FUNCTIONAL

Simple fMRI Postprocessing Suffices for Normal Clinical Practice

S. González-Ortiz, L. Oleaga, T. Pujol, S. Medrano, J. Rumí, L. Caral, T. Boget, J. Capellades, and N. Bargalló

ABSTRACT

BACKGROUND AND PURPOSE: Whereas fMRI postprocessing tools used in research are accurate but unwieldy, those used for clinical practice are user-friendly but are less accurate. We aimed to determine whether commercial software for fMRI postprocessing is accurate enough for clinical practice.

METHODS: Ten volunteers underwent fMRI while performing motor and language tasks (hand, foot, and orolingual movements; verbal fluency; semantic judgment; and oral comprehension). We compared visual concordance, image quality (noise), voxel size, and radiologist preference for the activation maps obtained by using Neuro3D software (provided with our MR imaging scanner) and by using the SPM program commonly used in research.

RESULTS: Maps obtained with the 2 methods were classified as “partially overlapping” for 70% for motor and 72% for language paradigm experiments and as “overlapping” in 30% of motor and in 15% of language paradigm experiments.

CONCLUSIONS: fMRI is a helpful and robust tool in clinical practice for planning neurosurgery. Widely available commercial fMRI software can provide reliable information for therapeutic management, so sophisticated, less widely available software is unnecessary in most cases.

ABBREVIATIONS: LI = laterality index; Tmax = maximal T; Tth = threshold T

fMRI is increasingly being used in the clinical setting. Whereas conventional MR imaging provides a structural view of the brain, fMRI enables the functional assessment of the regions responsible for sensory, motor, cognitive, and affective processes in both healthy patients and in patients with neurologic disease. Therefore, fMRI allows direct correlation between function and the anatomic structures responsible for this function, combining spatial and temporal resolution.

Although there are several potential clinical applications of fMRI (follow-up of the recovery of neural functions after stroke or head injuries, assessment of seizure disorders, monitoring of the effects of drugs, etc), the main clinical application of fMRI to date has been the location and evaluation of eloquent brain areas in planning surgery for brain pathology. Locating the brain area responsible for critical functions such as language, memory, or motor function is crucial in surgical planning, because although brain organization was traditionally thought to be reproducible among persons, it is now known to vary widely.1

One of the disadvantages of fMRI as a clinical tool is the time required for postprocessing. Experimental studies usually use sophisticated software for postprocessing, most commonly SPM (Wellcome Department of Imaging Neuroscience, London, UK) and FSL (http://www.fmrib.ox.ac.uk/fsl), though many programs are available for medical image analysis. However, postprocessing with these programs is too laborious for clinical practice, so image analysis software by the scanner manufacturer is commonly used. fMRI postprocessing involves 2 steps: the first, called image preprocessing, basically consists of movement corrections, section alignment, coregistration of structural and functional images, and, when necessary for grouped comparisons, image normalization to a standard brain; the second step basically consists of the statistical analysis of the signal changes. Research tools such as SPM and FSL allow interventions in every step and usually provide more information about the results of statistical analyses. This information is essential to rule out false-positive or false-negative activations. On the contrary, the image analysis software provided by manufacturers aims to be “user friendly” and is gen-
eraly more closed to intervention during the process. This is probably why manufacturers’ image analysis software is rarely used for scientific studies and why the results of studies done with research tools are often difficult to translate to clinical practice.

We aimed to determine whether fMRI postprocessing with the software provided by the scanner manufacturer was useful for clinical decision making. To this end, we compared Neuro3D (Siemens, Erlangen, Germany) with SPM in a wide variety of clinically useful paradigms.

MATERIALS AND METHODS

Participants
Ten healthy right-handed volunteers (4 men and 6 women; mean age, 29.3 years; age range, 25–50 years) without intracranial injuries or physical or neurologic diseases and with a high educational level underwent fMRI while performing a total of 63 tasks (11 related to verbal fluency, 11 to semantic judgment, 11 to oral comprehension, 6 to right-hand motor function, 6 to left-hand motor function, 6 to right-foot motor function, 6 to left-foot motor function, and 6 to orolingual motor function).

fMRI Experiments
We used a block-design experiment in which activation tasks were alternated with a period of rest. To study motor function, we performed 3 experiments: 1) finger tapping (left hand followed by a rest period and right hand followed by a rest period), 2) flexo-extension movements of the toes (of the left foot followed by a rest period and of the right foot followed by a rest period), and 3) orolingual movements (movements of the tongue and lips alternated with a rest period). To study language function, we had participants perform tasks related to verbal fluency (subvocalizing words beginning with the letters “F,” “A,” and “S,” followed by a rest period), semantic judgment (subvocalizing words related to bedroom, vehicle, and school, alternated with a rest period), and to oral comprehension (listening to short stories; in this paradigm, the rest period consisted of listening to the same stories but reading backward with the aim of overriding the activation of the primary auditory areas; after fMRI, participants were asked about the story to check comprehension). Each activation task was performed for 30 seconds, followed by 30 seconds of rest. Every session (baseline rest task followed by the function-activation task) was repeated 3 times. The duration of each experiment was 3 minutes. These paradigms are commonly used in our center and were chosen based on our clinical experience, neuropsychological tests, and the literature.

Image Analysis

SPM Analysis. We used SPM5 software for postprocessing and image analysis. Initially, the images were realigned, normalized (Montreal Neurological Institute template), and smoothed. In the statistical analysis, we used a family-wise error corrected for multiple comparisons (threshold $P < .05$). If no activation was found, the threshold was lowered to $P = .005$ uncorrected. In each paradigm, the Tmax and Tth were obtained. The Tmax corresponds to the anatomic area that has the highest T value (activation value), and the Tth corresponds to the T value below which no significant activation occurs, which depends on the $P$ value used in each individual analysis. Structural T1 3D and functional image coregistration were obtained. We reconstructed serial axial images with similar characteristics to those provided by Neuro3D, by using the images from SPM’s display tool. The automated anatomic labeling tool was used to represent the areas of activation from an anatomic perspective.

Neuro3D Analysis (Workstation MR Imaging Software). We used the software installed on the MR imaging workstation (Neuro3D) to process the images. This software automatically coregisters the anatomic and functional images obtained in each participant and applies a spatial filter for smoothing and movement correction. The statistical analysis provided a general linear model. The only option that can be changed by the user is the T value. This software does not show whether the T value is corrected by multiple comparison or how many clusters are observed in the activation area.

First, we used the Tth value previously obtained with SPM to obtain a comparable set of axial images for each paradigm. Then, to be more restrictive, we obtained another set of axial images by using the Tmax previously obtained with SPM as the Tth. This step enabled us to compare 2 ways of postprocessing to determine which would be better in clinical practice.

Image Comparison

Two radiologists, a senior neuroradiologist with expertise in fMRI and a junior neuroradiologist, assessed the studies.

First, to assess interobserver agreement, each radiologist evaluated the studies individually and the intraclass correlation coefficient was calculated for Tth; noise; Tmax; cluster size; preference for all of the paradigms; and laterality for verbal fluency, semantic judgment, and oral comprehension paradigms.

Three weeks later, the 2 radiologists evaluated the studies jointly and reached a consensus about discrepancies. Although similar axial reconstructed images were obtained with the 2 methods (SPM and Neuro3D), blinded reading between the 2 methods was not possible because the image presentation is different.

Comparison of Motor Paradigm Experiments

We visually compared the activation maps obtained by SPM with those obtained by the Tth Neuro3D. We classified paired maps as “overlapping” (when both activation maps coincided totally), “partially overlapping” (when there was partial coincidence in the 2 activation maps), or “not overlapping” (when the 2 maps were different). We also assessed the presence of areas of activation outside of the brain parenchyma or unexpected activation...
The intraclass correlation coefficient and the confidence interval at the 95% level were Tth 0.63 (0.45–0.76), noise 0.68 (0.52–0.79), Tmax 0.83 (0.74–0.89), cluster size 0.62 (0.44–0.75), preference 0.57 (0.38–0.71), verbal fluency 1 (1–1), semantic judgment 0.77 (0.36–0.93), and oral comprehension 0.93 (0.77–0.98). The mean overall intraclass correlation coefficient was 0.75 ± 0.15.

Motor Experiments
Table 1 shows the mean and range Tth and Tmax values obtained with SPM for all participants for each motor paradigm. The Tth was the same in both hand motor tasks (6.2) and similar in the foot motor task, (4.7 in the right foot and 4.1 in the left). The Tmax was higher during the hand motor task (13.3 left, 12.1 right) and lower during the foot motor task (7.78 right, 5.9 left). In the orolingual task, Tth was similar to the foot motor task, but the Tmax was closer to the hand task.

The highest Tth and Tmax values were observed in the area activated by the hand motor task. By contrast, the lowest Tth values were observed in the area activated in the foot motor task.

Table 2 summarizes the results of the comparison between analyses by use of Neuro3D and SPM. When the comparison was based on the functional maps generated with the Tth values, all of the results obtained by the 2 methods as the threshold were classified as “overlapping” (Fig 1) or “partially overlapping” (Fig 2) (30% and 70%, respectively) in all of the motor paradigms. “Noise” was more common in the Tth Neuro3D–processed images; however, it did not interfere with the interpretation of the results except in 2 cases: 1 during the orolingual motor task and 1 during the flexo-extension movements of toes of the left foot. When the comparison was based on the functional maps created with the Tmax, the 2 methods were more similar. Cluster size was often larger (in 70% of cases) in the Tth Neuro3D postprocessing analysis. The readers preferred the functional maps obtained with SPM in 70% of the studies.

Language Experiments
Table 3 shows the mean and range of the Tth and Tmax values obtained with SPM for all participants in the language paradigm experiments. The Tth and Tmax were less variable between different paradigms than on the motor tasks, though they were somewhat lower in the verbal comprehension task. In general, Tth and Tmax values were lower in language paradigm tasks than in motor tasks.

Table 4 summarizes the comparison between Neuro3D and SPM analyses. The functional maps obtained with the 2 methods (Tth Neuro3D and SPM) were classified as “partially overlapping” in 72.7% of cases, as “overlapping” in 12.1%, and as “not overlapping” in 15.2%. The paradigm with the most results classified as “not overlapping” (3 participants) was the “verbal fluency” paradigm, whereas the paradigm with the most concordant
results (2 “overlapping” and 9 “partially overlapping”) was the “oral comprehension” paradigm.

In approximately 20% of the word-generation paradigms (verbal fluency and semantic judgment), Tth Neuro3D functional maps had noise that made it difficult to interpret the areas activated (in 4 of the “verbal fluency” and in 3 of the “semantic judgment”). However, no noise affected the interpretation of activation in the “oral comprehension” paradigm.

Table 2: Comparison between Neuro3D and SPM analysis in motor paradigm experiments

|                | Right Hand | Left Hand | Right Foot | Left Foot | Oralng | Total |
|----------------|------------|-----------|------------|-----------|--------|-------|
| Tth            | Overl      | 2 (33.3%) | 4 (66.7%)  | 2 (33.3%) | 0 (0%) | 1 (16.7%) | 9 (30%) |
|                | Part.overl | 4 (66.7%) | 2 (33.3%)  | 4 (66.7%) | 6 (100%) | 2 (33.3%) | 21 (70%) |
|                | Not overl  | 0 (0%)    | 0 (0%)     | 0 (0%)    | 0 (0%) | 0 (0%) | 0 (0%) |
|                | Not pres   | 3 (50%)   | 3 (50%)    | 1 (16.7%) | 4 (66.6%) | 1 (16.7%) | 12 (40%) |
| Noise          | Pres. not inter | 3 (50%) | 3 (50%) | 5 (83.3%) | 1 (16.7%) | 3 (50%) | 15 (50%) |
|                | Inter      | 0 (0%)    | 0 (0%)     | 0 (0%)    | 1 (16.7%) | 2 (33.3%) | 3 (10%) |
| Tmax           | Overl      | 5 (83.3%) | 5 (83.3%)  | 4 (66.7%) | 4 (66.7%) | 5 (83.3%) | 23 (76.6%) |
|                | Part.overl | 1 (16.7%) | 1 (16.7%)  | 2 (33.3%) | 2 (33.3%) | 1 (16.7%) | 7 (23.3%) |
|                | Not overl  | 0 (0%)    | 0 (0%)     | 0 (0%)    | 0 (0%) | 0 (0%) | 0 (0%) |
|                | SPM        | 0 (0%)    | 0 (0%)     | 0 (0%)    | 0 (0%) | 0 (0%) | 0 (0%) |
| Cluster size   | Neuro3D    | 4 (66.7%) | 4 (66.7%)  | 4 (66.7%) | 5 (83.3%) | 5 (83.3%) | 22 (73.3%) |
|                | Equal      | 2 (33.3%) | 2 (33.3%)  | 2 (33.3%) | 1 (16.7%) | 1 (16.7%) | 8 (26.7%) |
|                | SPM        | 4 (66.7%) | 3 (50%)   | 4 (66.7%) | 6 (100%) | 6 (100%) | 23 (76.6%) |
| Preference     | Neuro3D    | 0 (0%)    | 0 (0%)     | 1 (16.7%) | 0 (0%) | 0 (0%) | 1 (3.3%) |
|                | Equal      | 2 (33.3%) | 3 (50%)   | 1 (16.7%) | 0 (0%) | 0 (0%) | 6 (20.1%) |

Note: inter indicates interfering; overl, overlapping; part.overl, partially overlapping; pres, present; orolng, orolingual.

FIG 1. Overlapping maps in the left finger-tapping paradigm. The first line of images is the Tth Neuro3D map, showing activation in the precentral area related to the primary motor cortex of the left hand (right side), in the left supplementary motor area, and smaller contralateral activation in the precentral primary motor cortex and premotor area. In the second line, the SPM map shows the same areas of activation.

FIG 2. Partially overlapping maps in the orolingual motor task. In the first line of images, the Tth Neuro3D map shows bilateral activation in the precentral area related to the primary motor cortex, in the premotor cortex, in the supplementary motor area, and in the postcentral bilateral area. In the SPM map (second line), the activation areas are smaller. Comparison between Tmax Neuro3D (third line) and SPM improves the classification to “overlapping.”
tion clusters in Th Neuro3D analyses were larger than in those in SPM in 90.9% of cases.

In the SPM analysis, the most common LI classification was “left dominance with right minor representation” (6/11), followed by “bilateral” (2/11), “left” (1/11), “right” (1/11), and “right dominance with left minor representation” (1/11). In the Neuro3D analysis, the LI was classified as “left dominance with right minor representation” (9/11), “right” (1/11), and “right dominance with left minor representation” (1/11).

The classification of LI by use of the 2 methods was concordant in 8 (72%) of the 11 participants. In 2 participants, SPM classified the LI as “bilateral,” whereas Neuro3D classified it as “left dominance with right minor representation.” In another participant, SPM classified the LI as “left,” whereas Neuro3D classified it as “left dominance with right minor representation.” In 2 participants, both methods classified the LI as “right dominance with left minor representation,” though both participants were right-handed.

**DISCUSSION**

Previous studies have established the value of fMRI to study the activation of motor and verbal speech areas during the planning of a surgical intervention and during the intervention itself through neuronavigation.4,5

In general, as shown in other studies, we found lower levels of activation for language tasks than for motor tasks, because cognitive tasks produce less signal change than motor or sensory tasks.6 Among motor tasks, hand movement resulted in the highest T values and foot movement in the lowest. In contrast, no significant differences in T values were found between different language tasks.

Comparing the 2 fMRI postprocessing methods, we found that, in most paradigms, the Neuro3D system was sufficient to assess the activations. In 20% of studies, we observed no visual differences in the postprocessing maps, and critical differences were absent in more than half. The interpretation by Neuro3D differed from that by SPM in only 7% of cases; in these cases, the differences were attributable to the Neuro3D software’s more imperfect correction of “noise” attributed motion artifacts or vessels. These results are in agreement with a previous study that also compares real-time fMRI provided by commercial software with SPM postprocessing.7 It is worth pointing out that in our study the postprocessing was off-line, as is done when using SPM. Online postprocessing in real time results in changing functional maps, depending on the data provided at the moment of acquisition.

Concordance between the 2 methods was better when the functional maps generated by Tmax were compared (Fig 2): in nearly all cases, studies initially classified as “not overlapping” were upgraded to “partially overlapping,” and those classified as “partially overlapping” were upgraded to “overlapping” after the Tmax maps were considered. This occurs because being stricter with the T threshold eliminates noise and nonspecific activations or false-positive activation in a particular paradigm. On the contrary, in 3 cases, use of the Tmax as a threshold resulted in down-grading concordance from “overlapping” to “partially overlapping”; however, none of the cases changed to “not overlapping.” This occurred because increasing the level of restriction can sometimes eliminate the activated areas.5 We recommend using the Tmax value to avoid false-positive results, though it is also important to be careful to avoid being too restrictive that the active areas of interest could be erased. The image quality in the Neuro3D analysis was more affected by noise, but the noise interfered in the assessment of the eloquent areas in only 10% of cases, all of which occurred with the language paradigms. Therefore, although the activation images obtained with SPM were “cleaner” and generally preferred by radiologists, the activation images obtained with Neuro3D were reliable for the location of eloquent areas. This means that this important objective can be reliably fulfilled on the MR imaging workstation. Nevertheless, in cases in which noise prevents the evaluation of specific activation, SPM postprocessing must be used. Furthermore, in most cases, the activated clusters were larger in the Neuro3D analysis than in SPM, and this difference must be considered when lesions are assessed and surgery is planned.

When we compared the 2 postprocessing methods for the assessment of the LI, we found strong agreement (72%) and no discordance regarding the dominant side. It is important to note that in participants with “right” or “right dominance with left minor representation,” the 2 methods were concordant. This result confirms that fMRI is a robust technique to assess language dominance. Our results suggest that standard MR imaging postprocessing software can reliably determine LI in the preoperative work-up, and more sophisticated postprocessing software used in research need only be used in participants in whom discordant laterality results are found with different paradigms or in other circumstances such as excessive noise that could prevent accurate assessment of the LI.

One limitation of this study was the small number of participants (n = 10) and that all participants were healthy volunteers. A larger study including patients in whom the anatomy might be disrupted is necessary to ensure more consistent results and to determine whether there are any advantages in neurosurgical management.

Another limitation was that our results were derived from a single commercial software package. We were unable to compare other packages because they were not available in our institution. We encourage other researchers who use other commercial packages to perform the same study.

Neuro3D does not allow interventions in every step of the process, and it provides less information about the statistical analyses that is important to rule out false-positive or false-negative activations.

Although the commercially available package may be perfectly acceptable for streamlining clinical workflow, its use should not preclude drawing on more rigorous research packages. Researchers and those without adequate experience in clinical fMRI should

| Table 3: Th and Tmax for language paradigm experiments: mean and range |
|---------------------|---------------------|---------------------|---------------------|
| Semantic Judgment | Verbal Fluency | Oral Comprehension |
| Th      | Tmax | Tth    | Tmax  |
| Th      | Tmax | Tth    | Tmax  |
| Th      | Tmax | Tth    | Tmax  |
| 4.4     | 6.6  | 4.1    | 5.9   |
| 2.6–6.5 | 3.9–11.4 | 1.6–6.5 | 2.7–9.0 |
| 1.6–4.3 | 2.0–9.8 | 110   | 119    |
not rely on commercially available packages exclusively. However, we believe that professionals with experience in fMRI should be able to identify complex cases that require SPM or other research programs.

CONCLUSIONS

Our results corroborate the findings of previous studies that show that fMRI can be a helpful and robust tool to locate eloquent areas (motor and language) when planning neurosurgery,8,9 even with commercial software.7

Readily available, user-friendly commercial software for analyzing fMRI data in clinical practice can provide key clinical information for the therapeutic treatment of patients, and more complicated, less readily available postprocessing programs used in research are necessary in few cases.

ACKNOWLEDGMENTS

The authors thank the Hospital Clinic i Provincial of Barcelona for the "Emili Letang" postresidency grant.

Disclosures: Nuria Bargallo—UNRELATED: Board Membership: Spanish Society of Neuroradiology, ESMOFIR, Comments: no money received; Consultancy: Jansen; Grants/Grants Pending: Fondo de Investigaciones Sanitarias (Spain);* Payment for Lectures (including service on speaker bureaus): Spanish Society of Neurology. [*Money paid to institution.]