Evaluation of a pre-surgical functional MRI workflow: From data acquisition to reporting

Evaluation of a pre-surgical functional MRI workflow: from data acquisition to reporting

Cyril R. Pernet[a,b], Krzysztof J. Gorgolewski[c], Dominic Job[a,b], David Rodriguez[a,b], Amos Storkey [e], Ian Whittle[a,d] & Joanna Wardlaw[a,b].

a, Brain Research Imaging Centre, The University of Edinburgh, UK
b, Centre for Clinical Brain Sciences, The University of Edinburgh, UK
c, Department of Psychology, Stanford University, Stanford, USA
d, Division of Clinical Neuroscience, NHS-Lothian, Edinburgh, UK
e, School of Informatics, The University of Edinburgh, UK

Author's copy, the final version is published in

International Journal of Medical Informatics, Volume 86, February 2016,

Pages 37-42, ISSN 1386-5056, http://dx.doi.org/10.1016/j.ijmedinf.2015.11.014.

Corresponding author:
Dr Cyril Pernet,
Centre for Clinical Brain Sciences (CCBS)
Neuroimaging Sciences
The University of Edinburgh
Chancellor's Building
Room GU426D
49 Little France Crescent
Edinburgh EH16 4SB
cyril.pernet@ed.ac.uk

Abstract

Purpose*:* present and assess clinical protocols and associated automated workflow for pre-surgical functional magnetic resonance imaging in brain tumor patients. ***Methods:*** Protocols were validated using a single-subject reliability approach based on 10 healthy control subjects. Results from the automated workflow were evaluated in 9 patients with brain tumors, comparing fMRI results to direct electrical stimulation (DES) of the cortex. ***Results:*** Using a new approach to compute single-subject fMRI reliability in controls, we show that not all tasks are suitable in the clinical context, even if they show meaningful results at the group level. Comparison of the fMRI results from patients to DES showed good correspondence between techniques (odds ratio 36). ***Conclusion:*** Providing that validated and reliable fMRI protocols are used, fMRI can accurately delineate eloquent areas, thus providing an aid to medical decision regarding brain tumor surgery.

Keywords*:* protocol, workflow, functional MRI, surgical planning, brain tumors

Highlights

- Operational definition of clinical fMRI protocols
- Validation of single subject fMRI protocol and dedicated data analysis
- Automated data analysis to aid brain tumour surgical planning
- Validation of fMRI results in patients vs. direct electrical stimulation

Introduction

In medicine, Magnetic Resonance Imaging (MRI) is typically used to image the structure of organs. MRI is however also used to obtain information about perfusion, diffusion, vascularization and physico-chemical state of tissues. Functional MRI (fMRI) is a technique that measures hemodynamic changes after enhanced neural activity [1], allowing to image non-invasively and with relatively high spatiotemporal resolution, the entire network of brain areas engaged when subjects undertake particular tasks [2]. Soon after its inception, fMRI has been used for clinical cases [3]. Nowadays, clinical research using fMRI encompasses many areas of neurology, from developmental, psychiatric, and dementia related disorders to strokes and brain tumors [4]. Despite the popularity of fMRI in cognitive and clinical research and its proven utility for surgical planning [5], it is not used extensively in day to day clinical practice. There are four main reasons for this: (i) fMRI requires special equipment, (ii) dedicated protocols must be in place, (iii) collected data have to be post-processed to obtain a final image, and (iv) results from analyses must be made available to the clinicians in a usable format.

fMRI delineates areas of the brain involved in motor or cognitive functions (so called eloquent areas) by asking patients to perform different tasks whilst image time-series are acquired. For motor related areas, a simple finger tapping (mapping the primary motor cortex) or more complex finger sequences (mapping the premotor cortex) may, for instance, be performed. For language areas, visual or auditory stimuli are presented whilst scanning, and patients perform different tasks such as reading, listening, repeating, etc. All patients must perform several trials, and crucially these trials must be synchronized with image acquisition. The tight coupling between stimulus presentation, task and image acquisition is mandatory for the statistical analysis, contrasting task periods versus rest periods, or contrasting different tasks periods against each other (Figure 1). This implies that MRI compatible equipment is available to interface between the scanner and the software used to design the tasks. The fMRI hardware, is also used to deliver instructions to the patient via MRI compatible headphones, screen, goggles, etc. and possibly also to record behavioral responses (via e.g. microphone, response pads), all of this in phase with the image acquisition. Typically, such equipment is available in research centers but not hospitals, constituting an obstacle to day-to-day application of fMRI.

In many university hospitals, research centers with fMRI equipment are present on site (and even sometimes in or next to the clinical department), and therefore patients can be scanned without the need for transportation to a different location. To ensure good clinical practice, established fMRI protocols must however be in place. These protocols must allow the mapping of given brain areas with high specificity. Because, there are many possible tasks to map the same brain area [6,7] and these have also been developed for group studies, there are no ‘off-the-shelf’ protocols that can be used to elicit reliable activations at the single subject level. It is therefore mandatory to establish standards to define ‘good clinical fMRI protocols’ and create such protocols to establish fMRI as a clinical tool.

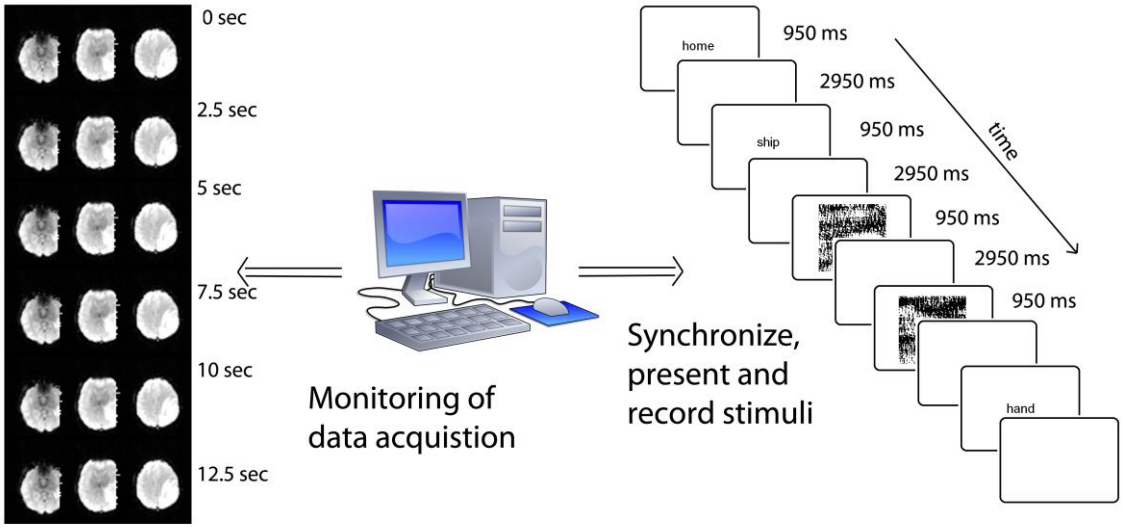

Figure 1. Illustration of hardware setting for fMRI: (i) on the left side is shown EPI data acquired by an MRI scanner with a repetition time of 2.5 sec; (2) in the middle a dedicated computer with specific hardware monitors the scanner data acquisition while presenting stimuli at specified times (3) on the right is a series of stimuli showed inside the scanner using MR compatible goggles, corresponding to our verb generation task. The synchronization between the MR images acquired (left) and the stimuli presented (right) is mandatory to contrast brain images acquired while the patient was seeing words vs. seeing noise stimuli.

Assuming that such protocols are in place, and can be run by trained radiographers, the data must be processed before reporting because, in contrast to standard structural imaging (e.g. T1 or T2 weighted images), there is no direct output from the scanner. Although scanner manufacturers offer fMRI acquisition mode, which in theory allow obtaining results after a scanning session, the schemas are extremely rigid and do not fit modern complex protocols. Off-line analyses must therefore take place, and this can take from half an hour to several hours depending on the length of the processing pipeline, number of tasks performed, the complexity of the analyses, and the hardware used. Such analysis also requires expert knowledge. Together, these constitute another strong deterrent to clinical fMRI. We believe that this complexity in data analysis can be overcome by creating automated analysis workflows that (i) allow checking data quality and analysis and, (ii) output 'ready-to-use' reports and images.

The images produced by fMRI software and the images used in the clinical environment have different format. This might seem trivial but it is still a major problem. Data coming out of the MRI scanner are in the DICOM format (http://dicom.nema.org/) but researchers using fMRI typically convert them to NIfTI format (Neuroimaging Informatics Technology Initiative http://nifti.nimh.nih.gov/) because it has many advantages for research use, and in particular it facilitates interoperability among software. In addition to changing format, data are often de-identified, making the conversion back to DICOM and its use on clinical PACS and other tools like neuro-navigation difficult.

Having previously developed a set of tasks suitable for patients [8], we present here (i) a validation of those protocols, showing higher within than between-subjects reliability and (ii) an automated analysis pipeline from data transfer to reporting, fitting with the busy day-to-day clinical practice. Pipeline analysis and optimization can take many different forms, but this is out of the scope of this article. We focus here on the implementation of such pipeline using open source software and the clinical validity of the results obtained.

Materials and Methods

Participants

All participants (healthy controls and patients) signed a written informed consent for this study that was approved by the NHS Lothian ethics committee.

Protocol validation in healthy controls

We investigated the within-subject reliability of the 5 tasks described in Gorgolewski et al. (2013 – [9]): one motor task to map the hand, foot and lips regions of the motor cortex, three language related tasks to map the auditory cortex, Wernicke and Broca areas, one attentional task to map the intra-parietal cortex (IPC). For each task, ten healthy participants (four males and six females, of which three were left-handed and seven right-handed according to their own declaration, with a median age at the time of first scan of 52.5 years; min=50, max=58 years) underwent two separate sessions and the reliability of activation maps was assessed. We consider a protocol as clinically valid when areas of activations are more reliable within-subjects than across subjects. Concretely that means that despite the same region of the brain being activated across subjects (for instance the hand area), the obtained maps must be more similar when repeated over two sessions in a given subjects than across subjects.

For every subject, T1 volumes from both sessions were coregistered, resliced and averaged using SPM8 [10]. A Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) template was created using averages from all subjects [11] and a brain mask was estimated from each average using the FSL [12] Brain Extraction Tool (BET). The first 4 volumes, during which the scanner reaches steady state, of every EPI sequence were discarded and the remaining images were slice time corrected. Motor sequences of 3 left–handed subjects were flipped along the Z–Y plane. For every subject, all slice time corrected volumes from all tasks and sessions were realigned and resliced to their mean to remove motion artefacts. The mean volume was coregistered to the T1–weighted volume between session average. The obtained affine transformation from the mean EPI to the T1 was then applied to headers of realigned files. Each EPI volume was then normalized using the DARTEL template and corresponding flow field and smoothed with 8mm full width half-maximum Gaussian kernel. Each session was analyzed separately, with a General Linear Model [13] being used to fit a design matrix consisting of an autoregressive filtering matrix (AR1) and task, realignment (6 parameters), high pass filter (128 Hz), and artefacts (one per artefact) regressors. Only voxels within previously estimated brain mask were included in model fitting. For an overview of pre-processing and first level analysis see Gorgolewski et al. (2013 - [8]).

For each session and task, contrasts of interest were computed and the resulting statistical parametric maps were thresholded using an adaptive technique, developed for single-subject analyses [14]. Dice coefficients were next computed within-subject (i.e. between the 2 sessions) and between-subjects (i.e. the average of all within-session map overlaps across subject pairs). The Harrell-Davis estimates of the median were then compared using percentile bootstraps: 95% confidence intervals and p-values are reported for each map separately and statistical significance was obtained using a False Discovery Rate (FDR) correction over all measures of location (q=5% [15]). Since, the validation is based on healthy volunteers, a-priori location of eloquent areas may be determined. Results are thus reported for the full brain and for task specific region of interest (ROI), namely the motor cortex, Wernicke area, Broca area, and the right IPC. These ROI were constructed using probability maps available in the anatomy toolbox [16,17] and were resliced to DARTEL template dimensions.

Automated analysis pipeline

A fully automated analysis workflow (Figure 2) was created to analyze tasks showing high within-subject reliability: (1) motor task showing activations for the hand, foot and lip areas, (2) auditory language task showing activations of the auditory cortex and Wernicke area, (3) the verb generation task showing activations of the supplementary motor area and Broca area. The workflow relies on multiple open source software packages as well as Matlab®.

From a data handling and format perspective, the pipeline can be split into three parts: (i) obtain raw DICOM image, anonymize and convert to NIfTI; (ii) do the actual data analysis; (iii) transform images back to DICOM re-attaching patient's information. From the scanner, identifiable DICOM data are stored on a secure server. The data are then de-identified using DICOM Confidential [18] and converted to NIfTI using dcm2nii from the MRIcron suite [19] and transferred onto a data-processing server. DICOM Confidential is a policy-driven DICOM de-identification toolkit developed in Java. It provides great flexibility regarding both what data should be protected and how to do it. Although the library comes with a number of ready-made classes that cover the most usual requirements, users can develop their own Java classes to accommodate their needs. Using DICOM-confidential allows us to strip identifiers from the DICOM headers and simultaneously generate a mapping file where the patient identification number is linked to a randomly generated pseudo-identifier that substitutes it in the header of MRI files. Once transferred, the de-identified and converted data (now in NIfTI format) are processed via a Matlab based script that calls dedicated pipelines (one for anatomical images and one per fMRI task). Each pipeline relies on the PSOM engine (20) which allows 1 – to avoid stopping the workflow by capturing errors if any and 2 – returning reports of the data analysis steps performed. Each pipeline relies on SPM to analyze the fMRI data [10] and updates a report file initiated in a script. At the end of each pipeline, the result images are 'printed' as series of slice pictures in JPEG format which are then converted to DICOM, re-attaching the patients' CHI. This last conversion is performed using the jpg2dcm java utility [21]. All the results are then copied onto a shared NHS/University server.

Workflow evaluation

Each pipeline calls routines from SPM [10] for the data analysis in conjunction with custom functions to determine e.g. outlier data points. Critically, the statistical thresholding depends on the Gamma-Gaussian Mixture model of the t values to set a cluster forming threshold, allowing separating activations from noise with a balance between the type I and type II error rates [14]. This thresholding technique is at the heart of the data analysis, since controlling the type II error rate is crucial for surgery. Here we evaluated the robustness of this approach by (1) comparing the activation pattern found in 9 patients (see table 1 for demographic information) with direct electrical stimulation (DES - [22] ,[23]) performed during surgery and (2) comparing the adaptive thresholding technique with a more standard fixed threshold approach, i.e. thresholding each map at p=.001 uncorrected to set a cluster forming threshold. In both cases, once a map was formed, a cluster-wise false discovery rate with q=0.05 was used, to control false positive clusters [24]. DES was used as ground truth and true/false positives and negatives activations were defined in reference with DES effects (Figure 3). The association between fMRI and DES was tested using a two-tailed Fisher's exact test. The data are available through the UK data service (http://ukdataservice.ac.uk/) under collection name: 'A neuroimaging dataset of brain tumour patients' [25].

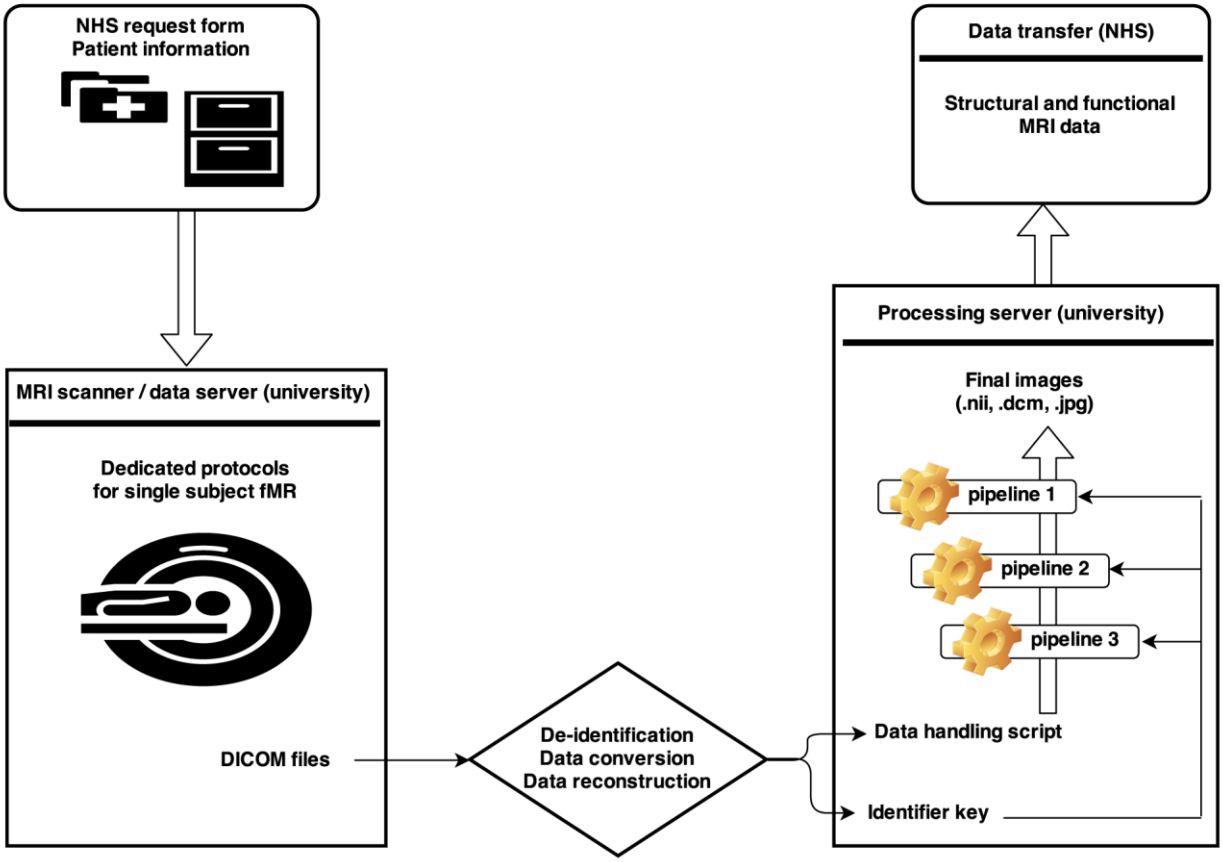

Figure 2. Data workflow. Once data are acquired on the scanner (DICOM files) they are de-identified and converted to NIfTI. Data processing and storage is done on de-identified data. Patient ID is re-attached only for the final images (.dcm) that are transferred onto a shared university/NHS server.

Sub-ID	Sex	Age	Pathology	Tumor Location	Handedness
18638	F	35	Astrocytoma type II	Right primary somatosensory area	R
19227	M	31	Astrocytoma type II	Right primary somatosensory area	R
19567	M	42	Astrocytoma type II	Left Pre-Motor area	R
19723	F	27	Astrocytoma type II	Left temporal cortex	L
18428	F	25	Astrocytoma type III	Right Supplementary Motor Area	R
18756	M	42	Glioblastoma Multiform	Right primary motor area	R
18675	M	68	Meningioma	Left primary motor area	R
17904	M	43	Oligodendrocytoma type II	Left Supplementary Motor Area	R
18863	M	75	renal cell carcinoma	Left Pre-Motor Area	L

***Table 1**. Patients' demographics with sex, age, pathology obtained from histology, tumor location, and patient handedness.*

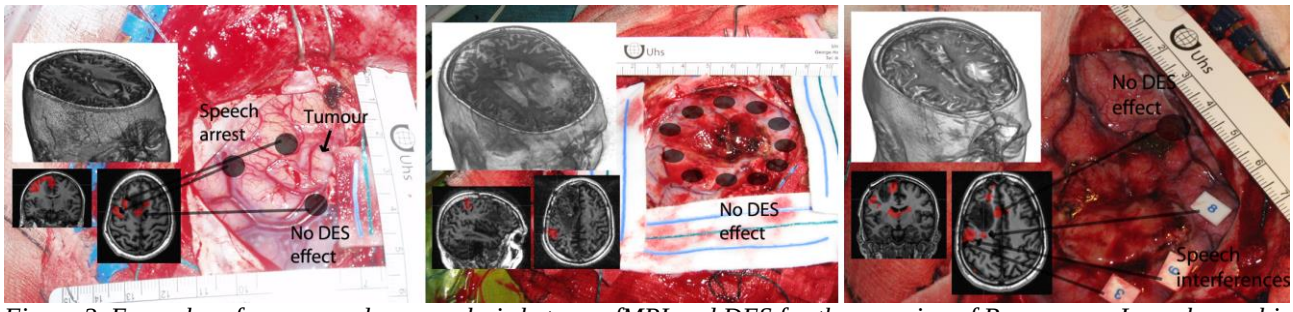

Figure 3. Examples of correspondence analysis between fMRI and DES for the mapping of Broca area. In each panel is displayed a render of the structural T1 image showing the tumor location, fMRI maps showing where activations were observed, and a picture taken during surgery with indication of where DES was performed. On the left is shown patient 17904 in which fMRI revealed two true positives and one false positive, relative to DES. In the middle is shown patient 18863 in which fMRI revealed a true negative, i.e. no activation of Broca area (in both fMRI and DES) that concurs with fMRI activations observed more posteriorly than in healthy individuals. Finally, on the right is shown patient 19567 in which fMRI revealed two true positives and a false negative (fMRI activation but no DES effect).

Results

Protocol validation

For the motor task, in which participants moved alternatively their hand, foot or lips, 10 out of 10 subjects show significant activations in both sessions and a significantly higher within-subject than between-subjects reliability was observed (Table 2).

	Fingers (>others)		Foot (>others)		Lips (>others)	
	Whole brain	Motor cortex	Whole brain	Motor cortex	Whole brain	Motor cortex
Within	0.61	0.76	0.53	0.71	0.46	0.85
Between	0.33	0.56	0.31	0.55	0.23	0.65
Difference CI	[0.07 0.39]	[0.03 0.34]	[0.08 031]	[0.07 0.26]	[0.08 0.32]	[0.1 0.27]
p value	<=0.009	<=0.01	<=0.0006	<=0	<=0.003	<=0

Table 2. *Within vs. Between subjects median dice coefficient for the 3 contrasts computed in the motor task, with percentile bootstrap 95% confidence interval of the difference and FDR corrected p value. Results are reported for the overlapping results across the whole brain and restricted to the motor cortex.*

For the three language tasks, higher reliability within-subject than between-subjects was also observed. For the overt word repetition task, 9 out of 10 subjects show significant activations in both sessions. The full brain pattern of activity and Wernicke area show more reliable significant activations within than between-subjects (Table 3). For covert and overt verb generation tasks, both full brain pattern of activity were more reliable within than between-subjects. Only the covert verb generation task showed reliable results for Broca area (Table 3). In the covert verb generation task, 10 out of 10 subjects show significant activations in both sessions, while in the overt verb generation task only 7 out of 10 subjects show significant activations in both sessions. Comparison of the two tasks (paired bootstrap test) over Broca area and its homologue show a significant difference (p=0.001) in favor of the covert verb generation task with 6 out of 7 subjects showing a bigger between sessions overlap.

For the attentional task, subjects had to decide which of two lines are bigger or smaller. This so-called landmark task is an fMRI substitute for the line bisection task used in clinical context. This task tests for the presence of unilateral spatial neglect by asking to mark with a pencil the centre of a series of horizontal lines. Here, no significant difference between within-subjects' overlaps and between-subjects' overlaps were observed (Table 4). Although 10 out of 10 subjects show significant activations in both sessions, locations were variables and show minimal overlaps.

	Overt word repetition		Overt verb generation		Covert verb generation	
	Whole brain	Wernicke area	Whole brain	Broca area	Whole brain	Broca area
Within	0.46	0.58	0.33	0.31	0.56	0.63
Between	0.22	0.34	0.07	0.05	0.24	0.32
Difference CI	[0.11 0.34]	[0.06 0.41]	[0.02 0.46]	[-0.01 0.5]	[0.06 0.42]	[0.05 0.49]
p value	0	<=0.013	<=0.07	=0.07	<=0.017	<=0.017

Table 3. *Within vs. Between subjects median dice coefficient for the 3 language tasks, with percentile bootstrap 95% confidence interval of the difference and FDR corrected p value. Results are reported for the overlapping results across the whole brain and restricted to the regions of interest (Wernicke or Broca area).*

	Landmark task	
	Whole brain	Right IPC
Within	0.11	0.06
Between	0.09	0.19
Difference CI	[-0.11 0.22]	[-0.25 0.12]
p value	=0.79	>0.6

Table 4. *Within vs. Between subjects median dice coefficient for the landmark task (contrast landmark trials vs. detection trials), with percentile bootstrap 95% confidence interval of the difference and FDR corrected p value. Results are reported for the overlapping results across the whole brain and restricted to the right intra-parietal cortex.*

Workflow evaluation

For all patients analyzed, the workflow processed the data without difficulties returning reports, lateralization indices for language tasks (when performed), and activation maps. A total of 21 activations maps were analyzed, showing a good correspondence between fMRI and DES. Five patients underwent DES and fMRI for motor region mapping with 10 sites tested. Five patients (four different patients and one also tested for the motor task) underwent DES and fMRI to map Broca area with also 10 sites tested. Finally, only 1 patient (one also tested for the motor task) underwent DES and fMRI to map Wernicke area.

Using the adaptive thresholding technique we found a significant association between fMRI and DES (odds ratio = 36 [2.6 481] F=0.0031 p=0) whilst the correspondence was not significant using a fixed threshold (odds ratio = 2.25 [0.36 18.87] F=0.24 p=0.64). Importantly, more true positives and less false negatives were observed using the adaptive thresholding (18/21 p=0.0007) than the fixed thresholding (13/21 p= 0.0392; table 5).

N = 21 from 9 patients	DES effect	DES no effect
BOLD activation	True positives 12 vs. 9	False positives 2 vs. 4
No BOLD activation	False negative 1 vs. 4	True Negative 6 vs. 4

Table 5. *Number of sites showing correspondences or differences between fMRI and DES. Results are reported for the adaptive vs. fixed thresholding techniques.*

Discussion

To validate fMRI tasks that can be used confidently in clinical practice, we assessed the reliability of activations maps using Dice coefficients. A typical issue related to reliability metrics is to establish what a good reliability value is. Here, we defined a protocol as valid when it showed a strong within-subject reliability that is, a higher within-subject map overlap than between-subjects map overlap. Whilst in previous research, the group maps were tested, our approach statistically tests how reliable a task is at the subject level. Our approach is particularly relevant in the clinical context since overall the same areas are expected to be activated across healthy participants, although in patients non-classical areas could take over various functions because of a slow growing tumor. Assuming that true activations are those that are reliable, then this approach allows deciding if a task is suitable for clinical purposes or not.

Results indicate that the motor task, the covert word repetition and verb generation are suitable. The overt verb generation showed some degree of reliability, although not for Broca area. This task was the same duration as the covert verb generation task, and thus lacked power since sparse sampling was used to allow participants to answer. It is likely that increasing the number of trials would increase its reliability, but total scanning time should also be considered. When working with clinical populations it is essential to have short scanning sessions (each tasks here last ~5 min to make them compatible with busy clinical departments, whilst some tasks in cognitive neuroscience can last up to an hour or more). The landmark task showed a strong activation around the right inferior parietal gyrus at the group level, an area also known to cause hemi-neglect if injured [26]. Yet, no clear activations were observed in this region at the subject level, thus exhibiting poor reliability. This may be explained by a high within-subject scanner noise relative to the present hemodynamic changes as well as the cognitive complexity of the task. Activation of the right IPC are observed at the group level because the within-subject variance is discarded. This last result also demonstrates that tasks which have been successfully used in cognitive neuroscience to study some brain regions (at the group level) can be useless for clinical application in individuals. This in turn highlights the need to within-subject reliability analyses to validate protocols.

The use of a dedicated analysis workflow in the clinical context demonstrates the feasibility of clinical fMRI with all patients successfully, automatically analyzed. In most cases, valuable information was gained from thresholded maps offering computer-aided medical decision. In some instances, looking at an unthresholded map may be useful to find additional areas, and such cases point to the need to still have an fMRI expert to go over the report and check activation maps, in liaison with the radiologist and neurosurgeon.

There was a good correspondence between fMRI and DES (only 3/21 discrepancies), which suggests that it might be possible in the future to substitute fMRI for DES. Indeed, despite being the gold standard, DES does not allow the surgeons to draw unequivocal conclusions about the role of stimulated areas' [23], just like fMRI. fMRI has however the advantage of being non-invasive, and to possibly reveal cases where surgery is too risky (eloquent areas around the tumor or even inside the tumor) thus providing invaluable information to surgeons and to patients alike.

Acknowledgements
Cancer Research UK through the Edinburgh Experimental Cancer Medicine Centre funded this study.

References

1. Ogawa S, Lee TM, Kay AR, Tank T. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci. 1990;87:9868–72.
2. Logothetis NK. What we can do and what we cannot do with fMRI. Nature. 2008;453:869–78.
3. Gallen CC, Bucholz R, Sobel DF. Intracranial neurosurgery guided by functional imaging. Surg Neurol. 1994;42(6):523–30.
4. Detre JA, Floyd TF. Functional MRI and Its Applications to the Clinical Neurosciences. The Neuroscientist. 2001 Feb 1;7(1):64–79.
5. Pillai JJ. The Evolution of Clinical Functional Imaging during the Past 2 Decades and Its Current Impact on Neurosurgical Planning. Am J Neuroradiol. 2010 Feb 1;31(2):219–25.
6. Friston KJ. Imaging neuroscience: Principles or maps? Proc Natl Acad Sci. 1998;95(3):796.
7. Anderson ML, Kinnison J, Pessoa L. Describing functional diversity of brain regions and brain networks. NeuroImage. 2013 Jun;73:50–8.
8. Gorgolewski KJ, Storkey AJ, Bastin ME, Whittle I, Pernet C. Single subject fMRI test–retest reliability metrics and confounding factors. NeuroImage. 2013 Apr;69:231–43.
9. Gorgolewski K, Storkey AJ, Bastin ME, Whittle I, Wardlaw J, Pernet CR. A test-retest fMRI dataset for motor, language and spatial attention functions. GigaScience. 2013;2:6.
10. Flandin G, Friston KJ. Statistical parametric mapping (SPM). Scholarpedia. 2008;3(4):6332.
11. Ashburner, J.T. A fast diffeomorphic image registration algorithm. NeuroImage. 2007;38(1):95–113.
12. Jenkinson, M., Beckmann, C.F., Behrens, T., Woolrich, M.W., Smith, S.M. FSL. NeuroImage. 2012;62:782–90.
13. Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp. 1994;2(4):189–210.
14. Gorgolewski KJ, Storkey AJ, Bastin ME, Pernet CR. Adaptive thresholding for reliable topological inference in single subject fMRI analysis. Front Hum Neurosci [Internet]. 2012
15. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(7):289–300.
16. Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage. 2005 May;25(4):1325–35.
17. Eickhoff SB, Paus T, Caspers S, Grosbras M-H, Evans AC, Zilles K, et al. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. NeuroImage. 2007 Jul;36(3):511–21.
18. Rodriguez Gonzalez D, Carpenter T, van Hemert JI, Wardlaw J. An Open Source Toolkit for Medical Imaging De-Identification. Eur Radiol. 2010;20(8):1896–904.
19. Rorden C. dcm2nii [Internet]. 2012. Available from: http://www.mccauslandcenter.sc.edu/mricro/mricron/dcm2nii.html
20. Bellec P, Lavoie-Courchesne S, Dickinson P, Lerch J, Zijdenbos A, Evans AC. The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine

for scientific workflows. Front Neuroinformatics [Internet]. 2012;6(7).
21. dcm4che2 DICOM Toolkit [Internet]. Available from: https://dcm4che.atlassian.net/wiki/display/d2/dcm4che2+DICOM+Toolkit
22. Krause F. Die operative Behandlung der Epilepsie. Med Klin Berl. 1909;5:1418–22.
23. Borchers S, Himmelbach M, Logothetis N, Karnath H-O. Direct electrical stimulation of human cortex — the gold standard for mapping brain functions? Nat Rev Neurosci. 2012 Jan;13(1):63–70.
24. Chumbley J, Worsley K, Flandin G, Friston K. Topological FDR for neuroimaging. NeuroImage. 2010 Feb;49(4):3057–64.
25. Pernet CR, Gorgolewski KJ, Job D, Rodriguez Gonzalez D, Whittle I, Wardlaw J. A structural and functional magnetic resonance imaging dataset of brain tumour patients. Sci Data. 2016;in press.
26. Corbetta M, Shulman GL, others. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci. 2002;3(3):215–29.