paired Student t test was performed to compare the mean reproducibility indices, laterality indices, and extraneous activation indices between the word-generation paradigm and the FBL paradigm.

Results

All subjects were able to perform the tasks for all the sessions. After the FBL paradigms were completed, all subjects were able to correctly respond to a series of questions specific to the text paradigm with 90% or greater accuracy. The group map for the word-generation paradigm demonstrated larger, more distributed areas of activation than those of the FBL paradigm (Figs 3, 4). Areas of activation on the word-generation group map included bilateral frontal lobes (left > right), anterior cingulate, left basal ganglia, left thalamus, and midbrain structures, as well as bilateral areas in the cerebellum. All of these areas were highly significant, surviving stringent corrections for multiple comparisons, demonstrating peak Z scores in excess of 8. The FBL paradigm produced activation in bilateral temporal lobes (left > right), the left frontal lobe, and the right cerebellum. These activations were much more focal than those of the word-generation paradigm. Although the activations were also highly significant, surviving stringent corrections for multiple comparisons, peak Z scores were less than those for the word-generation paradigms (scores < 7).

For the word-generation group map, peak activity in the left frontal lobe was in the left superior frontal gyrus (BA 6, Talairach coordinates = $[-4, 14.2, 50]$, $Z = 8.5$). Peak activity in the right frontal lobe involved the right insula and right inferior frontal gyrus (BAs 13 and 47, Talairach coordinates = $[43.6, 15.5, -0.8]$, $Z = 7.16$).

For the FBL group map, the peak activity in the left frontal lobe was in the inferior frontal gyrus (BA 9, Talairach coordinates = $[-51.5, 16.7, 22.2]$, $Z =$

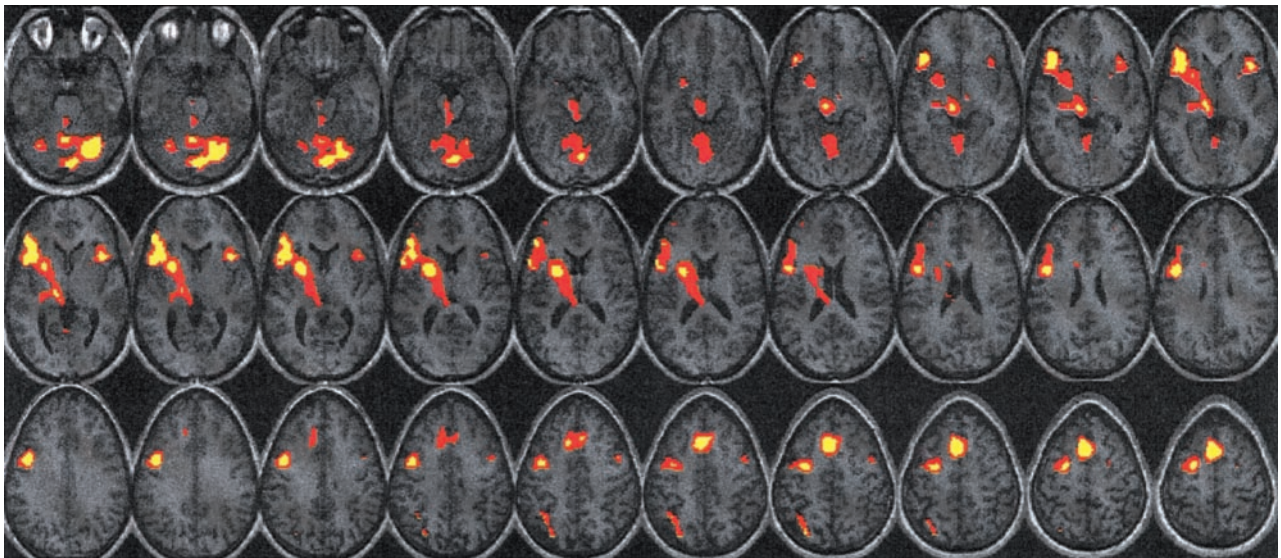


FIG 3. Group map for the word-generation paradigm ($P < .001$, height corrected). Images displayed in Talairach space (*right* side of the image is the *right* side of the subject). Bilateral frontal lobe activation is demonstrated (*left* > *right*), as well as distributed activations in the left basal ganglia, thalamus, midbrain, superior frontal lobe, anterior cingulate region, and bilateral cerebellum.

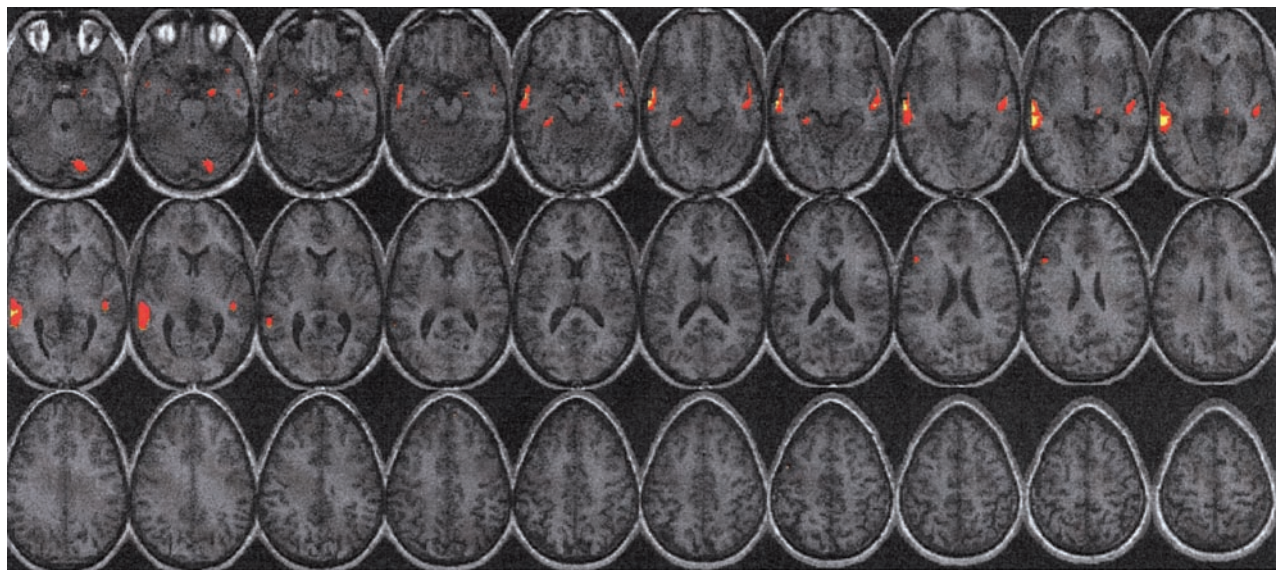


FIG 4. Group map for the FBL paradigm ($P < .001$, height corrected). Images displayed in Talairach space (*right* side of the image is the *right* side of the subject). Bilateral temporal lobe activation is demonstrated (*left* > *right*), and a focal area of activity in the left inferior frontal gyrus (BA 9) is present. Activation is more focal and less spatially distributed than with the word-generation paradigm. A small area of right cerebellar activity is also present.

6.23). Peak activity in the left temporal lobe was in the superior temporal gyrus (BA 21, Talairach coordinates = $[-63.4, -19.6, -3.2]$, $Z = 7.2$). Peak activity in the right temporal lobe was in the middle temporal gyrus (BA 21, Talairach coordinates $[51.5, -19.8, -7.4]$, $Z = 6.77$).

Reproducibility Indices

The mean global reproducibility indices for the word-generation and language paradigms were 0.75 (SD = 0.13) and 0.5 (SD = 0.14), respectively, with $P < .005$ (Table 1). The mean language area reproducibility indices for the word-generation and language paradigms were 0.85 (SD = 0.12) and 0.6

TABLE 1: Global reproducibility values*

Subject	Word-Generation	FBL
1	0.865517	0.577685
2	0.597006	0.450596
3	0.725085	0.599033
4	0.759399	0.696469
5	0.623490	0.374860
6	0.987388	0.434296
7	0.730123	0.258259
8	0.706395	0.570180
Mean (SD)	0.749 (0.127)	0.495 (0.141)

* $P < .005$

TABLE 2: Language area reproducibility values*

Subject	Word-Generation	FBL
1	0.994310	0.781844
2	0.657320	0.432581
3	0.841645	0.719868
4	0.842682	0.771116
5	0.768336	0.560528
6	0.959938	0.523434
7	0.987398	0.324236
8	0.777632	0.659780
Mean (SD)	0.854 (0.12)	0.597 (0.165)

* $P < .008$ **TABLE 3: Extraneous area reproducibility values***

Subject	Word-Generation	FBL
1	0.792541	0.433713
2	0.563576	0.424092
3	0.638156	0.472435
4	0.711660	0.617451
5	0.557372	0.269693
6	0.955450	0.380242
7	0.601508	0.213573
8	0.655076	0.491226
Mean (SD)	0.684 (0.134)	0.413 (0.127)

* $P < .002$ **TABLE 4: Laterality indices***

Subject	Word-Generation	FBL
1	0.282255	0.251434
2	0.112344	0.177413
3	0.377311	0.375264
4	0.237336	0.122048
5	0.235070	0.205566
6	0.256373	0.153030
7	0.313335	0.0541177
8	0.371002	0.210081
Mean (SD)	0.273 (0.085)	0.194 (0.095)

* $P < .06$

(SD = 0.17), respectively, with $P < .008$ (Table 2). The mean extraneous activation reproducibility indices for the word-generation and language paradigms were 0.68 (SD = 0.13) and 0.41 (SD = 0.13), respectively, with $P < .002$ (Table 3).

Laterality and Extraneous Activation Indices

The mean laterality indices for the word-generation and language paradigms were 0.27 (SD = 0.08) and 0.19 (SD = 0.1), respectively, with $P < .06$ (Table 4). The mean extraneous activation indices for the word-generation and language paradigms were 0.71 (SD = 0.04) and 0.63 (SD = 0.05), respectively, with $P < .004$ (Table 5).

Repeated Sessions in a Single Subject

Figure 5 demonstrates activation maps generated in a representative subject for the word-generation and FBL paradigms performed 1 week apart. The word-genera-

TABLE 5: Extraneous activation indices*

Subject	Word-Generation	FBL
1	0.734886	0.609496
2	0.742806	0.581206
3	0.651104	0.632916
4	0.773129	0.704500
5	0.741684	0.631687
6	0.668879	0.568074
7	0.664156	0.634116
8	0.707600	0.686904
Mean (SD)	0.711 (0.0446)	0.631 (0.047)

* $P < .004$

tion paradigm demonstrates a higher reproducibility with a large number of extraneous activations. The FBL paradigm demonstrates a lower reproducibility, but with less extraneous activations. Although the word-generation paradigm appears highly reproducible, the clinical utility (eg, for presurgical mapping), would be limited because of the extensive extraneous areas of significant activation. On the basis of the assumption that areas of substantial activation on the word-generation map are necessary for the task, almost no areas in either hemisphere can be safely resected without inducing a language deficit.

Discussion

In this study, we demonstrated that two commonly used paradigms to study language processing demonstrate good reproducibility over four test-retest sessions. However, the demonstration of reproducibility alone is insufficient to evaluate the utility of a paradigm for clinical use. For preoperative mapping, the ability to localize function is important. Paradigms that demonstrate large distributed areas of activation or areas of extraneous activation are not of much utility in preoperative localization. In this study, we called areas outside the dominant left temporal and left frontal lobes areas of extraneous activation. In fact, these areas may be relevant to language function; however, in keeping with the current models of language function (19) and the lack of electrical stimulation data on these extraneous sites, the assumption that they are not necessary for language function is likely justified. Although findings from previous fMRI studies (20–23) suggest bilateral representation of language in female subjects, the functional relevance of activations outside the dominant hemisphere has not been determined. Although we included equal distributions of male and female subjects, our sample sizes were too small to make any statistically relevant comparisons of sex-related differences. However, no significant differences were demonstrated between male subjects and female subjects in our group in terms of extraneous activation indices and laterality indices with either paradigm.

Prior studies (24–32) have been performed in attempts to validate the accuracy and reliability of fMRI by using cognitive and motor paradigms; however, no consistent method for determining reproduc-

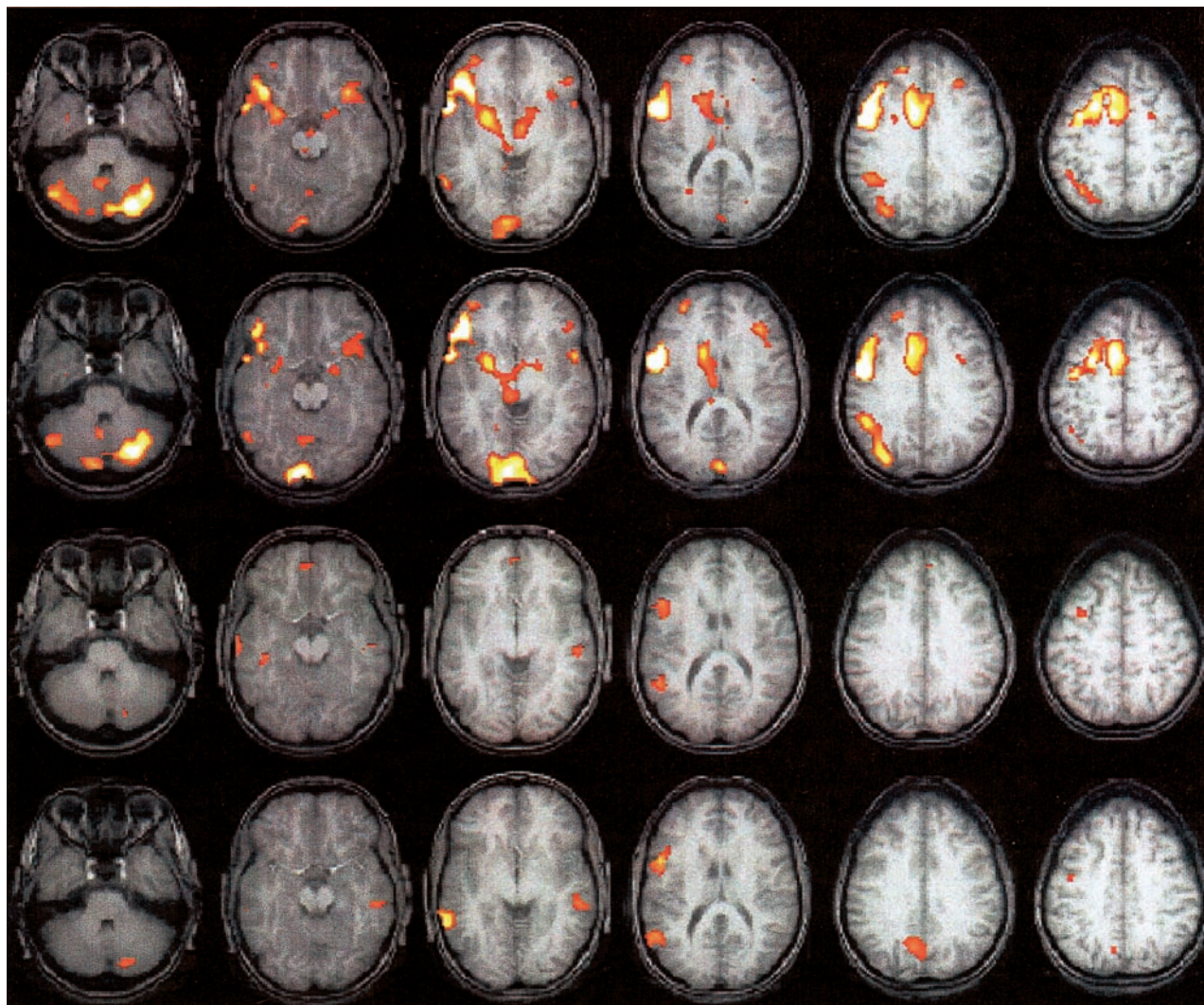


FIG 5. Word generation and FBL activation maps in one subject obtained 1 week apart ($P < .0001$ corrected for spatial extent at $P < .05$). Images displayed in Talairach space (right side of the image is the right of the subject).

Row 1 (from the top) Word-generation paradigm on day 1. Images demonstrate bilateral frontal, bilateral cerebellar, bilateral thalamic, bilateral occipital left parietal, and anterior cingulate regions.

Row 2, Word-generation paradigm performed 1 week later. Although the activations are highly reproducible, numerous areas of extraneous activation are also present.

Row 3, FBL paradigm on day 1. Images demonstrates activation of bilateral temporal and left frontal regions.

Row 4, FBL paradigm performed 1 week later. Although the reproducibility indices for this paradigm are lower, the degree of extraneous activation is also markedly reduced.

ibility exists. We tried to show that a variety of indices are necessary to characterize cognitive paradigms. We then demonstrated the use of these indices for two commonly used language paradigms. The paradigms we choose come from the general category of language paradigms; however, they are known to activate different regions, and the purpose was not to demonstrate that one paradigm should be used instead of another. Rather, these reproducibility indices can be used to determine if a particular paradigm is better than another at activating the regions it is designed to activate.

In a study examining the reliability of functional MR word-generation tasks, Brannen et al (30) reported a test-retest precision of 49% with the use of a concurrence ratio between repeated iteration of the

task. A selective region of the brain was examined (left inferior frontal gyrus) for computation of the concurrence ratios, and the analysis was limited to either one section demonstrating maximal activation or five sections encompassing the left inferior frontal region. This method enables a limited evaluation of the paradigm as a whole because it fails to take into account distributed activation areas that may or may not be reproducible. Invasive electrocortical stimulation (ECS) was also performed in a subset of the patients; good agreement was demonstrated between language function identified with ECS and that identified with fMRI. No coregistration of the fMRI with the intraoperative stimulation sites was attempted, however, and the agreement was based solely on BAs determined intraoperatively by visual inspection. In a study by

FitzGerald et al (32), fMRI findings with language paradigms were compared with ECS results in 11 patients. They found that, in 85% of comparisons, fMRI activation was localized to within 1 cm of the ECS-determined area of language function. These studies provide evidence that selected areas of activation on word-generation studies correspond to language areas identified with ECS and that selected frontal lobe activation areas are reproducible to some degree. These studies do not address the question of overall paradigm reliability and the relative utility of language paradigms in the preoperative identification of language eloquent cortex. Furthermore, no consensus exists about how the reliability or accuracy of fMRI paradigms should be determined. This lack of agreement is exemplified by the disparate methods based on some arbitrarily defined metric that are used in these studies to determine reproducibility or reliability.

Our study provides several surprising results. Both language tasks demonstrated good global reproducibility and similar laterality localizing ability. However, the word-generation task consistently had significantly higher indices for global reproducibility and language area reproducibility. Using these indices, one might prematurely conclude that the word-generation paradigm is superior to the language paradigm in terms of reproducibility. However, the clinical utility of the paradigm is also influenced by the localizing ability of the paradigm for the desired function. In general, the word-generation paradigm demonstrates larger and more distributed activations, including areas contralateral to the dominant hemisphere. Although we called these activations outside the dominant hemisphere extraneous, they are nevertheless highly reproducible. These activations may be involved in cognitive processes related to language function, such as memory function, or higher-level association areas; however, their true nature remains unknown at this time. This observation is especially true in areas of reproducible activation in the left thalamus, cerebellum, midbrain, and anterior cingulate.

Although the FBL paradigm had lower reproducibility values, the extraneous activation index was also significantly lower. Also, the FBL paradigm revealed significant activation in the dominant left temporal and left frontal lobes, which presumably corresponded to Wernicke's and Broca's areas, respectively. The word-generation paradigm did not activate the dominant temporal lobe. The FBL paradigm demonstrated bilateral temporal lobe activation in homologous areas, although the peak activation for the left temporal lobe extended more superiorly to the superior temporal gyrus. The right-sided temporal lobe activation is unlikely to have been caused by auditory activation, because auditory stimulation was present throughout the paradigm, and the nature of the paradigm ensured that the tonal and frequency content was matched between the baseline and active epochs. The right-sided temporal lobe activity may represent nondominant language-related areas involved in text comprehension. The focal left frontal lobe activation for the FBL paradigm overlapped the large activation cluster in the left frontal lobe identified with

the word-generation paradigm. However, the activation maxima within the cluster for the word-generation paradigm included more inferior, as well as superior, regions. These results indicate that, despite the lower reproducibility indices, the FBL paradigm is more clinically relevant because it provides equivalent lateralizing ability to the word-generation paradigm, more focal activations, and a significantly lower extraneous activation index. It also activates regions in both dominant temporal and frontal lobes that are known to be involved in language function. This determination is based on the assumption that the extraneous activation areas are not necessary for language function. Although this assumption is in keeping with current models of language function, the high degree of reproducibility associated with these activations raises significant questions regarding their nature and warrants further investigation.

The results of this study do not mean that the FBL paradigm is a better language paradigm than the word-generation paradigm. The choice of these two particular paradigms was somewhat arbitrary. For the purposes of this investigation, we could have compared the word-generation paradigm with a working-memory paradigm. In this study, we proposed a set of indices that can be used to characterize the reproducibility and reliability of clinically relevant fMRI cognitive paradigms in general. Using these indices, we demonstrated that, for what each paradigm is designed to do (ie, activate language specific areas while minimizing false positive activation), the FBL performs better than the word-generation paradigm. Among the category of paradigms expected to activate Broca area, the word-generation paradigm may very well be best.

Conclusion

Previous attempts at evaluating the reproducibility of cognitive paradigms in fMRI have focused on expected locations of activation and ignored regions of extraneous activation. We demonstrate the use of a set of indices that enable more complete evaluation of the reproducibility of clinically relevant paradigms. Importantly, these indices take into account cluster morphology, cluster proximity, and localizing ability.

Acknowledgement

The authors thank Pete Santago for valuable discussions on determining cluster overlap in serial studies.

References

1. Ogawa S, Lee TM, Kay AR, Tank DW. **Brain magnetic resonance imaging with contrast dependent on blood oxygenation.** *Proc Natl Acad Sci U S A* 1990;87:9868-9872
2. Le Bihan D, Jezzard P, Haxby J, Sadato N, Rueckert L, Mattay V. **Functional magnetic resonance imaging of the brain.** *Ann Intern Med* 1995;122:296-303
3. Jack CR Jr, Thompson RM, Butts RK, et al. **Sensory motor cortex: correlation of presurgical mapping with functional MR imaging and invasive cortical mapping.** *Radiology* 1994;190:85-92
4. Puce A, Constable RT, Luby ML, et al. **Functional magnetic resonance imaging of sensory and motor cortex: comparison with electrophysiological localization.** *J Neurosurg* 1995;83:262-270

5. Maldjian JA, Schulder M, Liu WC, et al. **Intraoperative functional MRI using a real-time neurosurgical navigation system.** *J Comput Assist Tomogr* 1997;21:910–912
6. Schulder M, Maldjian JA, Liu WC, Mun IK, Carmel PW. **Functional MRI-guided surgery of intracranial tumors.** *Stereotact Funct Neurosurg* 1997;68:98–105
7. Schulder M, Maldjian JA, Liu WC, et al. **Functional image-guided surgery of intracranial tumors located in or near the sensorimotor cortex.** *J Neurosurg* 1998;89:412–418
8. Oldfield RC. **The assessment and analysis of handedness: the Edinburgh inventory.** *Neuropsychologia* 1971;9:97–113
9. Maldjian JA, Gottschalk A, Patel RS, Pincus D, Detre JA, Alsop DC. **Mapping of secondary somatosensory cortex activation induced by vibrational stimulation: an fMRI study.** *Brain Res* 1999; 824:291–295
10. Friston K, Holmes A, Worsley K, Poline J, Frith C, Frackowiak R. **Statistical parametric maps in functional imaging: a general approach.** *Human Brain Mapping* 1995;2:189–201
11. Friston K, Ashburner J, Poline J, Frith C, Heather J, Frackowiak R. **Spatial registration and normalization of images.** *Human Brain Mapping* 1995;2:165–189
12. Holmes A, Friston K. **Generalisability, random effects and population inference.** *NeuroImage* 1998;7:S754
13. Kikinis R, Shenton M, Iosifescu D, et al. **A digital brain atlas for surgical planning, model driven segmentation and teaching.** *IEEE Trans Visualization Comput Graphics* 1996;2:2323–2241
14. Talairach J, Tournoux P. **Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System—An Approach to Cerebral Imaging.** New York, NY: Thieme Medical Publishers; 1988
15. Duncan J, Seitz RJ, Kolodny J, et al. **A neural basis for general intelligence.** *Science* 2000;289:457–460
16. Lancaster JL, Summerlin JL, Rainey L, Freitas CS, Fox PT. **The talairach daemon: a database server for talairach atlas labels.** *NeuroImage* 1997;5:S633
17. Lancaster JL, Woldorff MG, Parsons LM, et al. **Automated talairach atlas labels for functional brain mapping.** *Hum Brain Mapp* 2000;10:120–131
18. Haralick RM, Sternberg SR, Zhuang X. **Image analysis using mathematical morphology.** *IEEE Trans Pattern Analysis Machine Intell* 1987;9:532–550
19. Price CJ. **The anatomy of language: contributions from functional neuroimaging.** *J Anat* 2000;197:335–359
20. Kansaku K, Yamaura A, Kitazawa S. **Sex differences in lateralization revealed in the posterior language areas.** *Cerebral Cortex* 2000; 10:866–872
21. Shaywitz BA, Shaywitz SE, Pugh KR, et al. **Sex differences in the functional organization of the brain for language.** *Nature* 1995;373: 607–609
22. Vikingstad EM, George KP, Johnson AF, Cao Y. **Cortical language lateralization in right handed normal subjects using functional magnetic resonance imaging.** *J Neurol Sci* 2000;175:17–27
23. Phillips MD, Lowe MJ, Lurito JT, Dzemidzic M, Mathews VP. **Temporal lobe activation demonstrates sex-based differences during passive listening.** *Radiology* 2001;220:202–207
24. Yetkin FZ, McAuliffe TL, Cox R, Houghton VM. **Test-retest precision of functional MR in sensory and motor task activation.** *AJNR Am J Neuroradiol* 1996;17:95–98
25. Rombouts SA, Barkhof F, Hoogenraad FG, Sprenger M, Valk J, Scheltens P. **Test-retest analysis with functional MR of the activated area in the human visual cortex.** *AJNR Am J Neuroradiol* 1997;18:1317–1322
26. Manoach DS, Halpern EF, Kramer TS, et al. **Test-retest reliability of a functional MRI working memory paradigm in normal and schizophrenic subjects.** *Am J Psychiatry* 2001;158:955–958
27. Loubinoux I, Carel C, Alary F, et al. **Within-session and between-session reproducibility of cerebral sensorimotor activation: a test-retest effect evidenced with functional magnetic resonance imaging.** *J Cereb Blood Flow Metab* 2001;21:592–607
28. Binder JR, Swanson SJ, Hammeke TA, et al. **Determination of language dominance using functional MRI: a comparison with the Wada test.** *Neurology* 1996;46:978–984
29. Binder JR, Frost JA, Hammeke TA, et al. **Human temporal lobe activation by speech and nonspeech sounds.** *Cereb Cortex* 2000;10: 512–528
30. Brannen JH, Badie B, Moritz CH, Quigley M, Meyerand ME, Houghton VM. **Reliability of functional MR imaging with word-generation tasks for mapping broca's area.** *AJNR Am J Neuroradiol* 2001;22:1711–1718
31. Benson RR, FitzGerald DB, LeSueur LL, et al. **Language dominance determined by whole brain functional MRI in patients with brain lesions.** *Neurology* 1999;52:798–809
32. FitzGerald DB, Cosgrove GR, Ronner S, et al. **Location of language in the cortex: a comparison between functional MR imaging and electrocortical stimulation.** *AJNR Am J Neuroradiol* 1997;18: 1529–1539