




Automatic identification of atypical clinical fMRI results

J. Martijn Jansma^{1,2} · Geert-Jan Rutten² · Lenny E. Ramsey³ · T. J. Snijders¹ · Alberto Bizzi⁴ · Katharina Rosengarth⁵ · Frank Dodoo-Schittko⁶ · Elke Hattingen⁷ · Mar Jiménez de la Peña⁸ · Gord von Campe⁹ · Margit Jehna¹⁰ · Nick F. Ramsey^{1,11} 

Received: 6 March 2020 / Accepted: 30 July 2020 / Published online: 18 August 2020
© The Author(s) 2020

Abstract

Purpose Functional MRI is not routinely used for neurosurgical planning despite potential important advantages, due to difficulty of determining quality. We introduce a novel method for objective evaluation of fMRI scan quality, based on activation maps. A template matching analysis (TMA) is presented and tested on data from two clinical fMRI protocols, performed by healthy controls in seven clinical centers. Preliminary clinical utility is tested with data from low-grade glioma patients.

Methods Data were collected from 42 healthy subjects from seven centers, with standardized finger tapping (FT) and verb generation (VG) tasks. Copies of these “typical” data were deliberately analyzed incorrectly to assess feasibility of identifying them as “atypical.” Analyses of the VG task administered to 32 tumor patients assessed sensitivity of the TMA method to anatomical abnormalities.

Results TMA identified all atypical activity maps for both tasks, at the cost of incorrectly classifying 3.6 (VG)–6.5% (FT) of typical maps as atypical. For patients, the average TMA was significantly higher than atypical healthy scans, despite localized anatomical abnormalities caused by a tumor.

Conclusion This study supports feasibility of TMA for objective identification of atypical activation patterns for motor and verb generation fMRI protocols. TMA can facilitate the use and evaluation of clinical fMRI in hospital settings that have limited access to fMRI experts. In a clinical setting, this method could be applied to automatically flag fMRI scans showing atypical activation patterns for further investigation to determine whether atypicality is caused by poor scan data quality or abnormal functional topography.

Keywords Functional MRI · Motor cortex · Language · Brain function · Clinical fMRI

Introduction

Functional MRI (“fMRI”) is one of the most popular and widely used brain activation measurement tools in cognitive

neuroscience. fMRI is a non-invasive imaging modality with a spatial resolution that is high compared with other non-invasive functional imaging methods, such as EEG. It is also relatively easily accessible due to the wide availability of MRI

✉ Nick F. Ramsey
n.f.ramsey@umcutrecht.nl

¹ Brain Center, Department of Neurology & Neurosurgery, University Medical Center Utrecht, Utrecht, The Netherlands

² Department of Neurosurgery, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands

³ Department of Neurosurgery, McGovern Medical School, University of Texas Health Science Center, Houston, TX, USA

⁴ Neuroradiology Unit, Istituto Clinico Humanitas IRCCS, Rozzano, Milan, Italy

⁵ Institute for Experimental Psychology, University of Regensburg, Regensburg, Germany

⁶ Medical Sociology, Institute for Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany

⁷ Institute of Neuroradiology, Goethe University, Frankfurt, Germany

⁸ Diagnostic Imaging Department, Hospital Universitario QuirónSalud, Madrid, Spain

⁹ Department of Neurosurgery, Medical University of Graz, Graz, Austria

¹⁰ Division of Neuroradiology, Vascular and Interventional Radiology, Medical University of Graz, Graz, Austria

¹¹ Braincarta BV, Utrecht, The Netherlands

scanners, particularly in clinical centers. These characteristics are important reasons why the development of fMRI has had a strong impact on neuroscience.

It is also considered to carry a strong potential for clinical applications. Clinical applications of fMRI have mostly centered on presurgical use for patients with a brain tumor and patients with epilepsy. In both fields, several reviews have concluded that fMRI can provide important information for clinical care [1–5]. For instance, several studies have indicated that for language dominance, fMRI shows good agreement with invasive clinical measures such as the Wada test [6–9]. It has also been shown to be helpful in providing information about brain function topography prior to surgery for brain tumor patients [3, 10–16] and epilepsy patients [17], although some limitations have also been described [18].

One important issue limiting the use of fMRI in a clinical setting is the quality of the results which can vary considerably due to the complex process of fMRI acquisition. It is far from straightforward to determine, even for experts, whether the quality of an fMRI scan has been compromised by movement, task non-compliance, or other disruptions during acquisition. Additionally, brain activation patterns from neurological patients can deviate strongly from what is expected in location, extent, and magnitude of activated brain areas due to the neurological disorder [3, 19]. This could further complicate the problem of detecting invalid scans in a clinical setting. While certain measures indicating the quality of a scan are currently available, there is no general expert agreement on their value for accepting or rejecting a clinical fMRI activity map. The effect of some other factors, such as task compliance, can currently only be determined by expert evaluation. Thus, the determination of the quality of an fMRI scan currently requires expertise that is not readily available in clinical settings. A fast, automated, and objective first selection method for identifying atypical fMRI scans would greatly facilitate clinical use of fMRI.

In this report, we introduce a method that can provide a fast, automatic, and objective determination of the atypicality of an activation pattern. The approach is based on the notion that the performance of an fMRI task affects the whole brain, rather than only those parts that exceed a statistical threshold. We exploit this factor in our proposed method to determine an objective value describing the atypicality of an activation pattern. This value is based on a whole brain voxel by voxel correlation of each individual activity map with an independent template activity (“template matching analysis” or “TMA”; see Fig. 1).

In order to test the validity of the method, we analyzed fMRI from healthy controls collected at seven clinical centers in Europe with two standardized fMRI tasks to map hand motor and language function. Participating centers were all members of the European Low Grade Glioma Network (www.braintumours.eu). The tasks have previously been

described for presurgical localization of motor [11, 20, 21] and language function [22, 23]. Additionally, we used this dataset to artificially create atypical activity patterns due to incorrect analysis, incorrect orientation, or incorrect task execution. This set was used to test the feasibility of our method to automatically identify scans with atypical activation patterns.

We also applied our method on a series of 32 clinical brain tumor verb generation scans from consecutive patients considered for surgery at the UMC Utrecht. The main goal of including the patient data in the current manuscript was to provide proof of principle for our presented method, specifically to test if the presence of a tumor would not affect the full brain activation patterns in such a severe manner that all patterns would become atypical and thus invalidate our approach.

Method

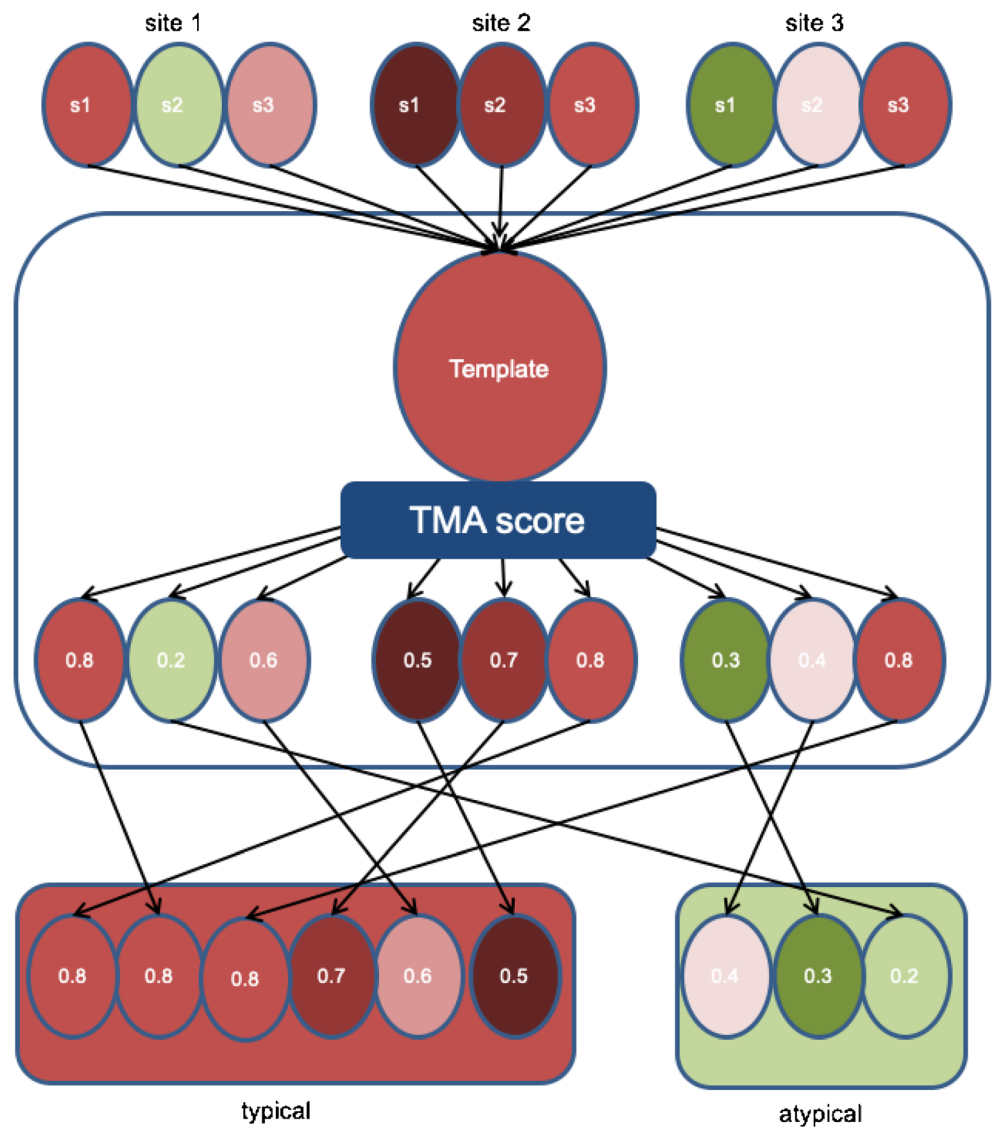
Subjects

For this study, fMRI data were included from centers in Austria (Graz), Germany (Frankfurt and Regensburg), Italy, (Milan), Spain (Madrid), and the Netherlands (Utrecht, Tilburg) (Table 1). This approach allows for assessment of differences between centers in situations that are most comparable to a real-life application of the protocols, hence with local language and routines.

Each site provided data from three male and three female healthy subjects aged between 18 and 30 years. All subjects were right-handed, with good eyesight (contact lenses were allowed), no history of neurological or psychiatric disease, and no use of medication other than contraceptives. All centers provided the data fully anonymized and devoid of any identifiable information. All participants gave approval for use of their anonymized data for the research. The Ethics Committee of the University Medical Center Utrecht determined that the healthy volunteer study did not require formal ethics approval, because all data were previously obtained and were fully anonymized, and all participants approved sharing of their fully anonymized data.

Additionally, clinical data from 32 consecutive patients who underwent surgery for low-grade glioma at the University Medical Center Utrecht over a period of 4 years (18 m/14 f, mean age 39 years, range 18–60 years) were also included. The patients at the University Medical Center Utrecht participated in a study that was approved by the Ethics Committee of the center, and signed informed consent, in accordance with the Declaration of Helsinki (2013). A general description of the tumor characteristics of the included patients is provided in Table 2.

Fig. 1 Schematic overview of the application of automatic assessment of typicality for clinical fMRI



Task protocols

The instructions to technical personnel were written in detail in a manual describing the fMRI procedure and data storage. The tasks were explained to the volunteers in accordance with

written instructions for the technicians. Volunteers did not practice the tasks before the fMRI scans. The two protocols were distributed on a DVD to be played on a DVD player and presented to subjects in the scanner. The menus and instructions on the DVD and the list of nouns for the verb generation

Table 1 Scan characteristics per site

Site	Country	Brand	Model	T	Pulse sequence	TE	TR	FA	Reps
1	Netherlands	Philips	Achieva	3.0	PRESTO	41	0.75	10	560
2	Netherlands	Philips	Achieva	3.0	PRESTO	27	1.5	10	280
3	Spain	GE	Signa HDxt	3.0	EPI	35	2.5	85	168
4	Italy	Siemens	Avanto AB	1.5	EPI	50	2.5	85	168
5	Germany	Siemens	Allegra	3.0	EPI	35	2.5	80	168
6	Germany	Siemens	Allegra	3.0	EPI	35	2.0	80	211
7	Austria	Siemens	MAGNETOM Trio	3.0	EPI	35	2.5	85	168

Table 2 Patient demographic data and tumor description

Subject	Demo				Tumor location						Tumor histology	
	TMA	Sex	Age	Size	FRO	TMP	OCC	PAR	INS	NS	Grade	Characteristics
1	0.63	M	41	x		L					III	Anaplastic oligo-astrocytoma
2	0.51	F	37	x		R			R		II	Low-grade oligo-astrocytoma
3	0.50	M	63						L		II	Low-grade oligo-astrocytoma
4	0.59	F	21	x		R					II	Low-grade astrocytoma
5	0.34	F	49			L					II	Low-grade oligo-astrocytoma
6	0.48	M	42		L						IV	Glioblastoma multiforme
7	0.54	F	30		L					L	I	Pilocytic astrocytoma
8	0.50	M	31		L						II	Low-grade astrocytoma
9	0.23	F	59		L				L		III	Anaplastic oligo-astrocytoma
10	0.43	F	38		L						II	Low-grade oligodendroglioma
11	0.40	M	20					L			III	Anaplastic oligo-astrocytoma
12	0.51	F	37	x	M						III	Anaplastic astrocytoma
13	0.31	F	42	x		L			L		II	Low-grade oligodendroglioma
14	0.38	M	43	x	L				L		III	Anaplastic oligo-astrocytoma
15	0.50	M	60			L					IV	Glioblastoma multiforme
16	0.62	M	25		L				L		II	Low-grade astrocytoma
17	0.50	M	42			L			L		II	Anaplastic oligo-astrocytoma
18	0.68	M	34		R						II	Low-grade astrocytoma
19	0.65	M	26		L						I	Ganglioglioma
20	0.47	F	57			R			R		III	Anaplastic oligo-astrocytoma
21	0.43	M	37	x	R	R			R		III	Anaplastic astrocytoma
22	0.07	M	18					L			I	Pilocytic astrocytoma
23	0.50	M	34		R				R		III	Anaplastic oligo-astrocytoma
24	0.34	F	38			L		L			I	Unknown
25	0.67	F	33			L					I	Ganglioglioma
26	0.33	F	46			L		L			II	Low-grade astrocytoma
27	0.46	M	40			L			L		I	Pilocytic astrocytoma
28	0.46	F	40		L						IV	Glioblastoma multiforme
29	0.34	F	38	x	L						II	Low-grade astrocytoma
30	0.56	M	41	x	R	R			R		II	Low-grade astrocytoma
31	0.44	M	24	x	L						II	Low-grade oligo-astrocytoma
32	0.29	M	51	x	L						IV	Glioblastoma multiforme

Demo demographical data: *F* female, *M* male; size: *x* clear anatomical displacement; tumor location: *FRO* frontal lobe, *OCC* occipital, *TMP* temporal lobe, *PAR* parietal lobe, *INS* insula, *NS* not specified, *L* left hemisphere, *R* right hemisphere, *M* midline

(VG) were translated to the local language for each site. All tasks started with 15 s of rest, followed by 7 tasks blocks of 30 s, interleaved with rest periods of 30 s, and ended with a 15-s rest period. Hence, the total time for each task was equal (7 min). The following instructions were provided to each site for each protocol:

1. Finger tapping (“FT”). Instructions to subject: “You will see a circle on the screen. Sometimes it is red and sometimes it is green. When it is red you will lie still and relax. When it is green the circle will flash on and off. You will then touch each of your fingers of your right hand with the

right thumb, one by one, at the rhythm of the green flashing circle. Do not move your arm. Your left hand stays relaxed for the whole task.” Frequency of the movement was 1 Hz.

2. Verb generation (“VG”). Instructions to subject: “You will see words, or black bars on the screen. When the bars appear you do nothing and lie still and relaxed. When a word appears it will be a noun. Think of what you can do with it and then imagine saying: ‘With that I can ...’ or ‘That I can ...’ For example: when you see the word ‘chair,’ imagine saying ‘That I can sit on.’ Do not speak because then you will move your head and the scans

become unusable. If you cannot think of what you can do with a word, skip it and continue with the next word.” Time per word was 3 s.

fMRI data acquisition

For each site, only the following scan parameters were standardized: FOV 256, scan matrix 64 by 64, slice thickness (including gap) 4 mm, 30 slices in transverse orientation parallel to the Sylvian fissure, one slice above the top of the brain. All centers used a quadrature head coil. Scanner equipment and scan protocol are detailed in Table 1. Field strength varied between 1.5 and 3.0 T. Scan time per volume varied between 0.75 and 2.5 s. Five centers used an EPI sequence; two centers used a PRESTO sequence [24].

fMRI analysis

All the fMRI datasets were sent to one site (UMC Utrecht, the Netherlands) for analysis after anonymization. All data processing and analysis was performed using IDL 8.2 (ITT Exelis Inc. McLean, VA), unless otherwise specified. All statistical tests were performed using SPSS 20.

Scans were registered to the last functional scan to correct for movement [25]. Scans were smoothed with a 3D Gaussian filter (full width at half maximum 12 mm) to minimize effects of functional anatomical differences between subjects. Scans were spatially normalized to a standard EPI template in MNI space (from the SPM5 template library) using linear transformation incorporated in FSL software [26]. All time series were normalized to a mean value of 100 per voxel to allow for comparisons between subjects. Temporal filtering was applied to remove low-frequency drifts (linear, first- and second-order slow trends [27]). All tasks were analyzed using a GLM, with the regressor of interest constructed using a canonical HRF model incorporated in SPM 5 (Wellcome Trust Center for Neuroimaging, London, UK), convolved with the boxcar input function of each protocol. For TMA, the b-maps generated by the GLM were used.

The TMA value that we introduce in this study is identical to the Pearson’s r value, calculated over the beta values from all voxels in a template activation pattern and an individual activation map, without application of any selection or any application of an activation threshold value. The TMA was calculated for each subject using a template based on data from the other six centers. Thus, each value was based on a comparison with an independent template (“leave-one-center-out” approach).

To examine the feasibility of TMA to automatically identify atypical activation patterns, we artificially created a large set of “atypical” scans (from the set of 42 scans) that included some common problems that can occur in a clinical setting: a

first set of atypical activation patterns was created by reversing the orientation over the y -axis to create an incorrect left-right orientation (“REVERSED”). A second set was created by using the activation pattern from the other task (“TASK”), and a third set was constructed by applying an incorrect analysis, using task regressors that were shifted 15 s (“SHIFT”). We categorized all fMRI scans from the healthy subjects as “typical.”

A brain tumor can affect activation patterns both physically by the tumor itself and indirectly due to for instance functional plasticity. If the majority of fMRI scans of brain tumor patients show highly atypical activation patterns, the usefulness of a TMA is strongly reduced. Thus, it is important to provide a proof of principle for the feasibility of TMA for scans of patients with a brain tumor. For this reason, we applied TMA on fMRI data of 32 tumor patients at the UMC Utrecht who had performed the VG task for clinical purposes and compared the values to those of the healthy controls. The clinical fMRI procedure involves the same VG task as used in the healthy volunteers and includes training before entering the scanner. For additional validation, we compared the patient results to atypical SHIFT scans, atypical TASK scans, as well as atypical REVERSED scans, using an independent samples t test after a Fisher Z transformation was applied to each TMA value to ensure normal distribution.

Automatic detection of atypical activation patterns

In the TMA approach, it is important to minimize incorrect identification of activity patterns, i.e., false negatives (when an atypical pattern is mistaken for being typical) and false positives (when a typical pattern is mistaken for being atypical). We examined for each task the false positive rate, after determining the lowest TMA value threshold that correctly identifies all atypical scans.

Results

Comparison of centers

Data from each site were compared with data from all other centers. For each site test (per task), a template was built of all data from the other six centers. For each subject in the tested site, the task b-map was compared with the template by means of the TMA calculation. Hence, from each single individual, two TMA values were obtained. These values were averaged per site and compared across centers.

Figure 2 presents group activation maps of each site and task, as well as the grand average activation pattern over all 42 subjects. One can appreciate the likeness of activity patterns across centers. The average TMA values for each site and task are displayed in Fig. 3. Results indicate that all subjects from

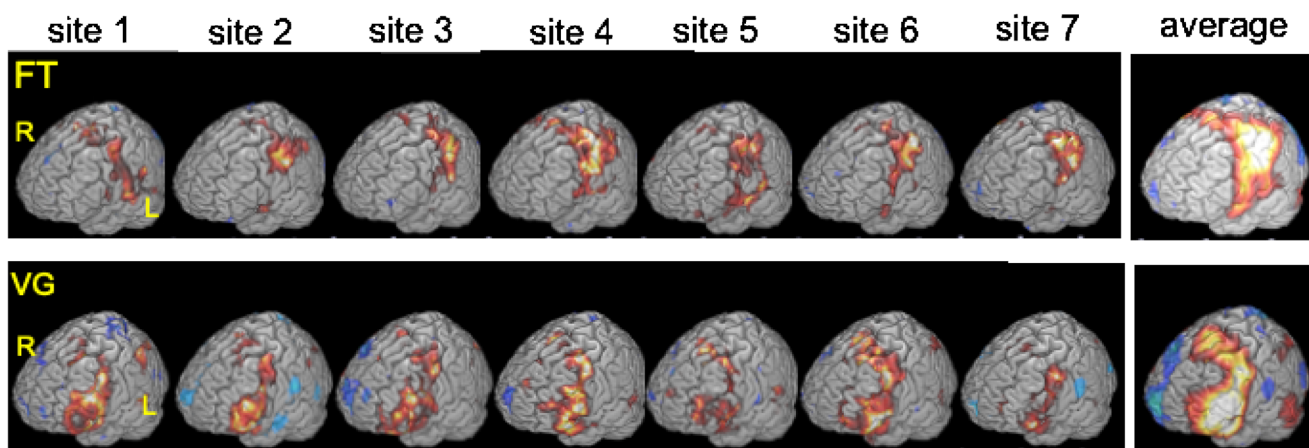


Fig. 2 Rendered group activation patterns per site and per task protocol ($|t| < 3.14$; $df = 6$; positive activity in red; negative activity in blue; FT, finger tap protocol; VG, verb generation protocol; L, left; R, right)

every site correlated high with the templates. This indicates that if a template is made based on a number of centers, the likelihood of that template being valid for evaluation of data from a new site is high.

All centers yielded a high mean TMA value for both tasks, ranging from 0.58 to 0.74 for the FT protocol and 0.66 to 0.72 for the VG protocol. TMA for individuals ranged from 0.36 to 0.82 for the FT protocol and from 0.42 to 0.78 for the VG protocol. We did find neither a difference in TMA values between the two protocols ($F(6,35) = 1.98$; $p = 0.22$) nor a site by protocol interaction ($F(6,35) = 1.33$; $p = 0.28$). Although all centers showed high TMA values, we did find a significant difference between centers in the mean values ($F(6,35) = 3.17$; $p = 0.009$).

Classification based on TMA

To assess feasibility of using TMA for assessment of pattern typicality, we compared TMA values of all individual subjects with TMA values computed from deliberately corrupted versions of the original scans. In Fig. 4, TMA values are displayed for all typical scans, as well as for atypical scans

due to incorrect orientation or incorrect task protocol. In Fig. 6, we provide “receiver operating characteristic curves” (“ROC curves”), showing the percentage correctly identified typical and atypical scans at a range of TMA threshold values for the FT and VG protocols. These figures clearly illustrate the power of the TMA method to correctly detect both typical and atypical scans for the FT and VG templates in healthy controls, as well for the VG template in patients (Fig. 5).

For the FT protocol, the average TMA value for a typical scan (correct task, correct analysis, correct orientation) was 0.65 (range 0.36–0.82). For atypical scans due to incorrect task protocol, the average TMA value was 0.24 (range –0.07 to 0.49). Of the typical scans, 95.2% had a higher TMA value than the highest value for an incorrect task protocol, so all atypical patterns could be detected, with a percentage of missed typical scans (i.e., typical scans misidentified as atypical or “false positives”) of 4.8% (or 2 in 42).

For incorrect orientation, the average TMA value was 0.27 (range –0.01 to 0.54). Of the typical scans, 85.7% had a TMA value higher than the highest value for a scan with incorrect orientation, so the percentage of false positives was 14.3% (6 in 42).

Fig. 3 TMA results for each subject per site (**a** FT (finger tapping), **b** VG (verb generation)). Horizontal bars represent mean TMA per site. While all centers showed high TMA values, we did find a significant difference between centers in the mean values ($F(6,35) = 3.17$; $p = 0.009$), which is likely the result of slight differences in signal to noise between the scanners of the different centers

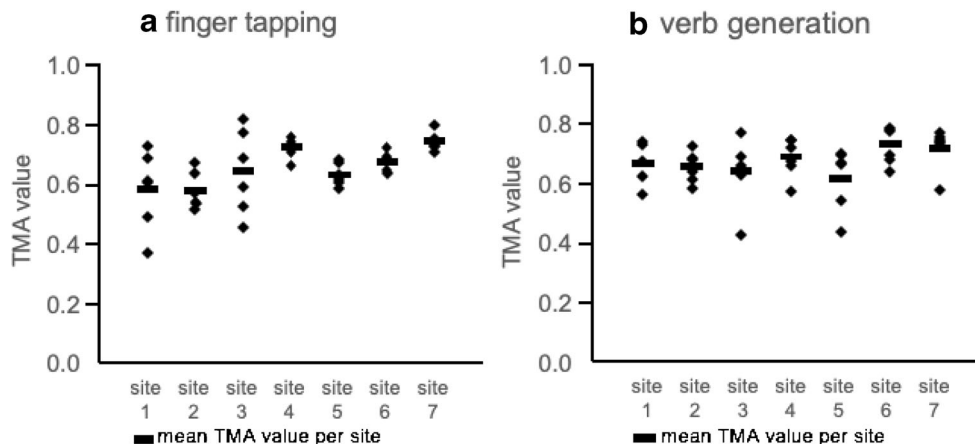
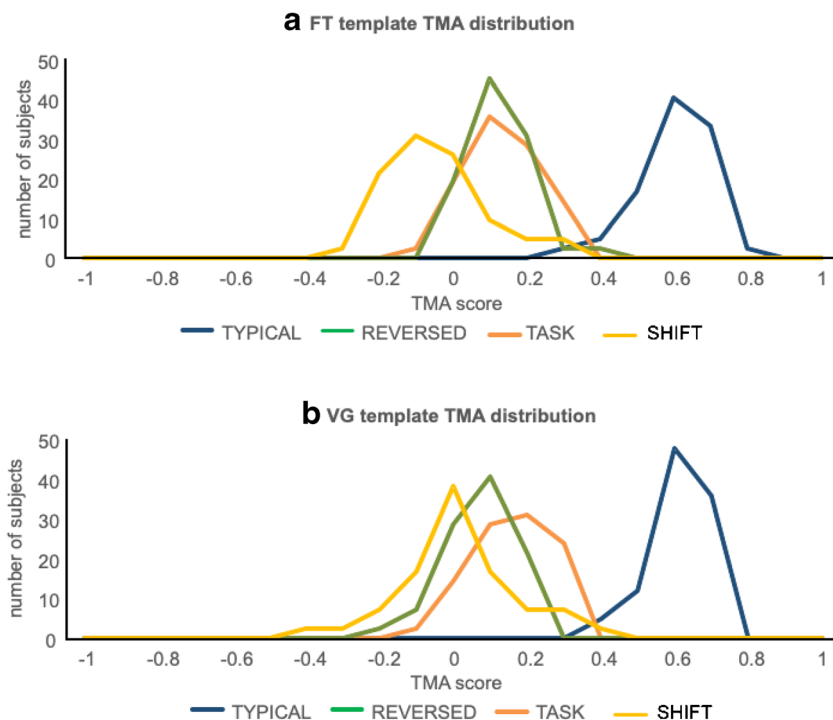


Fig. 4 Distribution of template matching analysis (TMA) values for typical scans (“TYPICAL,” blue), as well as for atypical scans due to reversed orientation (“REVERSED,” green), incorrect task protocol (“TASK,” orange), or incorrect analysis with a time-shifted regressor (“SHIFT,” yellow). **a** Finger tapping (FT) template. **b** Verb generation (VG) template



For incorrect analysis due to a shifted regressor, the average TMA value was 0.00 (range – 0.23 to 0.37). Of the typical scans, 97.6% had a TMA value higher than the highest value for a scan with incorrect analysis, resulting in a percentage of false positives of 2.4% (1 in 42).

Taken together, for the FT protocol, we were able to correctly classify 93.5% of the scans as typical, while detecting 100% of the atypical scans, using a TMA threshold of 0.49.

For the VG protocol, the average TMA value for a typical scan (correct task, correct analysis, correct orientation) was 0.67 (range 0.42–0.78).

For atypical scans due to an incorrect task protocol, the average TMA value was 0.25 (range – 0.10 to 0.47). Of the typical scans, 95.2% had a higher TMA value than the highest value for an incorrect task protocol, so the percentage of false positives was 4.8% (2 in 42).

For incorrect orientation, the average TMA value was 0.26 (range 0.06–0.51). Of the typical scans, 95.2% had a TMA value higher than the highest value for a scan with incorrect orientation, so the percentage of false positives was 4.8% (2 in 42).

For incorrect analysis due to a shifted regressor, the average TMA value was 0.06 (range – 0.25 to 0.43). Of the typical scans, 95.2% had a TMA value higher than the highest value

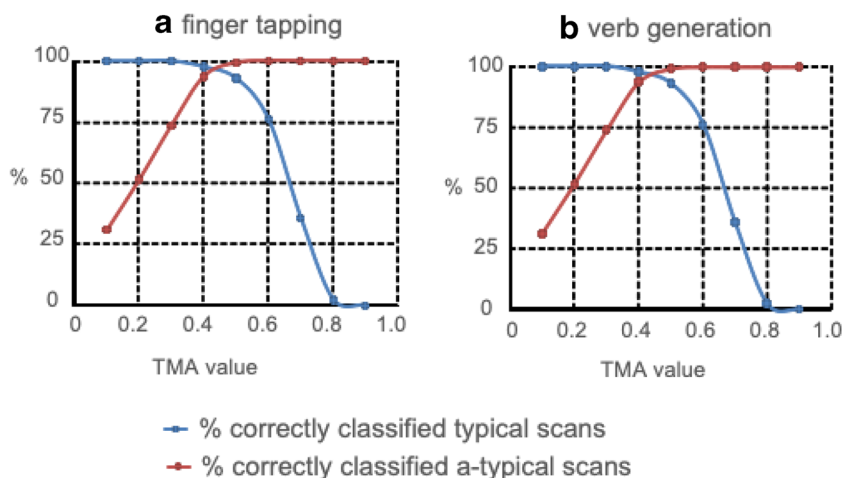


Fig. 5 Receiver operating characteristic curve for healthy subjects data of **a** FT and **b** VG tasks. Figures display the percent correct classification (y-axis), for a range of TMA values

for a scan with incorrect analysis, resulting in a percentage of false positives of 2.4% (1 in 42).

Taken together, for the VG protocol, we were able to correctly classify 96.4% of the scans as typical, while detecting 100% of the atypical scans, using a TMA value of 0.51.

In Fig. 6, we provide receiver operating characteristic curves (ROC curves), showing the percentage of correctly identified typical and atypical scans at a range of TMA threshold values for the FT and VG protocols. Figures 4 and 6 illustrate the ability of the TMA method to correctly distinguish between typical and atypical scans for the FT and VG templates in healthy controls.

Correlation between TMA and strength of activation

We also evaluated if the quality of the scan, approximated by the maximum t value of the activation pattern, was a factor that affected the TMA value. For this reason, we calculated, for both the VG and FT protocols, the Pearson correlation between the maximum t value and the TMA value. The maximum t value did not correlate with the TMA value for either protocol (FT: $r = 0.09$; $p = 0.58$; VG: $r = 0.14$; $p = 0.38$).

Template matching results for brain tumor patient data

We gathered 32 datasets of consecutive tumor patients who had executed the VG protocol at the UMC Utrecht for clinical purposes. Figure 5 shows the distribution of the TMA values of all patients, compared with typical healthy subject scans and the atypical sets of healthy subject scans. TMA values of the patient scans were significantly lower than those of healthy subjects ($t(72) = 8.64$; $p < 0.001$), but significantly higher than those of atypical SHIFT scans ($t(72) = 10.72$; $p < 0.001$), atypical TASK

scans ($t(72) = 6.91$; $p < 0.001$), and atypical REVERSED scans ($t(72) = 7.11$; $p < 0.001$).

Discussion

The main goal of this study was to evaluate a novel method for fast, objective, and automatic detection of atypical fMRI scans based on the activation patterns. The method is based on a template matching analysis (TMA). Activation patterns were collected from seven European clinical centers and two commonly used clinical fMRI tasks for localization of motor and language, and from patients with a brain tumor from the UMC Utrecht.

Results indicated that fast and automatic detection of atypical activation patterns appears to be a reachable goal in healthy subjects. For localization of motor activity, a detection rate of 93.5% of compromised scans could be achieved using the TMA approach, while for location of language activation, the VG protocol, a detection rate of 96.4% was achieved. The analysis of patient VG data suggests the feasibility of TMA in a clinical setting. Patients with a brain tumor did show slightly lower template matching values on average, which is likely due to the inclusion of poor-quality data due to movement and cognitive impairments, and anatomical and functional abnormalities related to the tumor. None of these confounders was considered for the purpose of this study, in order to obtain a conservative indication of the robustness of the TMA method.

Automatic detection of atypical activation patterns may prove helpful in clinical fMRI settings. While scans with a high TMA value can be forwarded for clinical use without need for expert quality assessment, experts can be consulted in cases where the scan has a low TMA value in order to determine the cause, such as movement, failure to comply with task rules by the patient or suboptimal analysis, versus abnormal activity due to severe anatomical deformation or

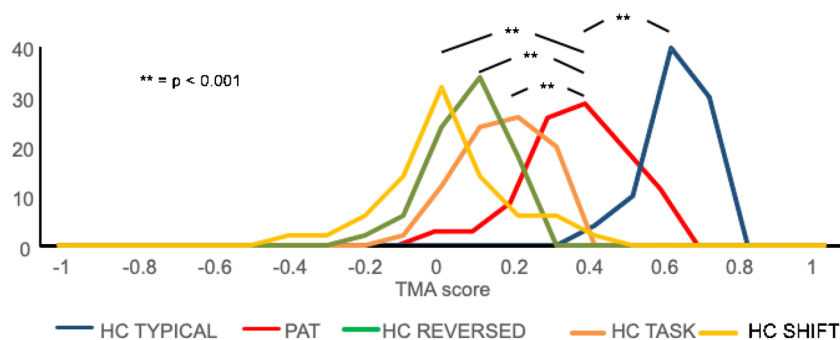


Fig. 6 Distribution of template matching analysis (TMA) values for the verb generation (VG) template for correctly analyzed healthy controls (“TYPICAL,” blue), tumor patient data (“PAT,” red), as well as for atypical scans due to reversed orientation (“REVERSED,” green), incorrect

task protocol (“TASK,” orange), or incorrect analysis with a time-shifted regressor (“SHIFT,” yellow). Patient TMA values were significantly higher than SHIFT ($t(72) = 8.33$; $p < 0.001$), TASK ($t(72) = 8.33$; $p < 0.001$), as well as REVERSED scans ($t(72) = 8.33$; $p < 0.001$)

functional plasticity. Thus, TMA may improve the quality of care by providing an objective and reproducible assessment of the fMRI activation pattern, and it can increase the cost efficiency by reducing the need for fMRI experts to analyze and interpret each fMRI dataset.

An alternative use of our method can be found in applications for large databases: TMA can automatically, at low cost and minimal manpower, identify datasets that are potentially compromised, thereby rapidly indicating potential site-specific problems and improving the quality of the database. This could facilitate studies that aim at gathering large databases over longer time or studies that need to pool fMRI data from various centers with potential systematic differences in quality.

It can be expected that the presence of a brain tumor changes the anatomy of a brain due to its effects on tissue integrity and displacement, as well as the functional anatomy, as it may disrupt functional networks. While the patients did display lower TMA values than healthy controls, the values were significantly higher than those of the atypical healthy subject scans (Fig. 5). A lower mean TMA value is to be expected since the set is likely to include poor-quality data (due to for instance movement, non-compliance, compromised cognitive function) as well as functional and large anatomical changes caused by the tumor.

One possible explanation for the limited tumor effects on TMA is that the protocols that were used evoked a pattern of activations and deactivations that are widespread over the brain. Thus, although the anatomy as well as functional activity pattern may change in the vicinity of the tumor, the activity pattern across the whole scan volume appears to be more stable and not severely affected by the tumor, resulting in only a somewhat lower than normal value for atypicality for most patients. Of note, task compliance and movement were not taken into account in order to obtain a conservative indication of robustness of TMA.

An efficient and reliable language protocol is a particularly important tool for presurgical planning as well as research [17, 28–31]. Clinical fMRI can locate the presence of crucial language activation in the immediate vicinity of a tumor. This would mandate a cautionary surgical approach and can be a reason to perform awake surgery or even advise against resective surgery. Importantly, fMRI may fail at the edge or within the tumor, due to abnormal vascular properties and brain tissue which can reduce the fMRI signal [32], and as such can only be used in conjunction with other clinical diagnostic modalities.

One of the important and possibly counterintuitive findings of this study is that the language protocol performed comparable with the motor protocol and showed robust results, despite the more complex nature of language production compared with motor performance and execution in four languages. This result suggests that the pattern of activity evoked

by language production is as consistent over subjects as that of motor production.

Our study also demonstrated that fMRI patterns do not differ strongly between centers. Previous multisite studies, applying simple motor protocols [33–38], visual protocols [39, 40], and sometimes cognitive protocols [41–43], have consistently reported promising results for the possibility to combine fMRI data across centers. Site variation typically appears to be much smaller than subject variation (see for instance Costafreda [44] for a review). However, previous multisite studies have predominantly evaluated location or strength of activation at a specific predefined area that was expected to be activated by the protocol. TMA indicates that whole-brain activation patterns associated with language as well as motor function are consistent over centers with a considerable range in hardware and scan protocols (Table 1).

The fMRI tasks were chosen based on extensive use across centers and validation against direct electrical stimulation in brain surgery [45, 46] and electrocortical mapping [47, 48]. The tasks are easy to understand and perform for most surgery patients and yield robust levels of activation. Apart from simplicity to promote patient compliance, characteristics that may be important in robustness include a block design (as opposed to a less sensitive event-related design [49]) and non-ambiguity of the task to avoid use of different cognitive strategies. Yet, patients should be tested before an fMRI scan to ascertain cognitive capability and ability to sustain attention.

The robustness of a standardized protocol can be applied to improve reliability of clinical fMRI, facilitate analysis, and interpretation of results. The quality of a fMRI scan is vulnerable and can be affected by many factors, such as subject movement, task compliance, and incorrect analysis steps. In many cases, it is difficult to objectively judge the quality of an fMRI scan, other than through visual inspection of the activity pattern, and as of yet, there is no commonly accepted metric for quality. The current results indicate that TMA provides a simple but effective metric that can be evaluated in future studies against expert quality assessment.

Robust protocols can also facilitate studies that are plagued by low power because of difficulties in recruitment due to for example low incidence rates of specific pathologies (e.g., low-grade glioma at a particular location) or surgical procedures [50]. It allows for the possibility to pool data from various centers for clinical fMRI studies, without substantial loss in statistical power. This approach can for instance be used for patient follow-up studies that aim at examining the association between surgery and patient outcome, especially in terms of the presence of functional plasticity, or for cohort studies, where large populations are required.

There are several limitations that have to be considered in interpreting the results of this study. Importantly, while the patient results do successfully indicate that presence of a tumor does not necessarily have an invalidating impact on

TMA, it will take more patient data and further expert evaluation in order to determine a firm TMA threshold for clinical use.

Our study did not evaluate the relevance of the information acquired by the task protocols but was limited to the reproducibility of the pattern across individuals. Our study did not evaluate differences in the level of the activity between centers, but instead focused on agreement between patterns. Possibly, there may have been differences in levels of activity between centers due to hardware and scan sequence differences. However, for clinical use, we do believe that the pattern of activity may be more informative than absolute levels.

Another limitation is that patients were not evaluated for quality of the scans, leaving the question about an appropriate TMA threshold for clinical use open. In the patient dataset used here, it is not known how many were of poor quality or how many exhibited lack of compliance due to cognitive or attentional deficits. This requires an extensive study comparing TMA in patient data with multiple quality measures including subjective evaluations by experts which did not fall within the scope of this proof-of-principle study. Moreover, we did not apply correction for abnormalities in anatomical structure caused by the tumor. But based on the current results, techniques that are able to correct for those, for instance, by non-linear “warping” techniques for normalizing a patient anatomical scan to a standard brain [51], are likely to further improve robustness of TMA. Finally, future studies as well as applications need to indicate if TMA can also be used to detect more subtle forms of atypical activation, for instance, if they are associated with poor task performance due to cognitive deficits.

As a final note, one could argue that head motion could increase the TMA value by generating false activations. We argue that motion artifacts can only elevate TMA if motion causes increases and decreases in various regions in such a way that they match the whole-brain template. Given the nature of data acquisition, this is highly unlikely since motion affects whole slices (and generally multiple) and is most often manifested along intensity edges (typically gray matter vs CSF) across the brain. Neither is likely to generate the template pattern. In effect, even though task-correlated motion may generate spurious significant activity along intensity edges, motion will reduce the TMA that is computed across the brain because it adds the image noise resulting from motion to the data.

In conclusion, we present a straightforward method for assessing typicality of fMRI activation patterns and validate it with correctly and incorrectly analyzed data from 42 healthy subjects from 7 centers. The results indicate that TMA is robust and detects the atypical activity patterns derived from the incorrectly analyzed quite well. The application of TMA to a

series of patient data suggests that the presence of localized anatomic abnormalities does not invalidate the method (although TMA values were lower than that of healthy subjects). Of note, it is of importance for clinical fMRI to optimize the scan procedure in order to avoid artifacts, preferably by applying some sort of quality assurance.

Funding information No funding was received for this study.

Compliance with ethical standards

Conflicts of interest statement Author NR declares a conflict of interest. He is the director, and owns stock, of startup company Braincarta BV. All other authors declare that they have no conflict of interest.

Ethics approval All procedures performed in the studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Castellano A, Cirillo S, Bello L, Riva M, Falini A (2017) Functional MRI for surgery of gliomas. *Curr Treat Options Neurol* 19:34
2. Dimou S, Battisti RA, Hermens DF, Lagopoulos J (2013) A systematic review of functional magnetic resonance imaging and diffusion tensor imaging modalities used in presurgical planning of brain tumour resection. *Neurosurg Rev* 36:205–214
3. Rutten GJ, Ramsey NF (2010) The role of functional magnetic resonance imaging in brain surgery. *Neurosurg Focus* 28:E4
4. Wang L, Chen D, Olson J, Ali S, Fan T, Mao H (2012) Re-examine tumor-induced alterations in hemodynamic responses of BOLD fMRI: implications in presurgical brain mapping. *Acta Radiol* 53:802–811
5. Bick AS, Mayer A, Levin N (2012) From research to clinical practice: implementation of functional magnetic imaging and white matter tractography in the clinical environment. *J Neurol Sci* 312:158–165
6. Rutten G (2002) fMRI-determined language lateralization in patients with unilateral or mixed language dominance according to the Wada test. *NeuroImage* 17:447–460

7. Bauer PR, Reitsma JB, Houweling BM, Ferrier CH, Ramsey NF (2014) Can fMRI safely replace the Wada test for preoperative assessment of language lateralisation? A meta-analysis and systematic review. *J Neurol Neurosurg Psychiatry* 85:581–588
8. Arora J, Pugh K, Westerveld M, Spencer S, Spencer DD, Todd Constable R (2009) Language lateralization in epilepsy patients: fMRI validated with the Wada procedure. *Epilepsia* 50:2225–2241
9. Binder JR (2011) Functional MRI is a valid noninvasive alternative to Wada testing. *Epilepsy Behav* 20:214–222
10. McKinsey RD, Moritz CH, Meyerand ME, Tome WA (2010) Assessment of multiple task activation and reproducibility in patients with benign and low-grade neoplasm. *Technol Cancer Res Treat* 9:319–326
11. Wengenroth M, Blatow M, Guenther J, Akbar M, Tronnier VM, Stippich C (2011) Diagnostic benefits of presurgical fMRI in patients with brain tumours in the primary sensorimotor cortex. *Eur Radiol* 21:1517–1525
12. Weng HH, Noll KR, Johnson JM, Prabhu SS, Tsai YH, Chang SW, Huang YC, Lee JD, Yang JT, Yang CT, Tsai YH, Yang CY, Hazle JD, Schomer DF, Liu HL (2018) Accuracy of presurgical functional MR imaging for language mapping of brain tumors: a systematic review and meta-analysis. *Radiology* 286:512–523
13. Tyndall AJ, Reinhardt J, Tronnier V, Mariani L, Stippich C (2017) Presurgical motor, somatosensory and language fMRI: technical feasibility and limitations in 491 patients over 13 years. *Eur Radiol* 27:267–278
14. Sair HI, Yahyavi-Firouz-Abadi N, Calhoun VD, Airan RD, Agarwal S, Intrapromkul J, Choe AS, Gujar SK, Caffo B, Lindquist MA, Pillai JJ (2016) Presurgical brain mapping of the language network in patients with brain tumors using resting-state fMRI: comparison with task fMRI. *Hum Brain Mapp* 37:913–923
15. Gould L, Mickleborough MJ, Wu A, Tellez J, Ekstrand C, Lorentz E, Ellchuk T, Babyn P, Borowsky R (2016) Presurgical language mapping in epilepsy: using fMRI of reading to identify functional reorganization in a patient with long-standing temporal lobe epilepsy. *Epilepsy Behav Case Rep* 5:6–10
16. Kuchcinski G, Mellerio C, Pallud J, Dezamis E, Turc G, Rigaux-Viode O, Malherbe C, Roca P, Leclerc X, Varlet P, Chretien F, Devaux B, Meder JF, Oppenheim C (2015) Three-tesla functional MR language mapping: comparison with direct cortical stimulation in gliomas. *Neurology* 84:560–568
17. Austermuehle A, Cocjin J, Reynolds R, Agrawal S, Sepeta L, Gaillard WD, Zaghoul KA, Inati S, Theodore WH (2017) Language functional MRI and direct cortical stimulation in epilepsy preoperative planning. *Ann Neurol* 81:526–537
18. Giussani C, Roux FE, Ojemann J, Sganzerla EP, Pirillo D, Papagno C (2010) Is preoperative functional magnetic resonance imaging reliable for language areas mapping in brain tumor surgery? Review of language functional magnetic resonance imaging and direct cortical stimulation correlation studies. *Neurosurgery* 66:113–120
19. Petrella JR, Shah LM, Harris KM, Friedman AH, George TM, Sampson JH, Pekala JS, Voyvodic JT (2006) Preoperative functional MR imaging localization of language and motor areas: effect on therapeutic decision making in patients with potentially resectable brain tumors. *Radiology* 240:793–802
20. Beisteiner R, Robinson S, Wurnig M, Hilbert M, Merksa K, Rath J, Hollinger I, Klingner N, Marosi C, Trattnig S, Geissler A (2011) Clinical fMRI: evidence for a 7T benefit over 3T. *Neuroimage* 57:1015–1021
21. Khorrami MS, Faro SH, Seshadri A, Moonat S, Lidicker J, Hershey BL, Mohamed FB (2011) Functional MRI of sensory motor cortex: comparison between finger-to-thumb and hand squeeze tasks. *J Neuroimaging* 21:236–240
22. Rutten GJ, van Rijen PC, van Veelen CW, Ramsey NF (1999) Language area localization with three-dimensional functional magnetic resonance imaging matches intrasulcal electrostimulation in Broca's area. *Ann Neurol* 46:405–408
23. Kunii N, Kamada K, Ota T, Kawai K, Saito N (2011) A detailed analysis of functional magnetic resonance imaging in the frontal language area: a comparative study with extraoperative electrocortical stimulation. *Neurosurgery* 69:590–596
24. Neggers SF, Hermans EJ, Ramsey NF (2008) Enhanced sensitivity with fast three-dimensional blood-oxygen-level-dependent functional MRI: comparison of SENSE-PRESTO and 2D-EPI at 3 T. *NMR Biomed* 21:663–676
25. Thevenaz P, Ruttimann UE, Unser M (1998) A pyramid approach to subpixel registration based on intensity. *IEEE Trans Med Imaging* 7:27–41
26. Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841
27. Tanabe J, Miller D, Tregellas J, Freedman R, Meyer FG (2002) Comparison of detrending methods for optimal fMRI preprocessing. *Neuroimage* 15:902–907
28. Jansma JM, Ramsey N, Rutten GJ (2015) A comparison of brain activity associated with language production in brain tumor patients with left and right sided language laterality. *J Neurosurg Sci* 59:327–335
29. Batouli SA, Hasani N, Gheisari S, Behzad E, Oghabian MA (2016) Evaluation of the factors influencing brain language laterality in presurgical planning. *Phys Med* 32:1201–1209
30. Hou BL, Holodny AI, Cooperman N, Gutin PH (2006) Reorganization of the cortical control of movement due to radiation necrosis. Case report. *J Neurosurg* 104:147–149
31. Tailby C, Abbott DF, Jackson GD (2017) The diminishing dominance of the dominant hemisphere: language fMRI in focal epilepsy. *Neuroimage Clin* 14:141–150
32. Hou BL, Bradbury M, Peck KK, Petrovich NM, Gutin PH, Holodny AI (2006) Effect of brain tumor neovasculature defined by rCBV on BOLD fMRI activation volume in the primary motor cortex. *Neuroimage* 32:489–497
33. Rath J, Wurnig M, Fischmeister F, Klingner N, Hollinger I, Geissler A, Aichhorn M, Foki T, Kronbichler M, Nickel J, Siedentopf C, Staffen W, Verius M, Golaszewski S, Koppelstaetter F, Auff E, Felber S, Seitz RJ, Beisteiner R (2016) Between- and within-site variability of fMRI localizations. *Hum Brain Mapp* 37:2151–2160
34. Friedman L, Stern H, Brown GG, Mathalon DH, Turner J, Glover GH, Gollub RL, Lauriello J, Lim KO, Cannon T, Greve DN, Bockholt HJ, Belger A, Mueller B, Doty MJ, He J, Wells W, Smyth P, Pieper S, Kim S, Kubicki M, Vangel M, Potkin SG (2008) Test-retest and between-site reliability in a multicenter fMRI study. *Hum Brain Mapp* 29:958–972
35. Costafreda SG, Brammer MJ, Vencio RZ, Mourao ML, Portela LA, de Castro CC, Giampietro VP, Amaro E Jr (2007) Multisite fMRI reproducibility of a motor task using identical MR systems. *J Magn Reson Imaging* 26:1122–1126
36. Maitra R, Roys SR, Gullapalli RP (2002) Test-retest reliability estimation of functional MRI data. *Magn Reson Med* 48:62–70
37. Gountouna VE, Job DE, McIntosh AM, Moorhead TW, Lymer GK, Whalley HC, Hall J, Waiter GD, Brennan D, McGonigle DJ, Ahearn TS, Cavanagh J, Condon B, Hadley DM, Marshall I, Murray AD, Steele JD, Wardlaw JM, Lawrie SM (2010) Functional magnetic resonance imaging (fMRI) reproducibility and variance components across visits and scanning sites with a finger tapping task. *Neuroimage* 49:552–560
38. Voyvodic JT (2006) Activation mapping as a percentage of local excitation: fMRI stability within scans, between scans and across field strengths. *Magn Reson Imaging* 24:1249–1261
39. Vlieger EJ, Lavini C, Majoie CB, den Heeten GJ (2003) Reproducibility of functional MR imaging results using two different MR systems. *AJNR Am J Neuroradiol* 24:652–657

40. Zou KH, Greve DN, Wang M, Pieper SD, Warfield SK, White NS, Manandhar S, Brown GG, Vangel MG, Kikinis R, Wells WM 3rd (2005) Reproducibility of functional MR imaging: preliminary results of prospective multi-institutional study performed by Biomedical Informatics Research Network. *Radiology* 237:781–789
41. Brown GG, Mathalon DH, Stern H, Ford J, Mueller B, Greve DN, McCarthy G, Voyvodic J, Glover G, Diaz M, Yetter E, Ozyurt IB, Jorgensen KW, Wible CG, Turner JA, Thompson WK, Potkin SG (2011) Multisite reliability of cognitive BOLD data. *Neuroimage* 54:2163–2175
42. Casey BJ, Cohen JD, O'Craven K, Davidson RJ, Irwin W, Nelson CA, Noll DC, Hu X, Lowe MJ, Rosen BR, Truwitt CL, Turski PA (1998) Reproducibility of fMRI results across four institutions using a spatial working memory task. *NeuroImage* 8:249–261
43. Gee DG, McEwen SC, Forsyth JK, Haut KM, Bearden CE, Addington J, Goodyear B, Cadenhead KS, Mirzakhani H, Comblatt BA, Olvet D, Mathalon DH, McGlashan TH, Perkins DO, Belger A, Seidman LJ, Thermenos H, Tsuang MT, van Erp TG, Walker EF, Hamann S, Woods SW, Constable T, Cannon TD (2015) Reliability of an fMRI paradigm for emotional processing in a multisite longitudinal study. *Hum Brain Mapp* 36:2558–2579
44. Costafreda S (2009) Pooling fMRI data: meta-analysis, mega-analysis and multi-center studies. *Front Neuroinformatics* 3:33
45. Rutten GJ, Ramsey NF, van Rijen PC, Noordmans HJ, van Veelen CW (2002) Development of a functional magnetic resonance imaging protocol for intraoperative localization of critical temporoparietal language areas. *Ann Neurol* 51:350–360
46. Roessler K, Donat M, Lanzenberger R, Novak K, Geissler A, Gartus A, Tahamtan AR, Milakara D, Czech T, Barth M, Knosp E, Beisteiner R (2005) Evaluation of preoperative high magnetic field motor functional MRI (3 Tesla) in glioma patients by navigated electrocortical stimulation and postoperative outcome. *J Neurol Neurosurg Psychiatry* 76:1152–1157
47. Hermes D, Miller KJ, Vansteensel MJ, Aarnoutse EJ, Leijten FS, Ramsey NF (2012) Neurophysiologic correlates of fMRI in human motor cortex. *Hum Brain Mapp* 33:1689–1699
48. Brunner P, Ritaccio AL, Lynch TM, Emrich JF, Wilson JA, Williams JC, Aarnoutse EJ, Ramsey NF, Leuthardt EC, Bischof H, Schalk G (2009) A practical procedure for real-time functional mapping of eloquent cortex using electrocorticographic signals in humans. *Epilepsy Behav* 15:278–286
49. Liu TT, Frank LR, Wong EC, Buxton RB (2001) Detection power, estimation efficiency, and predictability in event-related fMRI. *Neuroimage* 13:759–773
50. Thirion B, Pinel P, Meriaux S, Roche A, Dehaene S, Poline JB (2007) Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. *Neuroimage* 35:105–120
51. Crinion J, Ashburner J, Leff A, Brett M, Price C, Friston K (2007) Spatial normalization of lesioned brains: performance evaluation and impact on fMRI analyses. *Neuroimage* 37:866–875

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.