ELSEVIER

We propose guidelines for 'good imaging practice' (GIP) in the application of fMRI as a biomarker in drug development, based around mitigating risks associated with the complexities of the technique.

A procedural framework for good imaging practice in pharmacological fMRI studies applied to drug development #2: protocol optimization and best practices

Adam J. Schwarz[1,2,7], Lino Becerra[1,3,6,7], Jaymin Upadhyay[1,3], Julie Anderson[1,3], Richard Baumgartner[1,4], Alexander Coimbra[1,4], Jeff Evelhoch[1,4], Richard Hargreaves[1,4], Brigitte Robertson[5], Smriti Iyengar[1,2], Johannes Tauscher[1,2], David Bleakman[1,2] and David Borsook[1,3,6]

[1] Imaging Consortium for Drug Development, McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA
[2] Lilly Research Laboratories, Eli Lilly & Company, Indianapolis, IN 46285, USA
[3] P.A.I.N. Group, Brain Imaging Center, McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA
[4] Merck Research Laboratories, West Point, PA 19486, USA
[5] Sepracor Inc., Marlborough, MA 01752, USA
[6] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA

Functional magnetic resonance imaging (fMRI) experiments are more complex compared with standard radiological imaging, involving additional data streams and hardware along with complex analysis methods. Here, we propose guidelines based around mitigating risks associated with the complexities of the technique at the level of the individual imaging protocol, including workable and effective quality assurance/quality control procedures and rigorous, predefined, analysis pipelines. Our aim is to provide a framework for 'good imaging practice' (GIP), enabling these requirements to be addressed at an appropriate level of detail. The development of a procedural framework for GIP in pharmaceutical fMRI studies could lead to greater acceptance of the method within industry and facilitate validation and, eventually, qualification of the technique as an imaging biomarker.

Introduction

A general issue for the use of imaging biomarkers in drug development is to ensure that relevant aspects of the imaging procedures are captured at an appropriate level of detail in terms of both acquisition and analysis. Anatomical magnetic resonance imaging (MRI) methods are already widely used in clinical trials for drug development [1]; accordingly, imaging study management

Corresponding author:. Schwarz, A.J. (a.schwarz@lilly.com)
[7] These authors contributed equally to this article.

Adam J Schwarz received a Bachelor of Science with First Class Honors in Physics (1991) and a PhD in Electrical and Electronic Engineering (1995) from the University of Canterbury, New Zealand. Following a postdoctoral position at the Institute of Cancer Research in London, he developed an integrated fMRI application for Marconi Medical Systems. He then spent 5 years with GlaxoSmithKline in Verona, Italy, applying functional imaging to drug discovery in Psychiatry. Since joining Eli Lilly and Company in 2007, he has continued to pursue the validation and application of quantitative and functional imaging methods as biomarkers in drug development.

Lino Becerra earned his Bachelor of Science degree in Physics at the Universidad Nacional de Ingenieria, Lima, Peru and his PhD degree in Physics from the University of Illinois at Urbana-Champaign. He did his post-doctoral work at the F. B. National Magnet Lab at MIT and the Department of Radiology at Massachusetts General Hospital. He is an Associate Professor at Harvard Medical School, he has co-appointments in the Departments of Psychiatry at McLean Hospital and Massachusetts General Hospital (MGH), and Radiology at MGH and Children Hospital Boston. He is the Director of the Imaging and Analysis Group at the Brain Imaging Center, McLean Hospital and Co-Director of Pain Imaging and Analgesics Neuroscience Group (P.A.I.N. Group) at the same institution. He has published over 80 articles, reviews, book chapters and co-edited a book on Imaging in CNS Drug Development.

David Borsook is a Neurologist and Neurobiologist by training. Following a Pain Fellowship at the Massachusetts General Hospital, Harvard Medical School. He later directed the Pain Clinic at the hospital from 1994–1999. He has participated in a number of national and international pain initiatives including a WHO program in China. Since 2000 he has led a pain research group utilizing functional imaging to define brain metrics of chronic pain and analgesics. He was a co-founder of a company Descartes Therapeutics that was started to use imaging in drug development. Currently he directs the Center for Pain and the Brain at Harvard Medical School where imaging is being used to define biomarkers for chronic pain and analgesia in children and adults. He has published widely in the field including over 110 articles in peer-reviewed journals and co-edited a book on Imaging in CNS Drug Development.

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2011.03.011

and associated quality assurance/quality control (QA/QC) procedures are available for these methods and efforts at standardization have been undertaken [2,3]. The concept of an 'imaging charter' was recently introduced, in particular for trials within the purview of regulatory agencies such as the US Food and Drug Administration (FDA), to provide a means of documenting aspects specific to imaging procedures (especially the image 'reads' or analyses) along with a detailed procedures manual provided to the participating sites [1]. Dynamic and/or functional imaging MR approaches are also becoming more commonly used within an industrial setting and guidelines for their consistent use in this context have been developed. For example, consensus recommendations and open questions relating to the use of dynamic contrast-enhanced (DCE)-MRI specific to the evaluation of novel therapeutics have been developed and refined over the past 10 years [4,5]. The recent Critical Path Initiative of the FDA (http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm) provides a mechanism to establish the 'qualification' of biomarkers as fit for purpose in a particular context of use [6,7]. However this exercise can only proceed following the assessment of the performance characteristics, progress on method standardization and the availability of comparable data sets supporting the case. Functional MRI (fMRI) is not yet at this stage and, as mentioned above, the flexibility of the technique is such that this is likely to be paradigm (and possibly also population) specific. Nevertheless, the ongoing investigation of fMRI (and other emerging quantitative imaging methods) in the context of drug development relies on tight control of the experimental process.

In comparison to more routine imaging studies in which the acquisition can be performed by a technologist or radiographer familiarized with the protocol, fMRI experiments are more complex and a vast variety of paradigms and their implementation (both design and hardware) are in use by the imaging community. The flexibility of fMRI as a technique is such that any consensus on the specifics of the experiment is likely to be applicable only to a very precise paradigm, its implementation and the context of its application. To this end, efforts toward the standardization and characterization of robust, clearly defined paradigms and implementations, addressing specific disease areas or reliably activating specific brain regions, are of great utility. However, such efforts would require broad input from the field and would most sensibly follow indisputable evidence regarding the utility, reproducibility and strong evidence of the cognitive underpinnings of specific paradigms. Nevertheless, at a more general level than specific paradigms, as fMRI becomes more routinely applied in the drug development environment in the process of building the scientific case for its application, there is also an open need for best practice guidelines regarding site capabilities, study oversight, analysis and reporting at an appropriate level of detail so as to facilitate the implementation and maximize confidence in the conclusions drawn from the data.

The nature of the fMRI experiment is such that several factors can impact final data quality and reliability: (i) blood oxygen-level dependent (BOLD) fMRI acquisitions are time series, typically lasting approximately 5–10 min (although possibly longer), in which the signal changes of interest are small temporal variations (approximately 1–2% or less) often not much greater than background noise level; (ii) protocols often involve more than one fMRI scan in the imaging session; and (iii) the fMRI scanning session involves the synchronization, control, monitoring and collection of additional data streams (e.g. paradigm and/or stimulus presentation, subject feedback and physiological recordings) that are controlled and recorded by computers independent of the MR scanner system. These non-imaging data are crucial for data analysis and interpretation [8–10]; they must be synchronized with the imaging data and recorded and available for analysis in such a way that they can be accurately associated with the relevant fMRI scans. Moreover, (iv) the analysis of fMRI data typically involves format conversions and a complex sequence of image manipulations, for each of which different algorithms, software implementation and parameter choices are available. The addition of a pharmacological agent into the procedure brings additional constraints and requirements, especially for a new chemical entity (NCE). These can include the involvement of contract research organizations (CROs) or sponsor company representatives for safety oversight and clinical monitoring responsibility, pharmacy support for compound preparation or receipt of a compound prepared off-site, blood sampling and safety oversight in the MR facility, plasma sample analysis, third party QC oversight and study documentation and reporting requirements. In addition, the pharmacokinetic (PK) characteristics of the compound will impact the protocol design.

Although organizations such as the American College of Radiology (ACR) publish guidelines (http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/tech-standards-mp.aspx) for ensuring appropriate scanner performance for clinical use, requirements are more stringent for quantitative imaging in which the data are analyzed to generate computed, numeric endpoints rather than being interpreted or 'read' by a radiologist. A substantial body of work exists on QA procedures for (f)MRI, concentrating on procedures to monitor scanner performance regularly (e.g. parameters such as stability of the magnet and electronics, signal:noise ratio (SNR), geometric distortion, slice profiles, ghosting, temporal stability/drift [11–14], and contrast:noise ratio [15]) as well as routines to detect artifacts in fMRI time series data (e.g. those caused by subject motion) [13]. MR scanner manufacturers provide QA routines as a recommended part of regular scanner performance monitoring to ensure that specifications are met. The function Biomedical Informatics Research Network (fBIRN) consortium (http://www.nbirn.net/research/function/index.shtm) is examining many of these issues in the explicit context of multi-center fMRI, applied in particular to schizophrenia [12,16–18]. There are also recommendations for the reporting of fMRI study design and methods [19,20].

In a previous article [21], we examined pharmaceutical company and imaging site processes that would typically circumscribe industry-sponsored fMRI studies and proposed guidelines to help frame site qualification. Here, we examine the imaging procedures in greater detail and suggest recommendations relating to the effective implementation of protocol-specific procedures. Appropriate QA and QC procedures around data acquisition and analysis are an important part of good imaging practice (GIP) designed to maximize confidence in the results of the study. These guidelines are more generally applicable than are specific QA protocols or implementation details, of which many examples have been published [12,13,15] and can be considered appropriate for a given study, especially if already in practice at the site(s) involved.

Optimizing processes for fMRI protocol implementation

Logistics and communication

When designing and refining a specific protocol, logistic considerations need to be considered at an early stage to reconcile requirements and expectations from the different parties involved. This can include issues of subject recruitment and screening, clinical monitoring or inpatient housing of study subjects, timing of subject visits and the full schedule of events and compound preparation, delivery or encapsulation and administration. The operating hours of the fMRI facility and other requisite staff and functions (e.g. research pharmacy) need to be reconciled with desired timings of compound dosing and data acquisition informed by the pharmacokinetic profile of the compound. Effective communication between all parties is paramount; schedules and mechanisms for regular updates and resolution of issues arising during the trial should be agreed up front.

The site qualification exercise discussed in below is designed to ensure the presence of robust procedures at the site level that will provide a framework for the implementation of a specific imaging protocol (Fig. 1). We recommend that detailed consideration of the particular protocol at hand be fully defined in the study Preparation phase. This enables a clear definition of workable QA/QC procedures, reporting metrics and roles and responsibilities within the study team. Including a full specification of the data analysis pipeline at this point enables the full data flow to be checked and will help ensure efficient data analysis. High-level recommendations for planning, study execution and analysis are summarized in Box 1 and elaborated further in the remainder of this section.

Auditing the protocol: risk → mitigation → decision rule

A pragmatic, risk-based approach to establishing protocol-specific procedures of an fMRI trial to an appropriate level of detail is to walk through the proposed schedule of events, explicitly evaluate the risks associated with each step and consider practical measures to mitigate each one. These could include both QA measures to prevent or minimize the chance of the issue occurring and also QC mechanisms to ensure that any occurrence is detected and reported in a timely fashion. In the latter case, a decision rule

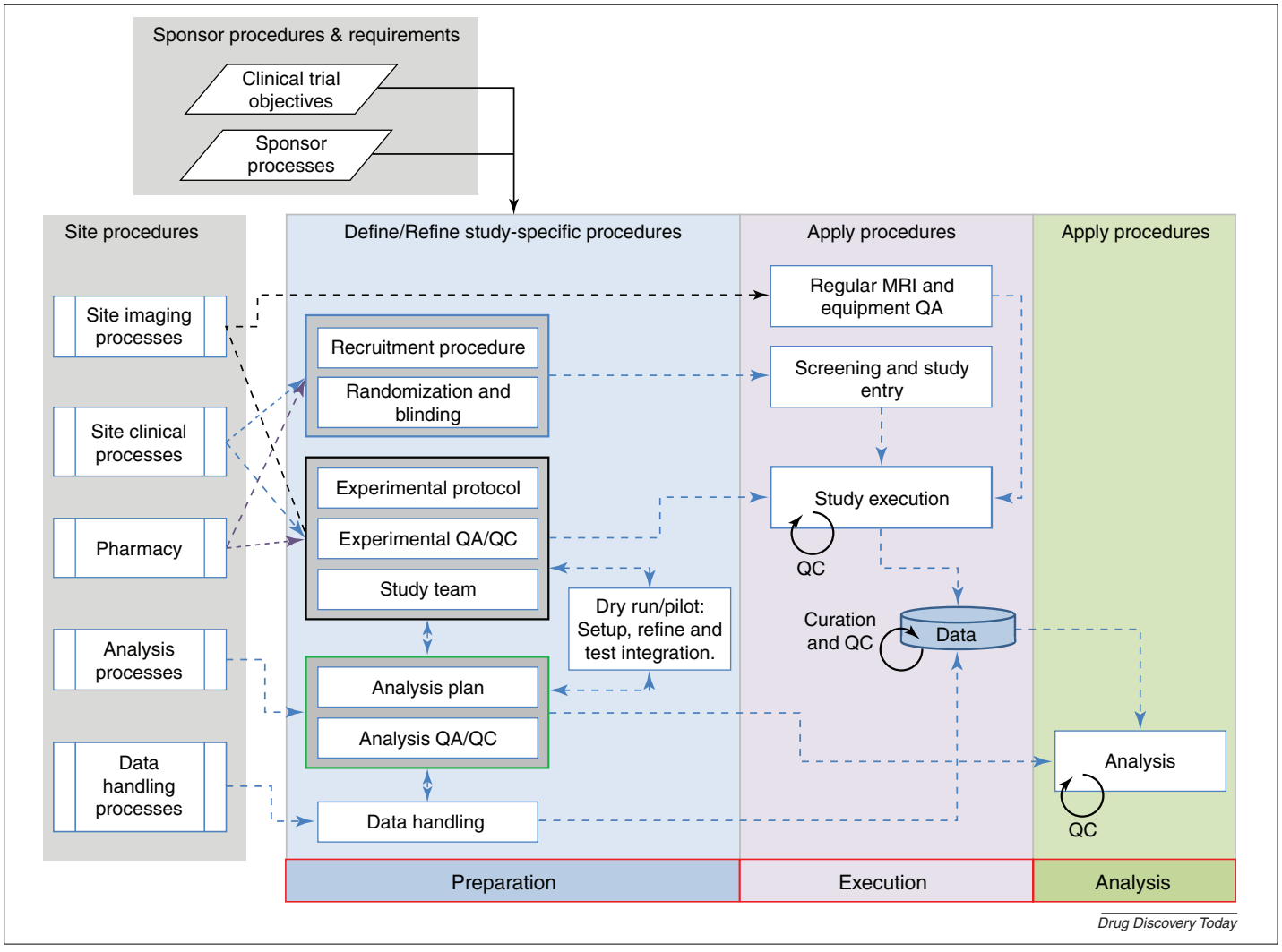

FIGURE 1

The study process. A breakdown of fMRI-related activities involved in the preparation, execution and analysis phases of a pharmaceutical fMRI study.

BOX 1

Summary recommendations for optimizing protocol implementation processes

Logistics and communication

- Reconcile logistical constraints between all parties in Preparation phase.
- Establish clear communication schedules and mechanisms for regular progress updates and resolution of issues arising.

Protocol implementation and risk assessment

- For any particular study, examine the fMRI process in detail in the Preparation phase with the aim of identifying potential risks to data quality (i.e. what could go wrong at each step).
- Develop mitigation strategies for each such risk identified and clear decision rules to guide consistent action for each eventuality.
- Perform a pilot run through the acquisition process to ensure team familiarity with all parts of the procedure and also to refine QA/QC checks and logging.

Imaging and acquisition procedures

- Establish scan-day and acquisition-related checks for all equipment (MR system and ancillary equipment and devices).
- Establish predefined subject- and acquisition-driven rescan/exclusion criteria.
- Assign well-defined roles and responsibilities to the study team and identify backup personnel in case of unavailability. Specify clear decision rules as to whether a scan proceeds based on the availability (or not) of study team members.
- Establish a mechanism and responsibility for documenting and regularly reporting QC results and exclusion decisions.
- Carefully evaluate potential risks of the study drug(s) at hand and develop safety strategies to ensure that an appropriate response will be available that is compatible with the magnet environment and fMRI equipment.

Data collection and handling

- Consider how ancillary data will need to be combined with the imaging data at the analysis stage, and implement storage locations and data recording parameters to facilitate this.
- Establish a consistent naming convention for all data files (imaging and ancillary).
- Establish data curation and routing responsibility and procedures to prepare data for analysis pipeline. Ensure procedures are in place to support required data transfers.

Pharmacy and pharmacology

- Reconcile compound preparation and delivery with compound administration and scanning schedule.
- Establish consistent time of key functional series acquisition with respect to compound dosing and pharmacokinetics.
- Include appropriate blood sampling in the scan-day protocol to enable peripheral compound exposure to be assayed.

Analysis

- Implement and document the image and statistical analysis pipelines in full detail alongside the acquisition protocol to ensure close alignment with study objectives and precise definition of endpoints. Where possible, include scripts and/or commands, masks [for volume of interest (VOI) definition] and design matrices to be used.
- Test the analysis pipeline before the start of data acquisition (e.g. on Pilot data) to ensure that all data expected for analysis will be acquired and in an appropriate format.
- Specify clearly how the primary endpoints will be computed and how the hypotheses of the protocol will be tested, including a description of how the anatomical localization of the effect of the compound will be determined.
- Specify QC steps and decision points in the analysis pipeline as precisely as possible, along with a mechanism for documenting and/or reporting them. Use objective measures from the data for QC wherever possible.
- Assign specific responsibilities for data analysis and QC.

Deviations (off-protocol imaging)

- Specify clear go/no-go criteria and how scanning sessions, individual scan series and subjects will be handled, especially as regards repeating a task or scanning session (e.g. exclusion or reschedule a replacement scan) in case of an imaging session deviating from the prescribed protocol.

based on clearly specified, preferably objective, criteria should be devised to determine how to proceed, and a mechanism for documenting the action incorporated in the study protocol. In general, this exercise would be valuable to perform for each new protocol; if similar protocols have been previously run at the site(s), then the exercise might be more one of refinement.

A key aim of this process should be to devise procedures that are workable and to reduce the complexity of the data integrity risk to relatively simple checklists and decision trees that are practical to implement and commensurate with effective communication of study progress. Once implemented, there is considerable value in robust and meaningful experimental oversight and well-defined operating and analysis pipelines. A formal dry run or full Pilot phase might be useful to 'road test' all components of the protocol and enable refinements of the acquisition and analysis procedures, including QA/QC mechanisms (Fig. 2) [22], especially if the protocol has not been previously executed in the specific form to be used in the drug study.

Imaging and acquisition procedures

MRI

Details of the MRI sequences and parameter values will depend upon the aims of the study and should be fully documented during the Preparation phase. If the protocol involves fMRI sessions comprising several functional scans, the order of these should be explicitly considered (e.g. constant versus randomized ordering). Although routine scanner QA should be in place (see below), there is still a risk of compromised image acquisition during the scanning session (e.g. owing to problems with the receive coil, control electronics or image reconstruction). If the scanner is also used for other research studies, there is an increased risk of some part of the transmit–receive–reconstruct chain requiring attention. In some cases, the issue can be remedied, but would probably

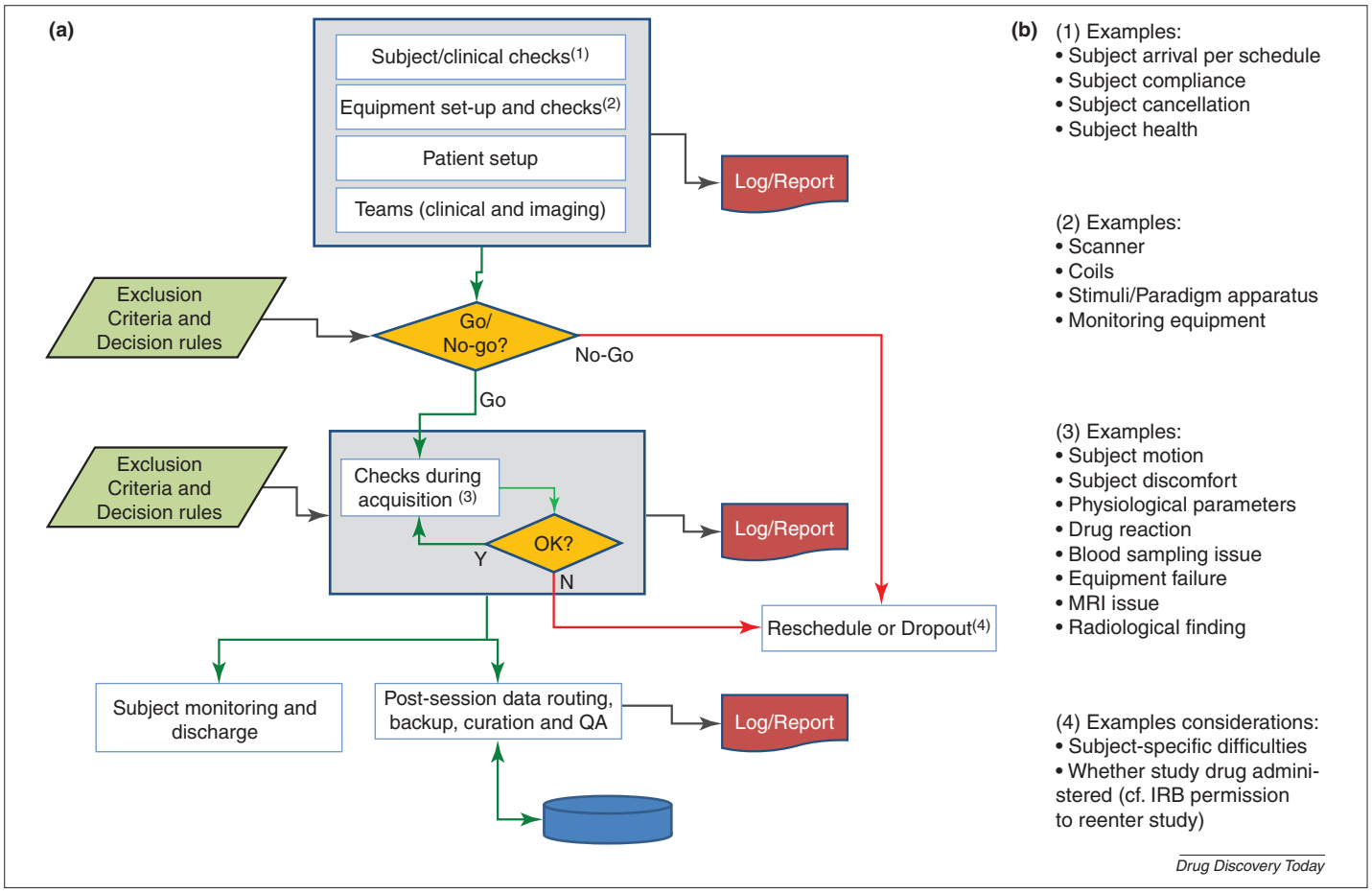

FIGURE 2

Scan-day implementation. **(a)** Flow diagram summarizing the fMRI scan-day process and points at which predefined decision rules and records of QC diagnostics could be applied. **(b)** Examples of what procedures might be in place to record and check implementation.

require a physicist or engineer to be present (see below) and the study procedures should specify a time limit beyond which the session does not proceed and the issue is resolved off-line. To ensure any such issues with the MR system are detected, it can thus be useful to have an explicit QC check either before or early during the MRI imaging protocol (e.g. following acquisition of the anatomical images). This could comprise a simple image SNR measurement, for example; or a qualitative verification of anatomical coverage and image quality (minimal distortion, signal dropout, among others) might be considered sufficient. A mechanism for regular reporting of the scan-day QC results and any actions or exclusion decisions should be implemented.

Many scanners offer the ability to display images in real time as they are being acquired; this adds another level of monitoring as sudden degradation of image quality can be detected early and potentially remedied. For example, if the head coil was not installed properly and suddenly loses some of the contacts, the effects on the images would be immediately visible, without the need to wait for a manual check following scan completion. Another advantage of this facility is that excessive motion can be detected while the scan is running; the subject can be reminded to stay still for the duration of the scan, and the scan restarted. This can reduce data attrition in the Analysis phase owing to motion-related exclusions.

Subject monitoring

The physiological parameters to be measured in the magnet should be clearly defined as deemed necessary or appropriate for the study. Some, such as end-tidal CO_2 ($ETCO_2$), heart rate and respiration rate can be measured while the subject is in the scanner and should be considered as part of the data acquisition protocol as a means of determining possible confounding contributions to the fMRI signal changes [8].

From a safety perspective, required subject monitoring will depend, to a large extent, on the compound(s) being tested. NCEs in the early phases of development will have stringent clinical monitoring requirements, and the specific side effects observed in preceding safety studies will inform on specific risks and monitoring requirements while the subject is in the scanner environment (e.g. having a nurse or physician present in the scan room while the subject is in the scanner [23]). In particular, compounds that might compromise the subject's mental or motor abilities might preclude the subject's ability to signal an alarm. For marketed compounds, the side effect profile will be better established and should likewise inform relevant monitoring procedures. The above-mentioned physiological parameters can also provide useful real-time feedback on the subject's physiological status and procedures to monitor these signals for deviation from normal limits should be considered. Other physiological signs, such as

impending vomiting, are subjective and not measurable. Vomiting in the prone position is dangerous, and in this eventuality, the subject needs to be removed from the scanner. Somnolence within the scanner might not be easily detected, but should be closely monitored. If this is a known side effect of the compound under study, this can be accommodated by explicit instructions to the subject, shorter scan runs enabling more frequent communication and more vigilant monitoring. Other medical conditions and issues related to monitoring that may be crucial to patient safety have recently been reviewed [23]. Subjects should be clearly instructed ahead of time about what to do if they are not feeling well and regular questions related to how they are feeling can be incorporated into the scan-day procedures. Subject monitoring requirements continue after the end of the scan and the logistics of this need to be reconciled if the subjects are housed or monitored at facilities separate from the scanning site (e.g. at a clinical CRO).

Another important aspect of subject monitoring is compliance with the protocol: is the subject following the instructions? Specific written instructions can be included in the visual paradigm presentation and subjects can be read a defined 'instruction' text throughout the protocol. Subject feedback (e.g. rating dials or button presses) enables performance to be monitored and acceptable compliance with the protocol should be defined as part of the scan-day QA/QC procedures. This should also include decision rules and exclusion criteria if compliance is compromised (see below).

Ancillary data

The non-imaging, ancillary electronic data to be collected in the trial should be clearly defined, along with their intended purpose. This could typically include the presentation and/or recording of the fMRI paradigm or applied stimulus, the recording of subject feedback data (e.g. button/joystick responses) and the recording of physiological data. Records of the applied paradigm and subject responses can be crucial for data analysis if they determine regressors for temporal modeling. Records of physiological parameters, such as heart rate, respiration rate, blood pressure or $ETCO_2$, can be evaluated independently to flag any potential confounding drug effects on physiology [8], or incorporated explicitly into the analysis pipeline [9,10].

The ancillary data streams are usually controlled and recorded by one or more computers or devices independent of the MRI scanner itself. Scan-day QC checks and exclusion criteria based on the ancillary data should be considered during protocol development. It is also useful to verify how these ancillary data streams are synchronized with the MRI data acquisition, including how the start of the fMRI acquisition coincides with the start of the paradigm delivery and/or clock synchronization between the two machines to ensure consistent time stamps.

Study team: roles and responsibilities

Successful scan-day implementation of the protocol involves a smooth interplay between many component systems, from patient arrival to drug challenge to imaging and subject discharge (Fig. 2). The specifics of this process should be walked through in detail during the Preparation phase for each protocol; this is greatly facilitated if based on a framework of site standard operating procedures (SOPs) (see below). These can provide templates for the development of protocol-specific checklists, equipment/data diagnostic readouts and reporting procedures. Once a protocol has been developed, it is advisable to conduct a 'dummy run' of the full protocol on a volunteer subject, including data acquisition on the scanner and data analysis (but omitting drug dosing). This enables the details of the scan-day protocol to be tested in action and provides an opportunity for them to be refined before commencement of the study proper.

The importance of teamwork in the clinical and research arenas has been a focus to improve the quality of processes [24–26]. Whereas more routine imaging acquisition can be performed by technologists and/or radiographers, fMRI is substantially more complex, requires parallel implementation of different activities and, thus, relies far more on a team of people with complementary expertise. Consistent and accurate application of all parts of the protocol is most assured by the involvement of a constant set of personnel trained on the specific protocol and experienced in relevant fMRI domains. Typically, several individuals must be involved during acquisition and so well-defined roles and responsibilities with respect to specific aspects of the study are advisable. These include overall study coordination, preparation of the subject (including any required medical intervention, e.g. cannulation), dosing, data acquisition on the scanner, control of paradigm presentation, QA/QC (e.g. defined responsibility for checklist sections), blood sampling and medical responsibility for safety monitoring. In this latter respect, it is advisable to have a dedicated clinician who is solely responsible for subject safety and who has no competing responsibilities.

Another important consideration of the study team is that not only it is important for the smooth execution of the protocol, but it also provides constant interaction with participant; the team is part of the study environment. To deal with the case where individuals might not be available on certain scanning sessions (e.g. owing to illness or vacation), suitably trained backup personnel should be defined.

Data collection and handling

It is useful to consider the collection and handling of all data acquired in each session in the context of QC and analysis operations that will be performed on them. An integrated view is crucial, given the probable need to combine data from different sources (e.g. imaging time series and subject response files). The details of this (e.g. which software will be needed to access them for intermediate processing) are important to consider to ensure that the data flow enables these operations to be performed efficiently. Hence, the format of the data files that will be generated during the Execution phase and any conversion routines should be specified and tested well in advance of the Analysis phase.

This should include consideration of how the ancillary data will be combined with the imaging data (see below). Clarity on this aspect is especially crucial if analysis and/or data QC is to be performed by the third party. Procedures can be developed to ensure that all data are available and appropriately organized for the subsequent analysis. Consistent file naming conventions across the different data types and a structure for data organization that facilitates QC and analysis are useful mechanisms.

The study case report form (CRF) provides a mechanism for collecting high-level data and incorporating checks relevant to the fMRI examination [i.e. in addition to other drug trial-relevant

information, such as medical examination(s), vital signs, concomitant medications, drug screen results, time and details of compound dosing and pharmacokinetic collection details]. Given that the primary data are electronic (image or ancillary text files), the CRF is useful for capturing logistical aspects of the scanning procedure. These could include items such as the acquisition start time for each scan in the session, a reminder for the subject to empty their bladder and/or checks that any additional equipment required for the examination has been set up (e.g. infusion pump or physiological monitoring equipment). Given that all CRF fields will be entered into the study database, the additional imaging-related fields should be selected pragmatically, taking into account their impact on the data acquisition process and the information required in the final study database.

Pharmacy and pharmacology

Drug preparation

Logistics of compound formulation and delivery (if prepared off-site), stability of the compound, operating hours of the pharmacy, pharmacokinetic constraints in the protocol and the available imaging times all need to be reconciled. As with the imaging staff, consistency in personnel is advisable to ensure familiarity with the protocol and to minimize variability. The stability of the compound itself will dictate how preparation and dispensing can be arranged. Depending on stability, some compounds can be formulated ahead of time and stored overnight, whereas others need to be prepared fresh on the day of administration. This might impact the scheduling of dosing and scanning relative to pharmacy working schedules.

Dosing

For acute dosing paradigms, to minimize additional variability and enable precise study design, it is desirable to minimize variability in the timing of the fMRI scans relative to the drug administration. The acquisition paradigm should be defined relative to the known pharmacokinetics of the drug. This might be accomplished by incorporating a set time relative to dosing (e.g. T_{max}) in the protocol at which a specific functional acquisition is started. Time ahead of that trigger point should then be built into the protocol, allowing for the subject to be set up in the scanner, for any initial imaging preparation (e.g. anatomical imaging or shimming) to occur and some 'buffer' time to allow for unforeseen delays. The consistency of this timing can then be tracked as part of the study QC and a range of acceptable scan start timings can be established as 'go' criteria for further data acquisition in that scanning session. In any case, start times of all functional scans should be recorded for comparison with compound PK.

If the protocol uses an intravenous infusion, QA/QC procedures should be developed for the infusion pump and a protocol specified for calculating the dose of the compound to be delivered to each subject. The input required for this calculation (e.g. subject weight), how it is entered and where it is obtained should be clearly specified.

Pharmacokinetic sampling

A key aspect of integrating fMRI into drug studies is to combine the imaging results (a pharmacodynamic (PD) measure) with plasma measurements of drug exposure (PK). A simple single drug assay based on a post-scan blood sample is probably the minimal procedure that should be considered to confirm that drug was present. For acute dosing studies and with novel drug candidates especially, more frequent sampling before and during the scanning session proper enables more usual PK parameters to be calculated for correlation with imaging measures and possibly more sophisticated PK/PD modeling. This requires the definition of blood withdrawal procedures for collecting blood samples while the subject is in the magnet and for sample storage and analysis. The timing of blood draws needs to be built in to the imaging protocol and recorded, as these should occur between scans. For chronic or subchronic dosing studies, blood sampling can probably be scheduled outside the scanner.

Often, it is more cost effective to have plasma samples analyzed in batches; this should be considered in the Preparation phase within the context of the trial to ensure these data are available when required (e.g. for an interim analysis). As with the MRI and ancillary fMRI data discussed above, a standardized method of labeling can minimize errors in linking PK data to the imaging results.

Analysis

An important aspect of the protocol audit exercise is to include explicit and detailed consideration of the image analysis pipeline during the Preparation phase to ensure all required information will be available in the appropriate form for efficient and streamlined analysis. Small details that could otherwise complicate the analysis phase can be resolved in this way and help ensure consistent treatment of data. Guidelines for thorough reporting of fMRI results have been published [19], and these can also provide an excellent framework to help shape the analysis plan. Moreover, if paradigm or response data recorded by a secondary computer are to be combined with the imaging data for analysis, it is useful to think through how these data files will be manipulated in the analysis pipeline. Details such as relative temporal sampling times of ancillary and imaging data (e.g. re-binning of physiological traces to match image data or conversion of event-based records to time vectors), synchronization of the two computers and consistent truncation of the data time series (e.g. truncation of initial time points to match those removed from image series for signal intensity stabilization) should be considered.

Primary versus secondary/exploratory endpoints

An important consideration is the conceptual separation between the primary, secondary and exploratory outcomes of the trial. These will shape how the acquisition protocols are designed, the computation of the primary and secondary imaging endpoints and the statistical tests to be performed on those endpoints to test the study hypotheses. Exploratory analyses of the data and pursuit of unexpected findings remain an important part of the science within this context, but must simply be considered *post hoc* and conceptually separate from the specification and hypothesis testing of primary endpoints. In the case of trade-offs being required in the Preparation phase, priority is given to optimal delivery of the prespecified endpoints.

Detailed analysis, plan and pipeline

In prospective clinical trials, it is crucial that hypotheses are explicitly stated *a priori* (http://www.ema.europa.eu/pdfs/human/ich/036396en.pdf). This has implications for the planning

of fMRI data analysis, which involves a sequence of many image manipulations, for most of which there are several widely used variants in terms of algorithm, parameter values and software implementation. Moreover, the order in which these steps are applied might differ slightly depending on the software environment and the analysis philosophy used.

Our recommendation is that all steps, hardware, software, functions/modules, version, parameter values and models used should be clearly specified *a priori* in an image analysis plan or fMRI review charter. Sufficient detail should be captured to enable a full re-analysis by an independent investigator if required, given the same data and the same software, to generate identical numeric endpoints.

Another key aspect to be captured is how the operations will be performed (e.g. subject by subject in turn or by batch processing) and the extent to which the operations rely on manual interaction with a graphical interface or are automated by means of batch scripts or software that permits and/or requires the specification of configuration files. The latter option minimizes repetitive user input and, hence, is less prone to human error; additionally, the scripts and/or configuration files provide both a record of the operations performed and the means to easily rerun or modify the analysis, should the need arise, or for *post hoc* data mining or exploratory analyses. Consider ease of scripting (e.g. constant number of characters in filename fields) and capturing processing steps incrementally in the filename to provide a simple record of the operations already performed on a given data file.

QC and data-driven exclusions

For all the potential variants in the details, a given fMRI analysis pipeline comprises a set sequence of operations on the data and lends itself to clear specification of QC checkpoints. These can include fully automated quantitative summaries based on the data (e.g. head motion excursions calculated by the motion correction algorithm). Any QC-driven exclusion or corrective action decisions should occur before the temporal modeling steps [i.e. blinded to the calculation of the final imaging endpoint(s)]. Figure 3 illustrates graphically an example fMRI spatial 'preprocessing' pipeline (i.e. not including temporal modeling and statistical inference) showing a defined sequence of operations, information logged from the imaging data and the integration of checks and decision points. However, the complexity of the fMRI experiment and its analysis is such that expert qualitative assessment of difficult-to-quantify aspects (e.g. brain extraction or spatial normalization 'quality') might also remain part of the QC process. The same recommendation applies to both, namely that the decision rules are specified *a priori* and the results of the checks, decisions (e.g. exclusions) and/or any corrective action taken are recorded and reported.

Anatomically specific hypotheses

In most fMRI studies outside the drug development context, the primary analysis is performed at the image level, generating statistical parametric maps identifying the anatomical distribution of voxels exceeding the statistical threshold. Most standard statistical tests (*t*-tests, ANOVA, ANCOVA, among others) have been implemented in fMRI software suites that enable these to be performed efficiently in a massively univariate sense at each of the tens of thousands of voxels comprising the image data. Many of these algorithms also take into account local spatial correlations (e.g. Gaussian Random Field Theory) to determine appropriate multiplicity correction. This whole-brain, image-based inference tests the null hypothesis globally across the whole brain; however, in pharmaceutical applications of fMRI, it is often of interest to test hypotheses that are more anatomically specific (i.e. regarding drug action on the response in a specific brain structure or set of structures [27]).

For such anatomically or circuit-specific hypotheses to be tested rigorously, the way in which a given structure is specified and how the response is related to that location should also be prespecified. An example might be the computation of a summary value from all voxels within a predefined region from each subject. Most simply, this could comprise the average BOLD signal change (amplitude corresponding to the contrast of interest) across all voxels within a specified anatomical structure. Other prespecified summary statistics from within a given region could also be extracted [28]. This can be useful in the case of patient studies, where differences between individuals might reasonably be expected to alter the precise brain location at which the effect of interest is localized (e.g. in a pain population where different subjects might have pain originating in different parts of the body). When the location of the effect at the group level is expected to be relatively focal but cannot be prespecified to a specific coordinate, an alternative approach is to perform a voxel-wise analysis restricted only to a predefined region of the brain, often referred to as 'small volume correction'; this does not generate a single numeric value from each subject but enables a regional hypothesis to be applied with a reduced multiple comparison penalty, at the expense of ignoring effects elsewhere in the brain [29]. Whatever the approach, the coordinate, atlas or mask to be used to define the anatomical structure and/or location of interest should be prespecified (including how the mask itself is thresholded if a probabilistic atlas is used) along with how the individual voxel responses will be combined. In other words, all parts of the hypothesis, including anatomical information, should be independent of the data being tested [30,31].

The approach of calculating the average voxel response within a predefined mask is particularly amenable to applications in drug development as it provides numeric endpoints that can be easily transferred to the sponsor and analyzed by an assigned statistician compatible with standard trial analysis processes, greatly facilitating integration of fMRI data with other trial information. It also enables a logistical separation of the image analysis (derivation of endpoints) and statistical analysis (inference on endpoints).

Blinding

Another important consideration is blinding of the data for the personnel performing the analysis. Those involved in data acquisition might be effectively unblinded owing to manifest effects of the compounds being studied (e.g. behavioral or physiological effects) and, in many cases, might also be involved in data analysis. Although the analysis steps should be predefined (and ideally deterministic), it is advisable to ensure that the analysis personnel cannot link a given data set to a subject or treatment condition. This could be achieved by using different personnel for acquisition and analysis, or via a naming convention and analysis scheduling that prevents inadvertent unblinding.

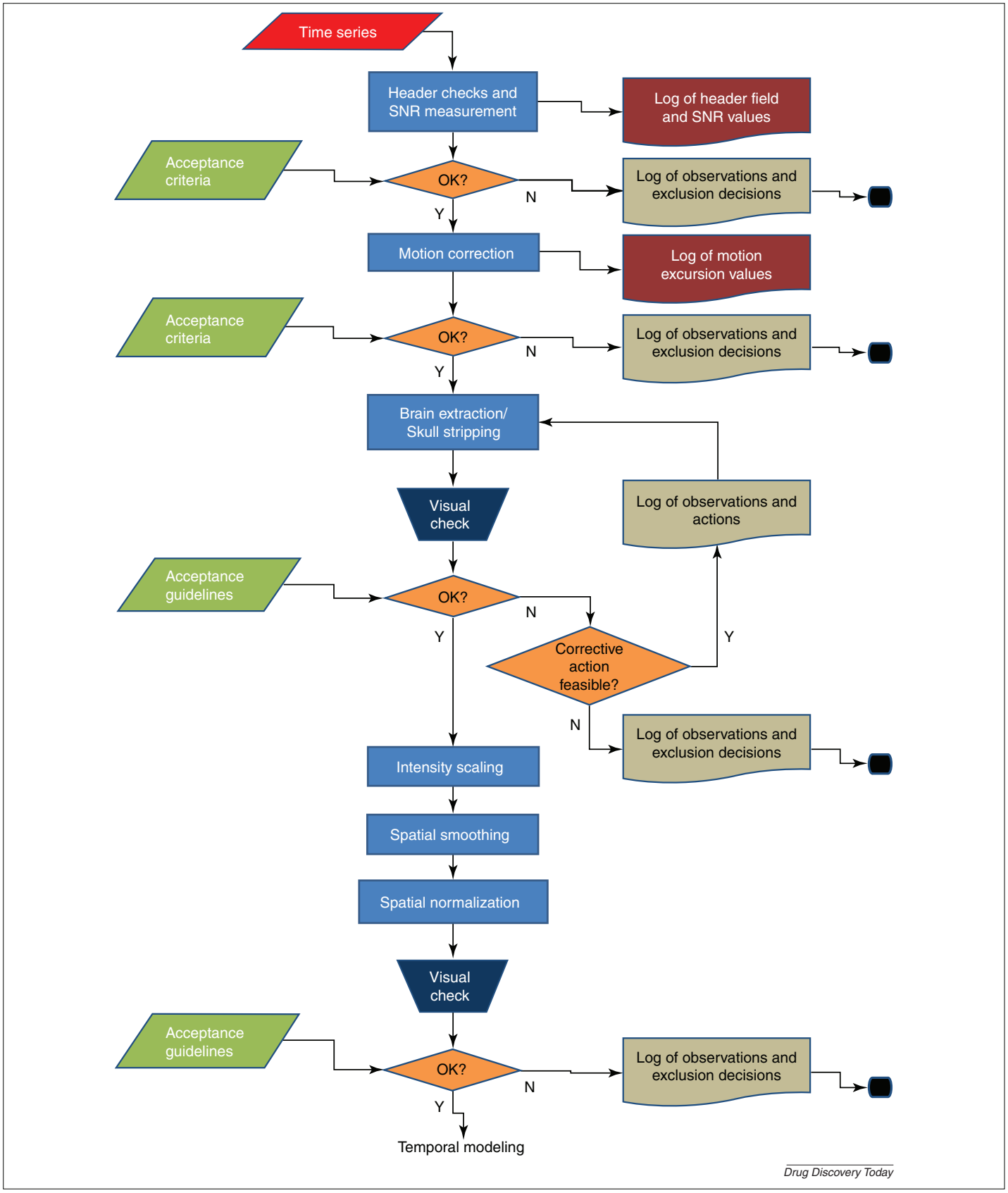

FIGURE 3

Preprocessing and QC pipeline. An example pipeline for the spatial preprocessing of fMRI time series data, illustrating both automated and manual QC checks and parameter logs. Data-driven exclusions should be based on predefined, preferably objective, criteria and applied before computation of final endpoints via temporal modeling.

Deviations (off-protocol imaging)

Any imaging study must contain defined procedures to deal with off-protocol imaging. Although standard trial parameters, such as the window of scan timing relative to drug administration, can be captured in the study protocol, the imaging procedures themselves bring a further level of detail related to the scan acquisition parameters. Although these should be specified before the study commences (and, on modern scanners can be saved for convenient loading when required), incorrect acquisition parameter values might be mistakenly applied or a modification of the parameter values might be required during the study. Typically, these are cases in which usable data are acquired, but which deviate in some way from the desired specifications. Along with screening (trial entry and scan-day checks) and data quality (e.g. SNR or motion) criteria, this represents the third category of monitoring criteria that needs to be carefully considered in the study Preparation phase. Deviations at this level of detail might not constitute a protocol violation *per se*, but at the very least should be logged and discussed with the sponsor.

It is also probable that the fMRI acquisition protocol will include more than one functional scan; what to do, for example, when one fMRI series is in some way corrupt (e.g. the subject does not comply with the paradigm instructions), but the others acquired in the same session are acceptable? Or when a subject does not complete all scanning sessions in a within-subject design? Enrollment of a replacement subject might be necessary. These eventualities need careful consideration in terms of their impact on timelines, planned analyses and expected outcomes, in particular those that involve combining data from different functional conditions. For cognitive paradigms in particular, the presence of learning effects might be a substantial confound if tasks are repeated; ideally, data should be available to demonstrate the lack of learning effects on the fMRI signal if task (or session) repeats are to be admitted in a given study. *A priori* decision trees can provide prespecified rules to deal with such situations.

Clear go/no-go criteria

Although much preparation goes into the process of having a 'perfect run' for each imaging session, go/no-go criteria should be defined for each protocol being implemented to determine whether the scan proceeds and/or if the subject or their data from a particular session or scan are excluded. For application to pharmaceutical trials, there is the expectation that exclusion criteria be clearly specified *a priori*. With respect to issues arising within the scan-day process, these can be classified into the following domains:

(i) **Subject issues**: these include (a) time of arrival for the study; (b) appropriate pre-imaging compliance (e.g. food restrictions, medication use, recreational drug use or other exclusion criteria defined in the protocol); (c) other illness (e.g. allergy, or change in medical condition); (d) unexpected radiological findings (e.g. cyst or tumor present in the brain); (e) non-compliance with protocol procedure; (f) drug side effect; (g) abnormal movement; (h) halting the scan (e.g. need to use the bathroom) or other processes that might delay the temporal specifications of imaging relative to drug intake. Depending on the stringency of predetermined data acquisition, it might be considered acceptable for certain experimental measures to be omitted; for example, missed blood samples for plasma assays. These should be captured as fully as possible in the Preparation phase in the definition of acceptable off-protocol acquisition. Another eventuality to be defined in the protocol and agreed with the ethical review board is whether (and how many times) a subject can return for a repeat of the imaging session if they have already been dosed and the scanning session needs to be aborted. If the subject needs to be unblinded owing to an adverse event, they would probably be excluded.

(ii) **Team issues**: if a full team (originals or replacements) is not present, the session should not proceed because the integrity of QA/QC, patient safety or simply the SOP implementation might be compromised.

(iii) **Magnet issues**: magnet operation should be in compliance with the site quality standards and any additional checks defined in the specific protocol (e.g. additional phantom scans on the imaging day). In addition, an acceptable time window for the resolution of any magnet or MRI equipment problems during imaging or on the scan day should be prespecified. If the issue cannot be resolved within this time, the imaging should not proceed.

(iv) **Equipment issues:** faulty operation of physiological monitoring equipment might be considered a no-go decision depending on the drug and on the satisfaction of the clinical team overseeing patient safety. For protocols where physiological measures are important (e.g. breathing and $ETCO_2$ for opioids), a no-go decision should be made for reasons of both safety and data QA. Similarly, equipment for psychophysical measures might fail to meet standards of operation before the study. Frequently, these can be fixed at the time of scanning but, as for MRI equipment, specific timeliness for problem resolution should be established during the Preparation phase to provide a clear decision rule for a go/no-go decision on data acquisition.

Discussion

Our aim here was to consider some of the complexities involved in the acquisition and analysis of fMRI data, and additional factors to take into account when a therapeutic drug candidate is part of the experiment. We have proposed a set of guidelines to help the robust design and execution of fMRI protocols. Whereas some might seem obvious to imaging specialists, and others to industry scientists, we feel that it is important to merge the exigencies of the technique and its industrial application to align understanding and expectations around fMRI studies. These recommendations can be considered as one part of GIP in pharmaceutical fMRI, designed to maximize confidence in the results obtained by providing clear documentation, appropriate QA/QC and meaningful checks of the study procedures, equipment and data. (Another important aspect of GIP, the site assessment process, was considered in [21].) This is important from an industry standpoint because many key details of the fMRI process do not fit naturally into standard clinical trial documentation templates. This gap can be addressed, in part, by capturing these details at an appropriate level in explicit procedure manuals, formal QC logs and detailed Image Review Charters/Analysis Plans, as is becoming standard for other quantitative imaging methods, especially those supported by specialist imaging CROs.

The complexities of fMRI can contribute a potential risk of incomplete data or for certain pieces of the data to be treated in an *ad hoc* fashion during analysis to compensate for an issue arising during data acquisition (i.e. non-uniform data handling as a result of deviations from on-protocol imaging). In experiments as complex as fMRI, rules governing exclusions must be balanced by the impact on the power of the study, which might be particularly acute in the case of small studies typical of fMRI. Specific and, where possible, objective data-driven decision rules and acceptance criteria should be defined and agreed upon along with a mechanism for logging and reporting the relevant details of the fMRI process and any exclusions or interventions. Data analysis is also expected to be fully specified before the Analysis phase and to be based on well-defined and well-characterized numeric endpoints, with a clear conceptual separation between *a priori* hypotheses being tested and *post hoc*, exploratory findings. Given the flexibility of fMRI and the wide variety of equipment and techniques used, it is difficult to prescribe a one-size-fits-all solution at a detailed level; rather, the precise details should be 'fit for purpose', appropriate to the study at hand and acceptable to the study team and all parties involved. It can be useful, if not essential, to test the resulting study 'algorithm' in a hands-on fashion in a pilot phase of the study allowing scope for refinement if necessary [22].

Furthermore, in many instances, the imaging site(s) will be chosen based on previous work and established procedures related to a particular specialization within neuroscience. In this case, if the protocol contains elements that map closely to procedures already well established at the site(s), the main burden might be on documentation (i.e. making the existing process visible). Nevertheless, a close examination of the protocol and potential data integrity risks is still likely to be useful and, in particular, the introduction of any novelty requires careful thought and planning. At the very least, it is likely that a novel pharmaceutical compound will be studied at the site and in the fMRI setting for the first time; pharmacology requirements based on the specific compound must always be carefully considered.

Given the complexity of fMRI and the inherent limitations of many parts of the process (e.g. residual intersubject differences in anatomy following warping into a standard coordinate space or a different impact on final data of isolated versus continual head motion during scanning) there is still an important role for QC decisions based on the informed judgment of a skilled practitioner. Automated processes and objective measures can help greatly but the final QC procedures must be pragmatic and best suited to the study at hand. Our key recommendations are that guidelines that are as precise as possible be established *a priori* for any qualitative QC and inclusion/exclusion decisions, that these decisions are documented and that they are made independently of knowledge of the final response value used for statistical inference. We did not explicitly consider aspects relating to multi-center studies, but our considerations apply equally to single- or multi-site trials. Clearly, in a multi-site context, factors relating to harmonization of procedures across the sites and potentially different scanner manufacturers become crucial and these have begun to be explicitly addressed [12,13,16,17,32,33]. Our emphasis here on bringing the fMRI process into line with standard industry practice is only one of many relevant issues regarding the use of fMRI to study central nervous system (CNS) drug effects [34]. Many other aspects are also important [35]; for example, how to deal with confounding effects on drugs on the vasculature directly or on neurovascular coupling [8] and demonstrating that a paradigm can produce reproducible responses [16,36–42].

Controlled experimentation is a crucial element of research and the establishment of agreed-upon approaches is fundamental to the wider, prospectively applied use of any new technology. The potential for evaluation of drug effects on CNS systems with fMRI provides an exciting opportunity in drug development, although at present this is occurring in the absence of standardized approaches, in contrast to other imaging methods more routinely used in registration trials [1–3]. With fMRI, there are additional complexities inherent in the method in terms of data acquisition, analysis and interpretation. Here, we have stressed these issues and provided a framework within which to approach these in a systematic fashion. Given that there are many academic institutions performing high-quality fMRI, an increasing number of imaging scientists employed in pharmaceutical companies and an increasing involvement of CROs in imaging, a consensus view on guidelines for GIP for fMRI in clinical studies of novel therapeutics would be extremely beneficial to help drive consistency of practice when used in this context and aid a more widespread site selection for robust fMRI trials. Recommendations at a finer level of detail regarding image acquisition, QA and analysis would be particularly useful but would more logically follow if and when specific paradigms find a niche in drug development. To draw a parallel with another functional imaging method, the development of consensus recommendations for the use of DCE-MRI in early-phase anti-angiogenic cancer drug trials [4,5] has been helpful in enabling primary endpoints to be specified consistently and compared across studies, drugs, laboratories and sponsors. Moreover, the ability of the field to reach such a consensus signaled a maturity of the technique that has also been helpful in its adoption by industry. Importantly, these consensus documents also highlighted key questions for which additional data were required to make a recommendation, thus helping to focus activity on important areas for drug development applications. For these same reasons, the case for a systematic application of fMRI to drug development would be strengthened if consistency in some aspects of the experiment could be established.

The guidelines proposed in the present (and its companion [21]) article were aimed at the process level and, thus, are not specific to any one paradigm or experimental set-up. We hope that this might stimulate discussion in the field toward broadly agreed recommendations applicable to the use of fMRI in drug development. Beyond this, the standardization of specific fMRI protocols represents another level of detail. Any such standards should be based on well-established and well-characterized paradigms that engage robustly specific brain systems with applicability to particular therapeutic areas. This would aid implementation in a multi-site setting and also at a wider range of sites beyond the leading academic centers and, hence, would be of great utility for use in drug development, where site selection and geographical jurisdiction might be constrained by the clinical development program of the drug candidate.

References

1 Steiger, P. (2009) Use of imaging biomarkers for regulatory studies. *J. Bone Joint Surg. Am.* 91 (Suppl. 1), 132–136
2 Jack, C.R., Jr *et al.* (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691
3 Jack, J. *et al.* (2010) Update on the Magnetic Resonance Imaging core of the Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dement.* 6, 212–220
4 Leach, M.O. *et al.* (2005) The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. *Br. J. Cancer* 92, 1599–1610
5 Leach, M.O. *et al.* (2003) Assessment of antiangiogenic and antivascular therapeutics using MRI: recommendations for appropriate methodology for clinical trials. *Br. J. Radiol.* 76, S87–S91
6 Lesko, L.J. (2007) Paving the critical path: how can clinical pharmacology help achieve the vision? *Clin. Pharmacol. Ther.* 81, 170–177
7 Woodcock, J. and Woosley, R. (2008) The FDA critical path initiative and its influence on new drug development. *Annu. Rev. Med.* 59, 1–12
8 Iannetti, G.D. and Wise, R.G. (2007) BOLD functional MRI in disease and pharmacological studies: room for improvement? *Magn. Reson. Imaging* 25, 978–988
9 Glover, G.H. *et al.* (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn. Reson. Med.* 44, 162–167
10 Jones, T.B. *et al.* (2008) Integration of motion correction and physiological noise regression in fMRI. *NeuroImage* 42, 582–590
11 Bourel, P. *et al.* (1999) Automatic quality assessment protocol for MRI equipment. *Med. Phys.* 26, 2693–2700
12 Friedman, L. and Glover, G.H. (2006) Report on a multicenter fMRI quality assurance protocol. *J. Magn. Reson. Imaging* 23, 827–839
13 Stocker, T. *et al.* (2005) Automated quality assurance routines for fMRI data applied to a multicenter study. *Hum. Brain Mapp.* 25, 237–246
14 Simmons, A. *et al.* (1999) Quality control for functional magnetic resonance imaging using automated data analysis and Shewhart charting. *Magn. Reson. Med.* 41, 1274–1278
15 Olsrud, J. *et al.* (2008) A two-compartment gel phantom for optimization and quality assurance in clinical BOLD fMRI. *Magn. Reson. Imaging* 26, 279–286
16 Friedman, L. *et al.* (2008) Test–retest and between-site reliability in a multicenter fMRI study. *Hum. Brain Mapp.* 29, 958–972
17 Ford, J.M. *et al.* (2009) Tuning in to the voices: a multisite FMRI study of auditory hallucinations. *Schizophr. Bull.* 35, 58–66
18 Potkin, S.G. and Ford, J.M. (2009) Widespread cortical dysfunction in schizophrenia: the FBIRN imaging consortium. *Schizophr. Bull.* 35, 15–18
19 Poldrack, R.A. *et al.* (2008) Guidelines for reporting an fMRI study. *Neuroimage* 40, 409–414
20 Carter, C.S. *et al.* (2008) Optimizing the design and analysis of clinical functional magnetic resonance imaging research studies. *Biol. Psychiatry* 64, 842–849
21 Schwarz, A.J. *et al.* (2011) A procedural framework for good imaging practice in pharmacological fMRI studies applied to drug development #1: processes and requirements. *Drug Discov. Today* doi:10.1016/j.drudis.2011.03.011
22 Verma, A. *et al.* (2010) Incorporating functional MRI into clinical pharmacology trials. In *Imaging in CNS Drug Discovery and Development* (Borsook, D., ed.), pp. 153–162, Springer
23 George, E. *et al.* (2010) Evaluation of novel drugs using fMRI in early phase clinical trials: safety monitoring. *Drug Discov. Today* 15, 684–689
24 Sossalla, S. and Schmitto, J.D. (2009) Scientific teamwork – a particular approach. *Kardiol. Pol.* 67, 1421–1423
25 Taneva, S. *et al.* (2010) Decoding the perioperative process breakdowns: a theoretical model and implications for system design. *Int. J. Med. Inform.* 79, 14–30
26 Zeltser, M.V. and Nash, D.B. (2010) Approaching the evidence basis for aviation-derived teamwork training in medicine. *Am. J. Med. Qual.* 25, 13–23
27 Wise, R.G. and Tracey, I. (2006) The role of fMRI in drug discovery. *J. Magn. Reson. Imaging* 23, 862–876
28 Mitsis, G.D. *et al.* (2008) Regions of interest analysis in pharmacological fMRI: how do the definition criteria influence the inferred result? *Neuroimage* 40, 121–132
29 Barch, D.M. and Mathalon, D.H. Using brain imaging measures in studies of procognitive pharmacologic agents in schizophrenia: psychometric and quality assurance considerations. *Biol. Psychiatry* 2011 Feb 18 [Epub ahead of print] doi:10.1016/j.biopsych.2011.01.004
30 Vul, E. *et al.* (2009) Puzzlingly high correlations in fMRI studies of emotion, personality and social cognition. *Perspect. Psychol. Sci.* 4, 274–290
31 Kriegeskorte, N. *et al.* (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535–540
32 Suckling, J. *et al.* (2008) Components of variance in a multicentre functional MRI study and implications for calculation of statistical power. *Hum. Brain Mapp.* 29, 1111–1122
33 Colombo, P. *et al.* (2004) Multicenter trial for the set-up of a MRI quality assurance programme. *Magn. Reson. Imaging* 22, 93–101
34 Logothetis, N.K. (2008) What we can do and what we cannot do with fMRI. *Nature* 453, 869–878
35 Haller, S. and Bartsch, A.J. (2009) Pitfalls in FMRI. *Eur. Radiol.* 19, 2689–2706
36 Costafreda, S.G. *et al.* (2007) Multisite fMRI reproducibility of a motor task using identical MR systems. *J. Magn. Reson. Imaging* 26, 1122–1126
37 Gountouna, V.E. *et al.* (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 Zandbelt, B.B. *et al.* (2008) Within-subject variation in BOLD-fMRI signal changes across repeated measurements: quantification and implications for sample size. *NeuroImage* 42, 196–206
39 Smith, S.M. *et al.* (2005) Variability in fMRI: a re-examination of inter-session differences. *Hum. Brain Mapp.* 24, 248–257
40 Bosnell, R. *et al.* (2008) Reproducibility of fMRI in the clinical setting: implications for trial designs. *NeuroImage* 42, 603–610
41 Tjandra, T. *et al.* (2005) Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. *NeuroImage* 27, 393–401
42 Caceres, A. *et al.* (2009) Measuring fMRI reliability with the intra-class correlation coefficient. *NeuroImage* 45, 758–768