



# A Hitchhiker's Guide to Functional Magnetic Resonance Imaging

José M. Soares<sup>1,2\*</sup>, Ricardo Magalhães<sup>1,2</sup>, Pedro S. Moreira<sup>1,2</sup>, Alexandre Sousa<sup>1,2,3</sup>, Edward Ganz<sup>1,2</sup>, Adriana Sampaio<sup>4</sup>, Victor Alves<sup>3</sup>, Paulo Marques<sup>1,2</sup> and Nuno Sousa<sup>1,2,5</sup>

<sup>1</sup> Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal, <sup>2</sup> ICVS/3B's - PT Government Associate Laboratory, Braga, Portugal, <sup>3</sup> Department of Informatics, University of Minho, Braga, Portugal,

<sup>4</sup> Neuropsychophysiology Lab, CIPsi, School of Psychology, University of Minho, Braga, Portugal, <sup>5</sup> Clinical Academic Center – Braga, Braga, Portugal

## OPEN ACCESS

**Edited by:**

Amir Shmuel,  
McGill University, Canada

**Reviewed by:**

Xin Di,  
New Jersey Institute of Technology,  
USA

Gerard R. Ridgway,  
University of Oxford, UK  
Boris Bernhardt,

Montreal Neurological Institute and  
Hospital, Canada

**\*Correspondence:**

José M. Soares  
josesoares@med.uminho.pt

**Specialty section:**

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
*Frontiers in Neuroscience*

**Received:** 14 July 2016

**Accepted:** 25 October 2016

**Published:** 10 November 2016

**Citation:**

Soares JM, Magalhães R, Moreira PS,  
Sousa A, Ganz E, Sampaio A, Alves V,  
Marques P and Sousa N (2016) A  
Hitchhiker's Guide to Functional  
Magnetic Resonance Imaging.  
*Front. Neurosci.* 10:515.  
doi: 10.3389/fnins.2016.00515

Functional Magnetic Resonance Imaging (fMRI) studies have become increasingly popular both with clinicians and researchers as they are capable of providing unique insights into brain functions. However, multiple technical considerations (ranging from specifics of paradigm design to imaging artifacts, complex protocol definition, and multitude of processing and methods of analysis, as well as intrinsic methodological limitations) must be considered and addressed in order to optimize fMRI analysis and to arrive at the most accurate and grounded interpretation of the data. In practice, the researcher/clinician must choose, from many available options, the most suitable software tool for each stage of the fMRI analysis pipeline. Herein we provide a straightforward guide designed to address, for each of the major stages, the techniques, and tools involved in the process. We have developed this guide both to help those new to the technique to overcome the most critical difficulties in its use, as well as to serve as a resource for the neuroimaging community.

**Keywords:** fMRI, hitchhiker's guide, acquisition, preprocessing, analysis

## INTRODUCTION

Introduced in the early nineties, functional Magnetic Resonance Imaging (fMRI) (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992; Bandettini, 2012a; Kwong, 2012) is a variant of conventional Magnetic Resonance Imaging (MRI) intended to measure brain activity and connectivity. It is a fundamentally non-invasive method, and one which provides a method to assess brain function with unparalleled spatial specificity. Amongst its attributes are high spatial resolution, signal reliability, robustness, and reproducibility.

Functional brain mapping is most commonly performed using the venous blood oxygenation level-dependent (BOLD) contrast technique (Ogawa and Lee, 1990; Ogawa et al., 1990a,b; Ogawa, 2012). The magnitude of the BOLD signal is an indirect measure of neuronal activity, and is a composite which reflects changes in regional cerebral blood flow, volume, and oxygenation. Functional MRI principles and basic concepts have been extensively described and reviewed in the literature (Le Bihan, 1996; Gore, 2003; Amaro and Barker, 2006; Norris, 2006; Logothetis, 2008; Buxton, 2009; Faro and Mohamed, 2010; Ulmer and Jansen, 2010; Poldrack et al., 2011; Bandettini, 2012b; Uğurbil and Ogawa, 2015). In summary, the basic concept underlying all fMRI measurement is that an increase in local neuronal activity stimulates both higher energy consumption and increased blood flow. The resultant indirect determination of brain function

is typically represented as a statistical map which reflects regional activity. Information transfer between neurons is a metabolically demanding process, which requires an increased flow of oxygenated blood, oxyhemoglobin. The local influx of oxygenated blood results in a net increase in the balance of oxygenated arterial blood to deoxygenated venous blood (associated with elevated deoxyhemoglobin). The increase in the oxy-/deoxy-hemoglobin ratio leads to an increase in the MRI signal compared to that of the surrounding tissue. It is important to note that as local neuronal activity increases, there is an intrinsic delay before regional vasodilation occurs and flow increases. This mechanism, which is a function of the properties of the local vascular network, is referred to as the hemodynamic response function (HRF) and has a time course of several seconds after the increase in activity. The BOLD signal can be characterized by the shape of this HRF, which reflects its vascular origin. Typically, fMRI software model the HRF with a set of gamma functions, commonly designated by canonical HRF, that is characterized by a gradual rise, peaking ~5–6 s after the stimulus, followed by a return to the baseline (about 12 s after the stimulus) and a small undershoot before stabilizing again, 25–30 s after (**Figure 1A**) (Miezin et al., 2000; Buxton et al., 2004; Handwerker et al., 2012). Occasionally, an initial dip is reported but its origin and implications are still under debate (Hu and Yacoub, 2012). This also highlights that, despite the good fit of the canonical HRF for most situations, the true HRF is known to present some variability. Whenever a researcher suspects that the canonical HRF is not good enough, it is common practice to include its temporal and dispersion derivatives in the model in order to estimate the variability in latency and shape, respectively (Friston et al., 1998; Calhoun et al., 2004).

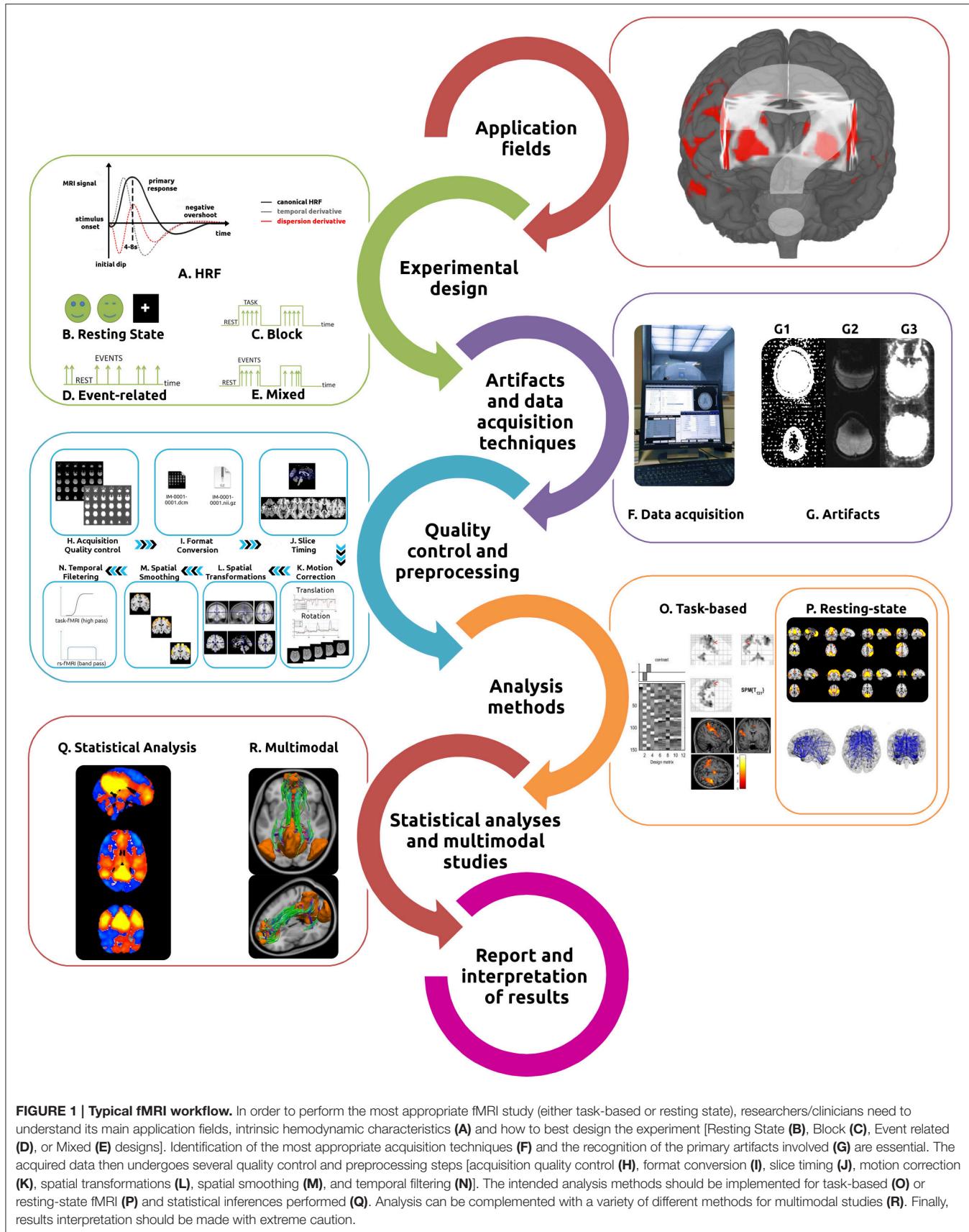
It has become an established practice in fMRI studies to investigate the differential neuronal responses to various forms of stimuli and activity during task performance. Typical investigations have compared periods of brain activation during a task with periods of a matched baseline task or a “rest” condition (Bandettini et al., 1992; Blamire et al., 1992; Vallesi et al., 2015). However, stimulus-evoked responses are only the tip of the iceberg in brain activity. More recently, a new perspective in functional imaging has brought with it the recognition that spontaneous/intrinsic brain activity is a fundamental aspect of normal brain function. Technical advances in neuroimaging methods have contributed to this paradigm shift, and have led to the recognition that the brain is more accurately considered a network of functionally connected (co-varying) and constantly interacting regions, requiring a focus on understanding patterns of connectivity as well as localized activation (Biswal et al., 1995; Carlson et al., 2003; Fox et al., 2005; Fox and Raichle, 2007; Raichle, 2009; Smith et al., 2011). For this reason, resting state fMRI (rs-fMRI) analyses rely upon spontaneous coupled brain activity to reveal intrinsic signal fluctuations in the absence of external stimuli or demands of imposed tasks (Damoiseaux et al., 2006; Fox and Raichle, 2007; Schölyvinck et al., 2010; van den Heuvel and Hulshoff Pol, 2010; Friston et al., 2014b). Complimentary approaches, combining rs-fMRI with functional deactivation (shifting from periods of stimulation to those of rest) also have been described to study functional activity transitions

(Greicius and Menon, 2004; Anticevic et al., 2012; Soares et al., 2016).

With its popularity steadily increasing among clinicians and researchers, the technique of fMRI has demonstrated great utility in the study of the functioning brain, both in health and disease. It is important to recognize, however, that it has an intrinsically complex workflow (summarized in **Figure 1**) which assumes broad knowledge of task design, imaging artifacts, complex MRI acquisition techniques, a multitude of preprocessing and analysis methods (in several software packages see **Tables 1–4**), statistical analyses, as well as interpretation of results. Several papers and books describing the main technical issues and pitfalls related to both intrinsic and evoked activity have been published (Jezzard and Song, 1996; Le Bihan, 1996; Norris, 2006; Haller and Bartsch, 2009; Cole et al., 2010; Margulies et al., 2010; Poldrack et al., 2011; Davis and Poldrack, 2013; Lee et al., 2013; Uğurbil and Ogawa, 2015). However, given the complex nature of the data processing, constant methodological advances and the increasingly broad application of fMRI to both the clinical and research domains, we have sought to compile a practical “hitchhiker's guide,” containing essential information and primary references. These guides have proven to be important to assist in the optimization of data quality and interpretation of results (Soares et al., 2013). We also have provided an analysis of the principal software tools available for each step in the workflow, highlighting the most suitable features of each. Through this process it is our goal to enable investigators/clinicians to design and implement practical workflows which will lead to robust and reproducible results. In the following sections, information about each specific fMRI workflow step, from the current technique applications to the final results interpretation, will be discussed in detail. We have started by presenting a list of common software tools used for fMRI pipelines (**Table 1**), including both applications for general and wide-ranging purposes (e.g., AFNI, BrainVoyager, FSL, or SPM) as well as for very specific tasks (e.g., Marsbar and NBS).

## APPLICATION FIELDS

The use of the technique of fMRI has led to significant expansion of understanding in multiple areas of cognitive neuroscience (Cabeza, 2001; Raichle, 2001; Poldrack, 2008, 2012). It has, for example, been successfully used to study systems involved with sensory-motor functions (Biswal et al., 1995; Calvo-Merino et al., 2005), language (Woermann et al., 2003; Centeno et al., 2014), visuospatial orientation (Formisano et al., 2002; Rao and Singh, 2015), attention (Vuilleumier et al., 2001; Markett et al., 2014), memory (Machulda et al., 2003; Sidhu et al., 2015) affective processing (Kiehl et al., 2001; Shinkareva et al., 2014), working memory (Curtis and D'Esposito, 2003; Meyer et al., 2015), personality dimensions (Canli et al., 2001; Sampaio et al., 2014), decision-making (Bush et al., 2002; Soares et al., 2012), and executive function (Just et al., 2007; Di et al., 2014). Functional MRI has also been used as a tool in the study of topics as diverse as addiction behavior (Chase and Clark, 2010; Kober et al., 2016), neuromarketing (Ariely and Berns, 2010; Kuhn et al., 2016) and politics (Knutson et al., 2006), among others.



**FIGURE 1 | Typical fMRI workflow.** In order to perform the most appropriate fMRI study (either task-based or resting state), researchers/clinicians need to understand its main application fields, intrinsic hemodynamic characteristics (**A**) and how to best design the experiment [Resting State (**B**), Block (**C**), Event related (**D**), or Mixed (**E**) designs]. Identification of the most appropriate acquisition techniques (**F**) and the recognition of the primary artifacts involved (**G**) are essential. The acquired data then undergoes several quality control and preprocessing steps [acquisition quality control (**H**), format conversion (**I**), slice timing (**J**), motion correction (**K**), spatial transformations (**L**), spatial smoothing (**M**), and temporal filtering (**N**)]. The intended analysis methods should be implemented for task-based (**O**) or resting-state fMRI (**P**) and statistical inferences performed (**Q**). Analysis can be complemented with a variety of different methods for multimodal studies (**R**). Finally, results interpretation should be made with extreme caution.

**TABLE 1 | Software tools used for fMRI pipelines present in published studies.**

Software tools	URL	Main purpose
Analysis of Functional NeuroImages (AFNI; Cox, 1996, 2012)	<a href="https://afni.nimh.nih.gov/afni/">https://afni.nimh.nih.gov/afni/</a>	Preprocessing, analysis, and statistical analysis
Analyze 4D	<a href="http://analyze4d.com/">http://analyze4d.com/</a>	Region of interest and time-series analysis
AnalyzeFMRI (Bordier et al., 2011)	<a href="https://cran.r-project.org/web/packages/AnalyzeFMRI">https://cran.r-project.org/web/packages/AnalyzeFMRI</a>	Independent component analysis
BioImage Suite (Papademetris et al., 2006)	<a href="http://bioimagesuite.yale.edu/">http://bioimagesuite.yale.edu/</a>	Preprocessing, analysis, and statistical analysis (modules from AFNI)
Brain Connectivity Toolbox (Rubinov and Sporns, 2010)	<a href="http://www.brain-connectivity-toolbox.net/">http://www.brain-connectivity-toolbox.net/</a>	Graph theory analysis
BrainVISA	<a href="http://brainvisa.info/web/index.html">http://brainvisa.info/web/index.html</a>	Analysis and statistical analysis (preprocessing modules from SPM or FSL)
BrainVoyager (Goebel, 2012)	<a href="http://www.brainvoyager.com/">http://www.brainvoyager.com/</a>	Preprocessing, analysis, and statistical analysis
BROCCOLI (Eklund et al., 2014)	<a href="https://github.com/wanderine/BROCCOLI/">https://github.com/wanderine/BROCCOLI/</a>	Preprocessing, analysis, and statistical analysis (mainly non parametric) with GPU implementation
Cambridge Brain Analysis (CamBA)	<a href="http://www.bmu.psychiatry.cam.ac.uk/software/">http://www.bmu.psychiatry.cam.ac.uk/software/</a>	Analysis and statistical analysis
Cambridge Centre for Ageing and Neuroscience (Cam-CAN)	<a href="http://www.cam-can.org/">http://www.cam-can.org/</a>	Structural equation modeling analysis
Functional connectivity toolbox (CONN)	<a href="https://www.nitrc.org/projects/conn/">https://www.nitrc.org/projects/conn/</a>	Preprocessing, analysis, and statistical analysis
Cosmo multi-variate pattern analysis toolbox (CosmoMVPA)	<a href="http://cosmomvpa.org/">http://cosmomvpa.org/</a>	Multi-voxel pattern analysis
Configurable Pipeline for the Analysis of Connectomes (C-PAC)	<a href="https://fcp-indi.github.io/">https://fcp-indi.github.io/</a>	Preprocessing, analysis, and statistical analysis (based on AFNI, FSL, and ANTS)
Data Processing and Analysis for Brain Imaging (DPABI; Yan et al., 2016)	<a href="http://rfmri.org/dpabi">http://rfmri.org/dpabi</a>	Preprocessing, analysis, and statistical analysis
Data Processing Assistant for Resting-State fMRI (DPARSF; Yan and Zang, 2010)	<a href="http://rfmri.org/DPARSF">http://rfmri.org/DPARSF</a>	Preprocessing and analysis
A software for dynamic functional connectivity analysis of fMRI data (DynaConn)	<a href="http://softwarecircuits.weebly.com/dynacconn.html">http://softwarecircuits.weebly.com/dynacconn.html</a>	Dynamic functional connectivity
Dynamic brain connectome (DynamicBC; Liao et al., 2014)	<a href="http://restfmri.net/forum/DynamicBC">http://restfmri.net/forum/DynamicBC</a>	Dynamic functional connectivity
fMRI Advanced Normalization Tools (ANTsR; Avants et al., 2009)	<a href="http://stnava.github.io/fMRIANTS/">http://stnava.github.io/fMRIANTS/</a>	Preprocessing, analysis, and statistical analysis
FMRLAB	<a href="http://scnn.ucsd.edu/fmrlab/">http://scnn.ucsd.edu/fmrlab/</a>	Independent component analysis
Freesurfer (FSFAST; Fischl, 2012)	<a href="http://freesurfer.net/">http://freesurfer.net/</a>	Preprocessing, analysis, and statistical analysis
FMRIB Software Library (FSL; Jenkinson et al., 2012)	<a href="http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/">http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/</a>	Preprocessing, analysis, and statistical analysis
Group ICA Of fMRI Toolbox (GIFT; Calhoun et al., 2009)	<a href="http://mialab.mrn.org/software/gift/">http://mialab.mrn.org/software/gift/</a>	Independent component analysis
Group Iterative Multiple Model Estimation (GIMME; Gates and Molenaar, 2012)	<a href="https://www.nitrc.org/projects/gimme/">https://www.nitrc.org/projects/gimme/</a>	Structural equation modeling analysis
GLM Flex	<a href="http://mrtools.mgh.harvard.edu/">http://mrtools.mgh.harvard.edu/</a>	Statistical analysis
Granger Multivariate Autoregressive Connectivity (GMAC; Tana et al., 2012)	<a href="https://www.nitrc.org/projects/gmac_2012/">https://www.nitrc.org/projects/gmac_2012/</a>	Preprocessing, analysis (Granger causality mapping)
Generalized psychophysiological interactions (GPPI; McLaren et al., 2012)	<a href="https://www.nitrc.org/projects/gppi">https://www.nitrc.org/projects/gppi</a>	Psychophysiological interactions analysis
A user-friendly toolbox for comprehensive graph analyses of functional brain connectivity (GraphVar; Kruschwitz et al., 2015)	<a href="http://rfmri.org/GraphVar">http://rfmri.org/GraphVar</a>	Graph theory analysis
GRaph thEoretICAL Network Analysis (GRETA; Wang J. et al., 2015)	<a href="https://www.nitrc.org/frs/shownotes.php?release_id=2213">https://www.nitrc.org/frs/shownotes.php?release_id=2213</a>	Graph theory analysis
Graph Theory GLM (GTG)	<a href="https://www.nitrc.org/projects/metabolab_gtg/">https://www.nitrc.org/projects/metabolab_gtg/</a>	Preprocessing and graph theory analysis
Marsbar	<a href="http://marsbar.sourceforge.net/">http://marsbar.sourceforge.net/</a>	Region of interest analysis
Multivariate Granger causality (MVGC; Barnett and Seth, 2014)	<a href="http://www.sussex.ac.uk/sackler/mvgc/">http://www.sussex.ac.uk/sackler/mvgc/</a>	Granger causality mapping analysis

(Continued)

**TABLE 1 | Continued**

Software tools	URL	Main purpose
Network Based Statistic (NBS; Zalesky et al., 2010)	<a href="https://www.nitrc.org/projects/nbs/">https://www.nitrc.org/projects/nbs/</a>	Graph theory analysis and statistical analysis
NeuroLens	<a href="https://www.nitrc.org/projects/nldo/">https://www.nitrc.org/projects/nldo/</a>	Preprocessing, analysis, and statistical analysis
Nipy (Millman and Brett, 2007)	<a href="http://nipy.org/">http://nipy.org/</a>	Preprocessing, analysis, and statistical analysis
Nitime	<a href="http://nipy.org/nitime/">http://nipy.org/nitime/</a>	Time-series analysis
Pattern Recognition for Neuroimaging Toolbox (PRONTO)	<a href="http://www.mlnl.cs.ucl.ac.uk/pronto/prtsoftware.html">http://www.mlnl.cs.ucl.ac.uk/pronto/prtsoftware.html</a>	Multi-voxel pattern analysis
MultiVariate Pattern Analysis in Python (PyMVPA; Hanke et al., 2009)	<a href="http://www.pymvpa.org/">http://www.pymvpa.org/</a>	Multi-voxel pattern analysis
Structural equation modeling for fMRI (SEM)	<a href="http://dslink333.dyndns.org/SEM.htm">http://dslink333.dyndns.org/SEM.htm</a>	Structural equation modeling analysis (toolbox for SPM)
Statistical non Parametric Mapping (SnPM; Nichols and Holmes, 2002)	<a href="http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/snpm">http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/snpm</a>	Non-parametric permutation/randomization analysis (toolbox for SPM)
Statistical Parametric Mapping (SPM; Friston et al., 2007)	<a href="http://www.fil.ion.ucl.ac.uk/spm/">http://www.fil.ion.ucl.ac.uk/spm/</a>	Preprocessing, processing and statistical analysis
The Decoding Toolbox (Hebart et al., 2014)	<a href="http://www.bccn-berlin.de/tdt">http://www.bccn-berlin.de/tdt</a>	Multi-voxel pattern analysis (toolbox optimized for SPM)

**TABLE 2 | Most common software tools used to program and present fMRI stimuli.**

Software tools	URL	OS	GUI	Programming skills required	Availability
A Simple Framework (ASF) (Schwarzbach, 2011)	<a href="https://code.google.com/p/asf/">https://code.google.com/p/asf/</a>	Windows (Matlab)	✗	✓	Open-source
BOLDSync (Joshi et al., 2014)	<a href="http://www.nbrc.ac.in/faculty/pravat/BOLDSync.php">http://www.nbrc.ac.in/faculty/pravat/BOLDSync.php</a>	Windows, Linux (Matlab)	✓	✗	Open-source
Cogent 2000	<a href="http://www.vislab.ucl.ac.uk/cogent_2000.php">http://www.vislab.ucl.ac.uk/cogent_2000.php</a>	Windows (Matlab)	✗	✓	Open-source
E-Prime (Psychology Software Tools, Pittsburgh, PA)	<a href="http://www.pstnet.com/">http://www.pstnet.com/</a>	Windows	✓	✗	Commercial
FLXLab	<a href="http://flxlab.sourceforge.net/">http://flxlab.sourceforge.net/</a>	Windows, Linux, OS X	✗	✓	Open-source
Inquisit	<a href="http://www.millisecond.com/">http://www.millisecond.com/</a>	Windows, OS X	✓	✓	Commercial
NordicAktiva	<a href="http://www.nordicneurolab.com/">http://www.nordicneurolab.com/</a>	Windows	✓	✗	Commercial
Paradigm	<a href="http://www.paradigmexperiments.com/">http://www.paradigmexperiments.com/</a>	Windows	✓	✗	Commercial
Presentation® (Neurobehavioral systems)	<a href="https://www.neurobs.com/">https://www.neurobs.com/</a>	Windows	✓	✓	Commercial
Psychophysics Toolbox (Brainard, 1997)	<a href="http://psychtoolbox.org/">http://psychtoolbox.org/</a>	Windows, Linux, OS X (Matlab/Octave)	✗	✓	Open-source
PsychoPy (Peirce, 2008)	<a href="http://www.psychopy.org/">http://www.psychopy.org/</a>	Windows, Linux, OS X	✓	✓	Open-source
PsyScope X (Cohen et al., 1993)	<a href="http://psy.ck.sissa.it/">http://psy.ck.sissa.it/</a>	OS X	✓	✓	Open-source
Stim 2	<a href="http://compumedicsneuroscan.com/">http://compumedicsneuroscan.com/</a>	Windows	✓	✗	Commercial
SuperLab	<a href="http://www.superlab.com/">http://www.superlab.com/</a>	Windows, OS X	✓	✗	Commercial
Wake Forest Visual Experimentation Software (WaVE) (Meyer and Constantinidis, 2005)	<a href="http://www.wakehealth.edu/Research/Neurobiology-and-Anatomy/Wake-Forest-Visual-Experimentation-Software-(WaVE).htm">http://www.wakehealth.edu/Research/Neurobiology-and-Anatomy/Wake-Forest-Visual-Experimentation-Software-(WaVE).htm</a>	Windows (Matlab)	✓	✓	Open-source

Another area in which fMRI is expanding is in clinical neuroimaging, with applications which range from pre-surgical mapping/planning (Stippich, 2015; Lee et al., 2016) to functional characterization of a variety of disease states (Matthews et al., 2006; Bullmore, 2012), as well as in understanding plasticity, contributing to the study of drug development (Wise and Preston, 2010; Duff et al., 2015), and in the study of genetically

determined differences in function (Koten et al., 2009; Richiardi et al., 2015).

While these research and clinical-oriented fields of fMRI application usually require a task or stimulus evoked brain response, the defining attribute of not requiring active patient participation, triggered the use of rs-fMRI for additional research and clinical applications. This research methodology has been

**TABLE 3 | A list of the main preprocessing steps implemented by the common fMRI tools \***.

Software	Preprocessing steps									
	DICOM import	Slice Timing	Motion Correction	Motion outlier detection	Automated skull striping	Coregister	Normalization	Confounding removal	Spatial Smoothing	Temporal Filtering
AFNI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BiolImage Suite	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓
BrainVoyager	✓	✓	✓	✗	✓	✓	✓	✗	✓	✓
BROCCOLI	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓
CONN	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓
C-PAC	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
DPABI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DPARSF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
fMRIANTs/ANTsR	✗	✗	✓	✗	✗	✓	✓	✓	✓	✗
Freesurfer	✓	✓	✓	✗	✓	✓	✓	✗	✓	✗
FSL	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
GMAC	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓
GTG	✗	✓	✓	✓	✗	✗	✗	✓	✗	✓
NeuroLens	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗
Nipy	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SPM	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓

\*To the best of our knowledge at the date of submission, based on information gathered from the software manuals, main webpages and published papers.

applied to prognostic and diagnostic information (Fox and Greicius, 2010; Lang et al., 2014), treatment guidance (Rosazza and Minati, 2011; Castellanos et al., 2013), identification of functional fingerprints, discovery, and validation of biomarkers in the investigation of relation to cognitive, emotional, and social processes (Fox M. D. et al., 2014; Krishnadas et al., 2014; Finn et al., 2015).

Finally, preliminary work appears to offer promise for the use of real-time neurofeedback and/or brain computer interfaces methodologies (Caria et al., 2012; Weiskopf, 2012; Sulzer et al., 2013; Kadosh et al., 2016), in the treatment of disorders such as Obsessive Compulsive Disorder (Emmert et al., 2016), Depression (Young et al., 2014), and Schizophrenia (Scheinost et al., 2013; Cordes et al., 2015).

## EXPERIMENTAL DESIGN

The number of variables (such as the specific nature of the research question, availability of imaging instruments, demand of data handling, and cost) associated with each study makes it essential to optimize BOLD signal acquisition time and statistical efficiency of the analysis. There is not one optimal design which will encompass all fMRI studies. However, optimizing certain parameters can significantly improve the study efficiency and reliability of the final results. Some reviews and book chapters have already provided the basic fMRI experimental design concepts (Amaro and Barker, 2006; Friston et al., 2007; Filippi, 2009; Bennett and Miller, 2013; Maus and van Breukelen, 2013). The experimental designs used in fMRI are resting state and task-based.

## Resting State

Characterization of the resting state is the most straightforward experimental design in fMRI. The subjects are not performing any explicit task (**Figure 1B**). During acquisitions performed under these circumstances, consistent and stable functional patterns, which are reproducible across individuals, sessions, scanners, and methods can be identified and are known as Resting State Networks (RSNs) (Fox et al., 2005; Damoiseaux et al., 2006; Long et al., 2008; Choe et al., 2015; Jovicich et al., 2016). That said, the specific resting conditions and the duration of the acquisition both have an important effect on the final functional signals. The most traditional design consists of instructing the participants to keep their eyes closed, not to think about anything in particular and not falling asleep. Alternative approaches have included keeping the eyes open or keeping the eyes open while fixating upon an object in the visual field, such as a cross, during scanning. The most suitable approach depends on the research question and purpose. If reliability and consistency are of upmost importance, the eyes fixated condition should be preferred, except for the primary visual network whose connectivity is more reliable with the eyes open but not fixated condition (Yan et al., 2009; Patriat et al., 2013; Zou et al., 2015). On the other hand, if the focus is on obtaining higher functional connectivity (FC) strength, eyes open, either fixated or not, should be used (Yan et al., 2009; Van Dijk et al., 2010). The chosen approach can also have a significant impact on the topological organization (Xu et al., 2014), global signal amplitude (Wang X.-H. et al., 2015; Wong et al., 2016), and directionality (Zhang et al., 2015). Nevertheless, the different resting-state conditions present comparable results, and thus the choice of the condition should also take into account which is more comfortable/appropriate for

**TABLE 4 | A list of the main analysis methods implemented by the common fMRI tools\*.**

Software	Task-based fMRI						Resting-state fMRI							
	GLM	PPI	SEM	DCM	GCM	MVPA	Seed-Based	ReHO	ALFF	PCA	ICA	Clustering	Graph Theory	dFC
AFNI	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✗	✗
AnalyzeFMRI	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
BiolImage Suite	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Brain Connectivity Toolbox	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
BrainVoyager	✓	✓	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗
BROCCOLI	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
CamBA	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Cam Can	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
CONN	✓	✓	✗	✗	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓
CosmoMVPA	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
C-PAC	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✗	✗	✓	✗
DPABI	✓	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✗	✗	✗
DynaConn	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
DynamicBC	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
ANTs/ANTsR fMRI	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
FMRLAB	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
Freesurfer (FSFAST)	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
FSL	✓	✗	✗	✗	✗	✗	✓	✗	✓	✓	✓	✓	✗	✗
GIFT	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
GIMME	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
GLMFlex	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
GMAC	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
GPPI	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
GraphVar	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
GRETNA	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
GTG	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
Lipsia	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
MVGC	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
NBS	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
Neurolens	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Nipy	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Nitime	✗	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✓	✗
PRONTO	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
PyMVPA	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
SEM - Structural Equation Modeling (SEM) for fMRI	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
SnPM	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
SPM	✓	✓	✗	✓	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗
The Decoding Toolbox	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗

\*To the best of our knowledge at the date of submission, based on information gathered from the software manuals, main webpages and published papers.

the study population, keeping in mind that it should be consistent for all the study participants. Differences in scan length also have a demonstrable impact, with acquisition times of 5–7 min shown to yield a reasonable trade-off between time/robustness of RSNs FC (Van Dijk et al., 2010; Whitlow et al., 2011), 5.5 min shown to be acceptable in young children (White et al., 2014), but both increased reliability and greater in-depth analysis are possible with scans of ~13 min (Birn et al., 2013).

## Task-Based

When employing task-based fMRI studies, the way in which the stimuli are presented as a function of time is of upmost importance. The typical experimental designs are termed block (**Figure 1C**), event-related (**Figure 1D**), and mixed block/event-related (**Figure 1E**). The most simple task design, block-design, consists of presenting consecutive stimuli as a series of epochs, or blocks, with stimuli from one condition being presented

during each epoch, followed by an epoch of stimuli from another condition, or with rest/baseline epochs. Specific block duration depends on the type of stimulus, with 15–30 s the most commonly used range, although some researchers suggest an optimal length of 15 s (Maus and van Breukelen, 2013). The order of the conditions is also important, and these are recommended to be counter-balanced across subjects of the same study. Block design allows a straightforward approach, good statistical power, signal amplitude and robustness. However, because each block is of such long duration, the participant's rapid habituation to task as well as the inability to accurately define response-time courses are intrinsic limitations of this design (Dale and Buckner, 1997; Amaro and Barker, 2006; Dosenbach et al., 2006).

Event-related designs are intended to delineate the association between brain functions and discrete events (typically randomized and of short duration between 0.5 and 8 s), separated by an inter-stimulus interval (ISI, normally ranging from 0.5 to 20 s). By incorporating great task flexibility and participant's unpredictability, this design provides the means to detect transient variations in local hemodynamic response. It presents however a more complex analysis process and a decreased signal-to-noise ratio (SNR), the combination of which leads to diminished detection power (Dale, 1999; Miezin et al., 2000; Huettel, 2012; Liu, 2012). Two types of event-related designs can be implemented and are characterized by different ranges of ISI: slow event-related designs, where the individual stimuli are well-separated in time (usually by more than 15 s), which prevents the overlap of successive stimuli HRFs, and rapid event-related designs, where stimuli are closely spaced in time (less than the HRF of the previous stimulus) resulting in the overlap of their HRFs. The latter protocols allow higher stimulus frequencies, resulting in greater statistical power, as well as diminished participant anticipation and boredom (Amaro and Barker, 2006; Huettel, 2012). Additionally, the randomized or pseudo-randomized order of stimuli presentation also is of importance in minimizing habituation. For these rapid event-related designs, implementing variable ISIs (jittering) allows differential overlap of HRFs, reduces multicollinearity problems and may provide better characterization of each condition response (Dale, 1999). Alternative methods as m-sequences (Buracas and Boynton, 2002; Liu, 2004) and genetic algorithms (Wager and Nichols, 2003; Maus et al., 2010) also can be used in event-related experimental designs, in order to reach flexible trade-offs between estimation efficiency and detection power. Some tools which can facilitate the implementation of randomized design are Optseq2 (<https://surfer.nmr.mgh.harvard.edu/optseq/>), RSFGen (<http://homepage.usask.ca/~ges125/fMRI/RSFgen.html>) and the fMRI Simulator (<http://www.mccauslandcenter.sc.edu/crnl/tools/fmristim>).

Combining stimuli in discrete blocks (mixed block/event-related design) provides information about both sustained and transient functional activations during task performance. While the technique offers the advantages of both block and event-related designs, it involves more assumptions, has a poorer HRF estimation and decreased statistical strength of sustained signal, and requires more subjects in order to measure statistically

significant and sustained effects (Visscher et al., 2003; Amaro and Barker, 2006; Petersen and Dubis, 2012).

Independent of the experimental design, the specific way with which the study conditions are modeled (model specification) also plays an important role in the signal optimization process (Price et al., 1997; Friston, 2005; Amaro and Barker, 2006; Friston et al., 2007). The most basic comparison consists of subtracting two or more conditions (e.g., A – B), in which one is typically a control condition. Factorial designs expand this principle to two or more factors (e.g., different cognitive processes), each one with two or more levels. A simple example of such design would be the visualization of two different words in two different colors which would result in 4 conditions: the first word with the first color (A) the first word with the second color (B), the second word with the first color (C) and the second word with the second color (D). This design, not only enables the exploration of the effect of the two main factors (words and colors), but also their interactions, specifically how one factor affects the relation between the other factor and the response variables [e.g., (A – B) – (C – D)]. If the researcher is interested in assessing if the BOLD response to trials is modulated by a continuously varying parameter, a parametric design (e.g., A < A < A < A) would be more suitable. A typical example would be a study where the goal is to assess if the BOLD response increases/decreases linearly with the difficulty of the task. Choosing appropriate baselines and controls is of paramount importance since neural activity may vary unpredictably and overlap (or even exceed in amplitude) regions activated during the target task. A properly defined baseline should allow for maximum sensitivity in the detection of brain activity related to the study target (target isolation) while controlling for as many extraneous variables and unrelated confounds as possible (Stark and Squire, 2001; Peck et al., 2004; Diers et al., 2014). Generic recommendations include the use of multiple baseline conditions, scan times as long as possible (the more trials the better, with several shorter runs preferred over one long run), randomized conditions when possible, avoidance of comparison between trials widely separated in time and keeping participants engaged (Friston et al., 2007).

Several software tools can be used to implement the stated principles and present the task to the participants in the scanner (**Table 2**).

When designing a study involving both task-based and rs-fMRI, in order to avoid contamination of rs-fMRI with residual activity from previous task performance, it is recommended that one perform the resting state acquisition before the task-based or, at the minimum, after a suitable delay (Stevens et al., 2010; Tung et al., 2013).

## Power Analyses

The question about “how large is enough” is a matter of debate in the neuroimaging field to determine the appropriate study sample size. For example, in an attempt to establish the boundaries for an adequate sample size, sensitivity and sensibility analyses were conducted, demonstrating that sample sizes of at least 27 subjects provide adequate reliability for fMRI investigations (Thirion et al., 2007). Additionally, in a controversial technical note (Friston, 2012), it was suggested that

there is an optimal sample size, compared to which sample sizes could be either too small (studies with less than 16 subjects) or even, although less frequently, too large (studies with more than 32 subjects), under the arguments of reduced power or meaningless/trivial findings resulting from overpowered studies, respectively. It was mentioned that: on one hand, significant findings obtained in small samples ( $n = 16$ ) indicate large effects being stronger than the same level of significance obtained with larger sample sizes; on the other hand, the relevance of significant findings obtained with large samples can be illustrated with the magnitude of observed effect-sizes. However, criticisms have been outlined (e.g., Yarkoni, 2012), particularly focusing on the liberal assumptions (e.g., significance threshold) in which Friston's arguments were built. Furthermore, it was recently described that a substantial number of published studies are statistically under-powered (Button et al., 2013). In this context, it is important to highlight the use of power analyses as a means to obtain robust and meaningful findings in these studies. Power analyses refer to the probability of rejecting the null hypothesis (given that the alternative hypothesis is true) and allow the establishment of a sample size that will increase the confidence of detecting true effects (Ioannidis, 2008). Functional MRI studies are often characterized by low statistical power, primarily due to limited sample size and large number of comparisons (Murphy and Garavan, 2004). Calculations of power are rarely performed in fMRI research, possibly due to the uncertainty associated to the unknown variance of the BOLD response and also due to the difficulty in predicting expected effects (Guo et al., 2012). Software tools have been developed in order to facilitate calculation of the statistical power, both for estimating the number of subjects to be included in the study, and for the number of stimuli to be presented. In order to employ these tools, information about the mean activation, the variance, the Type I error rate, and the sample size must be provided (Mumford, 2012). The power calculation should use either the statistical images (t/F maps generated by simple study designs) from pilot studies (PowerMap software; Joyce and Hayasaka, 2012), the estimated parameters in specific regions-of-interest (fMRIPower tool) (Mumford and Nichols, 2008) or the prevalence of active peaks (NeuroPower; Durnez et al., 2016).

## DATA ACQUISITION TECHNIQUES AND ARTIFACTS

Performing effective fMRI studies requires a thorough understanding of specific MRI acquisition techniques and artifacts, and how to deal with them (Figures 1F,G). When the activity of a population of neurons within a voxel (minimum spatial resolution unit in each image, the volume element) changes, the associated hemodynamic response can be determined using T2\* weighted MRI acquisitions (details in Buxton, 2009; Hashemi et al., 2012). Detection of the BOLD signal is the most commonly used technique in fMRI, due primarily to its ease of implementation and inherent functional contrast. Alternative detection methods do exist and are based on the measurement of a combination of additional parameters

including: changes in cerebral blood volume (CBV), cerebral blood flow (CBF), and cerebral metabolic rate of oxygen (CMRO2) (Davis et al., 1998). The alternative methods are: calibrated BOLD, based on BOLD contrast but also taking into account physiological variation (e.g., hematocrit, oxygen extraction fraction, and blood volume) (Davis et al., 1998; Blockley et al., 2012); Arterial Spin Labelling (ASL) used to measure regional CBF by tracking intravascular water as an endogenous tracer (Williams et al., 1992; Buxton et al., 1998; Telischak et al., 2015); Vascular-Space-Occupancy (VASO) based on differences between blood and surrounding tissues and determined through dynamic measurement of local CBV (Lu et al., 2003; Lu and van Zijl, 2012); Venous Refocusing for Volume Estimation (VERVE), based on changes in venous cerebral blood volume (Stefanovic and Pike, 2005); Signal Enhancement by Extravascular Protons (SEEP) based on the determination of proton-density changes associated with cellular swelling (Stroman et al., 2003; Figley et al., 2010); and diffusion-weighted fMRI, which measures structural changes in the neural tissues related to cell swelling during activation (Le Bihan, 2012; Aso et al., 2013).

Functional MRI data are generally collected over the entire brain through the acquisition of sequential volumes (time-points), each one composed of a set of slices. The typical sequence used for fMRI studies is echo planar imaging (EPI), which is attractive due both to its imaging speed and BOLD contrast sensitivity, but also associated with inherent artifacts and diminished image quality (Stehling et al., 1991; Poustchi-Amin et al., 2001; Schmitt et al., 2012). EPI may be performed using gradient-echo, spin-echo, or combination techniques. When compared to spin-echo EPI, gradient echo acquisitions have higher BOLD sensitivity, imaging speed and versatility, and have been used in the majority of fMRI studies. On the other hand, spin-echo sequences have been proposed as a viable alternative when the goal is to obtain increased functional localization in the capillary bed (especially at high fields) and when specific regions of interest (ROIs) are less superficial regions such as for example the ventromedial frontal and anterior inferior temporal cortex are the primary focus of the study (Norris, 2012; Boyacioglu et al., 2014; Halai et al., 2014; Chiacchiaretta and Ferretti, 2015).

## Data Acquisition Techniques

In order to minimize artifact, and to obtain the most reliable data it is critically important to optimize the acquisition phase. There is no single "gold standard" fMRI protocol due to the great variability in parameters such as the MRI hardware vendor and configuration, field strength, scanning time available, specific regions under study and subsequent analyses intended. For this reason, we here confine ourselves to a series of suggestions based upon the use of a standard single-shot gradient-echo EPI 3 T fMRI acquisition. When defining an fMRI acquisition protocol, a reasonable strategy is to start from a well-characterized "standard" protocol usually provided by the vendor, and then to modify it according to the specific requirements of the study to be undertaken. A practical description of the parameters involved in a typical fMRI acquisition, and guide to how they should be reported, is provided in Inglis' checklist (Inglis, 2015).

While many characteristics of the individual MRI scanner and of the specific acquisition protocols have a strong impact on the fMRI results, magnetic field strength is amongst the most defining. The amplitude of signal usually associated with the BOLD contrast is very low (around 1% of baseline or less). With increased field strengths the sensitivity is increased as is the spatial resolution and SNR (Gore, 2003; van der Zwaag et al., 2009; Wald, 2012; Skouras et al., 2014), but all at the cost of increased artifact (Triaanafyllou et al., 2005). The majority of scanners currently in use, both in diagnostic and research centers, are units having field strengths of 1.5–3 T, but some research groups are already utilizing 7 T fields, and it is expected that the availability and use of such scanners will increase (Duyn, 2012). Typically, fMRI data are acquired using a series of 2D axial slices to cover the whole brain (one volume) and then the process is repeated to collect a number of volumes over time (time-series). Each volume can be acquired using either interleaved or sequential slice acquisitions. While interleaved acquisitions have less adjacent slice interference, they can be more vulnerable to spin history effects generated by head motion (Muresan et al., 2005). To reduce the influence of both these potential issues, most fMRI acquisitions utilize a gap between slices (around 10–25% of the total slice thickness). Slice acquisition also can be performed either in an ascending (foot-to-head) or descending order, with the former theoretically affected by excitation and saturation of in-flowing blood. Although no significant differences have been reported between the two directions, the most robust approach seems to favor the use of descending sequential acquisitions (Howseman et al., 1999).

An important trade-off in fMRI acquisition is between temporal and spatial resolution. Since the BOLD signal changes as a function of time, optimizing the temporal resolution is critical. Typical fMRI acquisitions with full brain coverage have repetition times (TRs) of 2–3 s (the time it takes to acquire one volume). For task-based studies, shorter TRs are usually chosen for event-related designs than for block designs, due to the relative lack of experimental power and greater importance of time-course information. Shorter TRs may lead to a significant reduction in SNR while longer TRs are theoretically associated with higher sensitivity to motion (Filippi, 2009; Wald, 2012; Craddock et al., 2013). Due to the necessity of optimizing temporal measurements, spatial resolution is usually sacrificed. With high-field strengths and/or if full brain coverage is not mandatory for the specific study, the TR can be made as low as 1 s, or even less. One way of increasing temporal resolution while still maintaining full brain coverage is to use a parallel imaging method, such as GRAPPA (Griswold et al., 2002), SENSE (Pruessmann et al., 1999), or multiplex-EPI (Feinberg et al., 2010). GRAPPA and SENSE work by reducing the time required for acquiring a single slice but increasing the sensitivity to motion. Thus, extra care should be taken, especially with participants prone to move a lot during scanning. On the other hand, multiplexed-EPI works by simultaneously acquiring more than one slice at a time (Feinberg et al., 2002). However, the simultaneous excitation of slices causes signal leaking from one slice to the other, which increases with the number of slices acquired simultaneously (i.e., the acceleration factor) and

also induces artifactual thermal noise correlations, critical for functional connectivity studies (Setsompop et al., 2013). The combination of both techniques can also be employed, further reducing the acquisition time and with revealed increased sensitivity to detect RSNs at moderate acceleration factors (Preibisch et al., 2015). Isotropic voxels are recommended (in-plane resolution and slice thickness with equal dimensions) because the folded cortex has no dominant orientation. At 3 T fields, typical voxel sizes range between 2.8 and 3.5 mm<sup>3</sup> (Wald, 2012; Craddock et al., 2013). Higher spatial resolution can be achieved at higher field strengths, but is associated with increased artifact (Olman and Yacoub, 2011). A square Field of View (FOV) ranging between 192 and 224 mm, with a matrix size of 64 and slice number of 30–36, is common at 3T. The most critical parameter when optimizing an fMRI protocol with respect to timing is the interval between slice excitation and signal acquisition, known as echo time (TE). The interval choice of TE in order to maximize the BOLD contrast depends on the tissue characteristics and the field strength and is ideally equal to the apparent tissue T2\*. The TE for 3 T field strength is typically around 30 ms (ranging from 25 to 40 ms) (Gorno-Tempini et al., 2002; Craddock et al., 2013; Murphy et al., 2013). The appropriate flip angle also is of relevance when optimizing the BOLD signal. One recommended practice is to select a flip angle equal to the Ernst angle (Ernst and Anderson, 1966) for gray matter. More recently, however, it has been shown that the use of much lower flip angles is possible, as long as physiological noise is the dominant noise source in fMRI time-series (Gonzalez-Castillo et al., 2011). For field strength of 1.5 T and a TR of 3 s, the Ernst angle is ~89°, resulting in the common choice of 90° for flip angle. For 3 T and a TR of 2 s, the angle is closer to 77° (Ernst and Anderson, 1966). These specifications are even more complex when a multicenter study is planned, and a number of considerations need to be taken into account in order to maximize reproducibility (Stöcker et al., 2005; Friedman and Glover, 2006; Glover et al., 2012; Keator et al., 2016).

Some important tips include: for studies involving both resting state and task-based fMRI, it is recommended that the same acquisition protocol be used, or at least, as similar as possible, in order to most accurately integrate and compare results (Ganger et al., 2015; Pernet et al., 2016); when performing task-based studies, it is of upmost importance to precisely synchronize scan acquisition with stimulus presentation. Such synchronization can be achieved through the use of manual configurations (e.g., sending triggers between the scanner and stimulus presentation software) or with integrated solutions such as the Lumina Controller (<http://cedrus.com/lumina/controller/>), SyncBox (<http://www.nordicneurolab.com/products/SyncBox.html>), SensaVue fMRI (<http://www.invivocorp.com/solutions/neurological-solutions/sensavue/>), or nordic fMRI solution (<http://www.nordicneurolab.com/products/fMRISolution.html>).

## Artifacts

The primary goal of any fMRI acquisition is to obtain the highest possible SNR and contrast-to-noise ratio (CNR) (Welvaert and Rosseel, 2013) while minimizing the impact of artifacts. The

artifacts in fMRI are usually related to the pulse sequence, gradient system hardware, acquisition strategy used as well as physiological noise. Three artifacts are characteristic of the traditional EPI pulse sequence: spatial distortions (**Figure 1G1**), signal dropouts (**Figure 1G2**), and ghosting (**Figure 1G3**). Geometric and intensity spatial distortions may result from static field inhomogeneity and appear locally either as stretched or compressed pixels along the phase-encoding axis, being worse at higher field strengths. A number of strategies have been suggested to correct the distortions, and include the use of shimming coils (Reese et al., 1995; Balteau et al., 2010), field mapping (Hutton et al., 2002; Zeng and Constable, 2002), point spread function mapping, or reversed phase gradients (Holland et al., 2010; In et al., 2015). Signal dropouts due to field inhomogeneities near air/tissue interfaces, particularly prevalent in the frontal and temporal lobes, also occur in EPI. The choice of an appropriate echo time (TE, described below; optimum BOLD contrast occurs when the TE matches the local T2\* of the tissue of interest), greater number of thinner (rather than lower number of thicker) slices, as well as optimizing slice tilt, the direction of the phase-encoding or the z-shim moment may all help to reduce these dropouts (Weiskopf et al., 2006; Balteau et al., 2010). Ghosting artifacts, which occur only in the phase-encoding direction, are triggered because odd and even lines of k-space are acquired with opposite polarity. Techniques such as implementing a multi-echo reference scan, two-dimensional phase correction or applying dual-polarity generalized autocalibrating partially parallel acquisitions (GRAPPA), can reduce the magnitude of these effects (Schmithorst et al., 2001; Chen and Wyrwicz, 2004; Robinson et al., 2013; Hoge and Polimeni, 2015). Hardware-related artifacts such as scanner and head coil heterogeneities, spiking, chemical shifts, and radiofrequency (RF) interferences all can significantly impact the fMRI image quality and compromise results (Bernstein et al., 2006; Poldrack et al., 2011). One approach for reducing the impact of these artifacts is to implement an Independent Component Analysis (ICA) or Robust Principle Component Analysis (RPCA) (Behzadi et al., 2007; Griffanti et al., 2014; Campbell-Washburn et al., 2016). Although hardware-related artifacts can, at least theoretically be fixed, participant related confounds will always be present. Participant's physiological confounds such as head motion (Power et al., 2012), cardiac, and respiratory "noise" as well as vascular effects all have a significant impact on the final fMRI results (Faro and Mohamed, 2010; Murphy et al., 2013). The most common and critical artifact in fMRI is head motion. Even though it is common to correct for subject motion during preprocessing (see preprocessing section), the best approach is to prevent motion as much as possible in the first place using comfortable padding and optimized head fixation (Edward et al., 2000; Heim et al., 2006), as well as to fully inform the subject in advance about scanner noise and the confining environment. Performing multi-echo acquisitions can also help reduce motion artifacts (Kundu et al., 2013). Cardiac pulsation and the respiratory cycle can have an impact similar to that of head motion. Due to the long repetition time (TR, see below) of standard BOLD EPI acquisitions (2–3 s) the fluctuations are aliased into low-frequency signals which may be mistaken

for neural activity-related BOLD oscillations, especially on rs-fMRI (Birn, 2012; Murphy et al., 2013; Cordes et al., 2014). A number of strategies have been used in an attempt to reduce these artifacts, and include the use of band-stop filtering, dynamic retrospective filtering (Särkkä et al., 2012), image-based methods (RETROICOR; Glover et al., 2000), corrections based on canonical correlation analysis (Churchill et al., 2012c) and through the use of externally recorded cardiac and respiratory waveforms as regressors (Falahpour et al., 2013).

Thorough understanding of the link between neural activity and the hemodynamic changes that give rise to the BOLD signal (neurovascular coupling), as well as the variation in its response, should help to reduce the inter-subject variability and increase the homogeneity and statistical power of the studies (D'Esposito et al., 2003; Handwerker et al., 2012; Liu, 2013; Phillips et al., 2016). One key feature is that as the signal increases (field strength, array coils), the physiological noise increases proportionally (Triantafyllou et al., 2005, 2006; Hutton et al., 2011). A great variety of software tools have been developed to minimize the impact of artifacts, for example the Artifact detection Tool (ART—[http://www.nitrc.org/projects/artifact\\_detect/](http://www.nitrc.org/projects/artifact_detect/)), the Physiological Artifact Removal Tool (PART—<http://www.mccauslandcenter.sc.edu/CRNL/tools/part>), the PhysIO Toolbox (<http://www.translationalneuromodeling.org/tnu-checkphysretroicor-toolbox/>), the ArtRepair Software (<http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html>), the FMRIB's-based Xnoisifier (FIX) (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIX>), and the RobustWLS Toolbox (<http://www.icn.ucl.ac.uk/motorcontrol/imaging/robustWLS.html>) (Diedrichsen and Shadmehr, 2005). While a significant problem in task-based fMRI, artifact identification and removal is even more complex with rs-fMRI. In the absence of an *a priori* hypothesis, it may be hard to distinguish the signal related to neural activity from the sources of noise, particularly when the artifacts are spatially or temporally correlated and may share a degree of spatial or spectral overlap with the RSNs. Whenever artifacts cannot be corrected, it may be necessary to adopt some alternative strategies such as the exclusion of the affected subject, volume or slice, or to limit the analysis to regions without significant artifacts.

## QUALITY CONTROL AND PREPROCESSING

Quality control and preprocessing procedures are key steps in the detection and correction of artifacts in fMRI, thus providing consistency and reliability to maps of functional activation. A variety of automated preprocessing pipelines have been described and implemented [e.g., DPABI, LONI (Rex et al., 2003), Nipype (Gorgolewski et al., 2011), BrainCAT (Marques et al., 2013) and C-PAC], but there is a lack of consensus about which workflow is the most effective. Several studies and reviews have explored the effects of preprocessing techniques on both task-based (Strother, 2006; Churchill et al., 2012a,b) and rs-fMRI results (Aurich et al., 2015; Magalhães et al., 2015). Herein we attempt to provide a practical guide to the most commonly used methodologies.

## Acquisition Quality Control and Data Conversion

The first quality control point comes during the acquisition phase. It is important to loop through the images using real-time display of the scanner, while it is still possible to repeat the acquisition and not lose data. Assessing the images using two different contrast settings, standard anatomical (to verify the appearance of the brain, gross head motion and spiking) and background noise contrast (to verify hardware issues and important small motion) is a wise strategy (**Figure 1H**). Following data acquisition, it is important to verify that all images have been imported and sorted correctly, and to ensure the same acquisition protocol has been used for all study participants. At this point, inspection of the scans to screen for obvious brain lesions (except for those specifically being studied) as well as visible artifacts can be performed using general-purpose viewers, such as Osirix, MRICro, RadiAnt, or ImageJ (Escott and Rubinstein, 2003; Rosset et al., 2004). Due to the absence of a standard file format, it is necessary to start by converting the original scanner data from DICOM format (Mildenberger et al., 2002; Liao et al., 2008; Mustra et al., 2008) to the most common file format used by fMRI preprocessing tools, the NIfTI format (Neuroimaging Informatics Technology Initiative, allows both separated \*.img and \*.hdr files or both combined on a single \*.nii file) (Poldrack et al., 2011), which is an extension from the Analyze 7.5 format (set of two files: \*.img containing the binary image data and \*.hdr with the metadata) (**Figure 1I**). In the NIfTI format most of the DICOM header information is discarded (e.g., patient information) and only basic acquisition information (e.g., TR, resolution, FoV, image orientation) is kept. Most of the fMRI processing packages include file converting tools, and several dedicated converters also are available [e.g., dcm2nii (<https://www.nitrc.org/plugins/mwiki/index.php/dcm2nii:MainPage>), MRIConvert (<https://lcni.uoregon.edu/downloads/mriconvert/mriconvert-and-mcconverter>) and NiBabel (<http://nipy.org/nibabel/index.html>)].

## Initial Stabilization, Slice-Timing, and Motion Correction

Upon beginning an acquisition, the scanner typically takes some seconds to completely stabilize its gradients, and the tissue being imaged requires some time to reach the necessary excitation. To remove the influence of these factors, it is common to discard the initial volumes (usually around the 10 initial seconds) of the fMRI acquisitions whether for task-based or rs-fMRI. Because fMRI volumes are acquired as 2D images, one slice at a time, and even though short and fixed TRs are utilized, there is an intrinsic delay between the real and the expected slice acquisition times, which may substantially decrease the ability to discern a given effect. The interval between the first and the last acquisition slice depends on the TR selected. Slice timing correction adjusts the time-course of voxel data in each slice to account for these differences by interpolating the information in each slice to match the timing of a reference slice (first or mean TR slice) (Calhoun et al., 2000; Sladky et al., 2011) (**Figure 1J**). The impact of using slice-time correction is described as quite variable,

depending on the type of study, ranging from very important for event-related designs (especially for time-course analysis), to less important for block designs, to having minimal effect on rs-fMRI. However, it seems that it never has a negative impact on the results (Henson et al., 1999; Sladky et al., 2011; Wu et al., 2011). In addition to the debate about whether or not to employ slice timing correction is the issue of, if used, when such correction ought be done, as this step can interact strongly with motion correction (described below). Common suggestions include: for interleaved acquisitions, it is usually performed before motion correction and for sequential acquisitions thereafter; for subjects with low head motion performed before motion correction and with high head motion after (it is recommended to keep the order consistent for all the study subjects) (Sladky et al., 2011). Nevertheless, the issue remains poorly addressed as slice timing and motion correction are two inextricably linked steps (Bannister et al., 2007). An alternative option is to perform slice timing and motion simultaneously through 4d realignments using the Nypipe 4d realignment function (Roche, 2011) or with the Seshamani data reconstruction framework (Seshamani et al., 2016). Additional methods exist for slice timing adjustment, such as adding regressors as nuisance variables (Henson et al., 1999) or altering the model rather than the data, as in dynamic causal modeling (DCM) (Kiebel et al., 2007), though that specific approach is not suitable for interleaved acquisitions.

Head motion during scanning is probably the most common and critical confound for both task and rs-fMRI studies, both of which are dependent upon precise spatial correspondence between voxels and anatomical areas over time (Friston et al., 1996; Satterthwaite et al., 2012; Maclaren et al., 2013; Zeng et al., 2014; Power et al., 2015). The most common strategy used to perform motion correction is first to realign each volume to a reference volume (mean image, first, or last volume) using a rigid body transformation (x, y, and z rotations and translations) (Jiang et al., 1995) (**Figure 1K**). While there is no standard rule about the motion threshold to be used, it is a rule of thumb to discard data sets with motion greater than the dimensions of a single voxel (Formisano et al., 2005; Johnstone et al., 2006). Because most traditional realignment strategies take into account each volume at a single point in time, and due to the fact that residual motion-induced fluctuations still are present in the data set and decrease the reliability and statistical sensitivity of the study, a different strategy was proposed. This technique was to include in the subject-level general linear model (GLM) the motion parameters estimated during the realignment step as “nuisance variables” (covariates of no interest), possibly also including the temporal derivatives of those variables (Friston et al., 1996; Johnstone et al., 2006; Power et al., 2012). Most of the commonly used fMRI packages include motion correction tools, and significant differences in their performance are not evident (Oakes et al., 2005; Morgan et al., 2007). Several groups have recently demonstrated that small head motion produces spurious but structured noise, which then triggers distance-dependent changes in signal correlations (Power et al., 2012, 2015; Satterthwaite et al., 2012; Van Dijk et al., 2012; Siegel et al., 2014). The method proposed to reduce these effects has been called Scrubbing, and is based on two measures

to capture the head displacements, Framewise Displacement (FD) or the brain-wide BOLD signal displacements (temporal Derivative VARiance—DVARS) derived from volume to volume measurements over all brain voxels (Power et al., 2012). After FD or DVARS calculation, a threshold is applied and, despite a lack of standardization, it is common to use  $FD > 0.2\text{--}1 \text{ mm}$  and  $DVARS > 0.3\text{--}0.5\%$  of BOLD signal in order to identify outliers. Scrubbing corrections can be implemented with several tools including the C-PAC, Artifact Detection Tools (Mazaika et al., 2007), DPARSF (Yan and Zang, 2010) and *fsl\_motion\_outliers* tool. By default, *fsl\_motion\_outliers* detects outliers if FD or DVARS exceeds the 75th percentile + 1.5 times the InterQuartile Range. The identified outliers are commonly regressed out later in the preprocessing pipeline (but before temporal filtering) from the data with a GLM where each outlier is entered as a nuisance regressor. Additional alternative motion correction strategies are available, such as the use of slice derived information (Beall and Lowe, 2014), task associated motion (Artifact Detection Tool), expansion to 24–36 motion regressors (Power et al., 2014), independent component analysis de-noising (Mowinckel et al., 2012; Griffanti et al., 2014; Pruij et al., 2015), and group-level motion covariates (Van Dijk et al., 2012). Furthermore, the use of non-gray matter nuisance signals (Behzadi et al., 2007; Jo et al., 2013) and regression of global signal (Power et al., 2014) have been shown to help reducing the impact of motion.

## Spatial Transformations

Performing spatial transformations to align the images from the individual's native space with those acquired from a different modality or subject [(co-)registration] or into a common standard space (normalization) is a fundamental step of the fMRI preprocessing (Brett et al., 2002) (**Figure 1L**). If homologous brain regions are not properly aligned between individuals, sensitivity is lost and leads to an increase in the false negatives rate. On the other hand, systematic normalization errors between groups may trigger false positive activations. In fMRI studies there are two main standard coordinate systems which have been used in order to reduce intersubject variability and to facilitate the reporting of results in the form of standard stereotactic (x,y,z) coordinates. These are the Talairach space, where the principal axis corresponds to the anterior commissure-posterior commissure (AC-PC) line, and which is based upon the brain of a single individual (Talairach and Tournoux, 1988), and the Montreal Neurological Institute (MNI) templates (there are several MNI templates available, being the MNI152 the most commonly used), which are based on the average of T1-weighted MRI scans of a large number of subjects (Mazziotta et al., 1995, 2001). These templates normally are associated with an atlas (Cabezas et al., 2011; Evans et al., 2012) and allow the localization of designated anatomical features in coordinate space, as well as the association of functional results to identified anatomical regions. The Automated Anatomic Labelling (AAL) (Tzourio-Mazoyer et al., 2002), the Talairach atlas (Lancaster et al., 2000), and the Harvard-Oxford atlas (Desikan et al., 2006) are amongst the most commonly used. It is important to note that the Talairach and MNI coordinates do not refer to the same

brain regions or structures (Laird et al., 2010), and it is frequently necessary to convert between the two (e.g., for meta-analyses). Available tools to implement transformation between the two coordinate spaces include the "icbm2tal" (Lancaster et al., 2007; Laird et al., 2010) (GingerALE, <http://www.brainmap.org/icbm2tal/>) and the "mni2tal" (Brett et al., 2002) (BioImage Suite, <http://bioimagesuite.yale.edu/mni2tal/>). Tools also are available to localize and label brain regions according to the MNI (MRIcron, <http://www.mccauslandcenter.sc.edu/mricro/mricron/>, Neurosynth, <http://neurosynth.org/>) or Talairach (Talairach software, <http://www.talairach.org/>, WFU\_PickAtlas, [http://www.nitrc.org/projects/wfu\\_pickatlas/](http://www.nitrc.org/projects/wfu_pickatlas/)) coordinates. Normalization strategies rely on optimization functions which maximize the similarity between two images (Jenkinson and Smith, 2001) by applying translations, rotations, and scaling in multiple axes. Transformations are usually divided into two subtypes: linear, applied uniformly along an axis and usually represented as affine matrices, and non-linear, defined locally (meaning that different points along an axis undergo unique transformations) usually defined by warp or distortion maps. Several deformation algorithms are available which can be applied to MRI registrations (Klein et al., 2009). An alternative registration method is the use of surface registration techniques, in which the functional time series are mapped onto cortical surface models [e.g., automatically implemented by Freesurfer (Fischl et al., 1999)], improving the computational efficiency and the mapping of the cortical surface, beneficial for subsequent processing and analysis steps (surface-based smoothing kernels and surface-registration can be used) (Klein et al., 2010; Khan et al., 2011).

In fMRI there are two commonly used processing streams for spatial normalization. In one, a single step strategy is used to normalize directly to a standard EPI template, while the other employs a multi-step method which first aligns to the matching structural image using rigid-body or affine transformations, following which the composite image is then registered to the reference space, using either affine or non-linear transformations (Poldrack et al., 2011). Complementary techniques for removing non-brain areas from the analysis and reducing the data size, such as skull striping or masking, may also help to improve the normalization step (Tsang et al., 2007; Andersen et al., 2010; Fischmeister et al., 2013). The choice of the optimal atlas template and mapping function depends on a multitude of factors and is influenced by age, gender, hemispheric asymmetry, normalization methodology, and disease-specificity (Crinion et al., 2007). Following the normalization step, it is always important to perform visual quality control, for example by displaying the fMRI data of each participant along with a reference EPI template.

## Spatial Smoothing and Filtering

The next preprocessing step normally implemented is that of spatial smoothing/filtering, a process during which data points are averaged with their neighbors, suppressing high frequency signal while enhancing low frequency ones, and results in the blurring of sharp edges (**Figure 1M**). Smoothing simultaneously increases the SNR and the validity of the statistical tests (from

random field theory) by providing a better fit to expected assumptions while reducing the anatomical differences. On the other hand, smoothing reduces the effective spatial resolution, may displace activation peaks (Reimold et al., 2006) and extinguish small but meaningful local activations depending on the filter parameters chosen (Yue et al., 2010; Poldrack et al., 2011; Sacchet and Knutson, 2013). The standard spatial smoothing procedure consists of convolving the fMRI signal with a Gaussian function of a specific width (as, spatially, the BOLD signal is expected to follow a Gaussian distribution). The choice of the proper size of the Gaussian kernel [Full Width at Half Maximum (FWHM)], which determines the extent to which the data is smoothed, will be dependent on specific features of the study undertaken, such as type of paradigm and inference expected, as well as on the primary image resolution. The amount of smoothing always should be the minimum necessary to achieve the intended results, and a reasonable starting point is a FWHM of twice the voxel dimension (care must be taken when using large smoothing kernels as they make the detection of smaller patterns of activation harder). The typical smoothing values used range between 5 and 10 mm for group analyses (Beckmann and Smith, 2004; Mikl et al., 2008; Poldrack et al., 2011). Alternative approaches to smoothing are the use of varying kernel widths (Worsley et al., 1996), adaptive smoothing (Yue et al., 2010; Bartés-Serrallonga et al., 2015), wavelet transforms (Van De Ville et al., 2007), and prolate spheroidal wave functions (Lindquist et al., 2006). Despite its common use, care must be taken when performing smoothing due to its effects on the final results (Geissler et al., 2005; Molloy et al., 2014), its interaction with motion correction (Scheinost et al., 2014) and impact upon analyses which are sensitive to the activation of individual voxels (such as ROI-to-ROI analysis, Regional Homogeneity and Multi-voxel Pattern Analysis). This step is not recommended for connectomic approaches in order to prevent the BOLD signal from extending across different regions of interest (Zuo et al., 2012; Tomasi et al., 2016).

A final step in the data preprocessing pipeline is temporal filtering (**Figure 1N**). This step is performed in order to remove the effects of confounding signals with known or expected frequencies. The use of frequency filtering (and/or spatial smoothing) may help attenuate noise and thus increase the SNR (White et al., 2001). Functional MRI time-courses often manifest low-frequency drifts which may reduce substantially the statistical power of the results. It is therefore of great relevance to attempt to identify which frequencies are those of interest and which are noise (Kruggel et al., 1999). For example, fMRI noise may be associated with slow scanner drifts ( $\sim <0.01$  Hz), as well as cardiac ( $\sim 0.15$  Hz) and respiratory ( $\sim 0.3$  Hz) effects (Cordes et al., 2001, 2014). The most frequently used filters for task-based fMRI acquisitions are high-pass filters (typically  $\sim 0.008\text{--}0.01$  Hz, 100–128 s), generally deployed with a rough rule of using a cut-off value at least 2 times that of the fundamental task-frequency (the interval between one trial start and the next one). With rs-fMRI the standard strategy is to apply a band-pass filter (0.01–0.08 Hz) following the reports of Biswal and colleagues (among others), which have shown that spontaneous BOLD

low frequency ( $\sim <0.1$  Hz) fluctuations were physiologically meaningful and reflect spontaneous neural activity (Biswal et al., 1995; Fransson, 2005; Shirer et al., 2015). Nevertheless, high frequency signals ( $>0.1$  Hz) have also been shown to present functional significance (Chen and Glover, 2015; Gohel and Biswal, 2015). Exploring such frequency band requires extra caution in controlling for physiological sources of noise (e.g., respiratory and cardiac effects) as these are known to present frequencies greater than 0.1 Hz. This can be achieved using simultaneous monitoring of pulse oximetry, electrocardiogram and/or breathing belt.

Normally associated with the filtering step, detrending methods also are used to reduce the effects of noise (Tanabe et al., 2002; Friman et al., 2004). Complex spatio-temporal filters (Kriegeskorte et al., 2010), cross-validation (Ngan et al., 2000), adaptive filters (Deckers et al., 2006), matched-filter acquisitions (Kasper et al., 2014), bilateral filtering (Rydell et al., 2008), and multifiltering (Hui et al., 2013) are alternatives to the standard fMRI temporal filters.

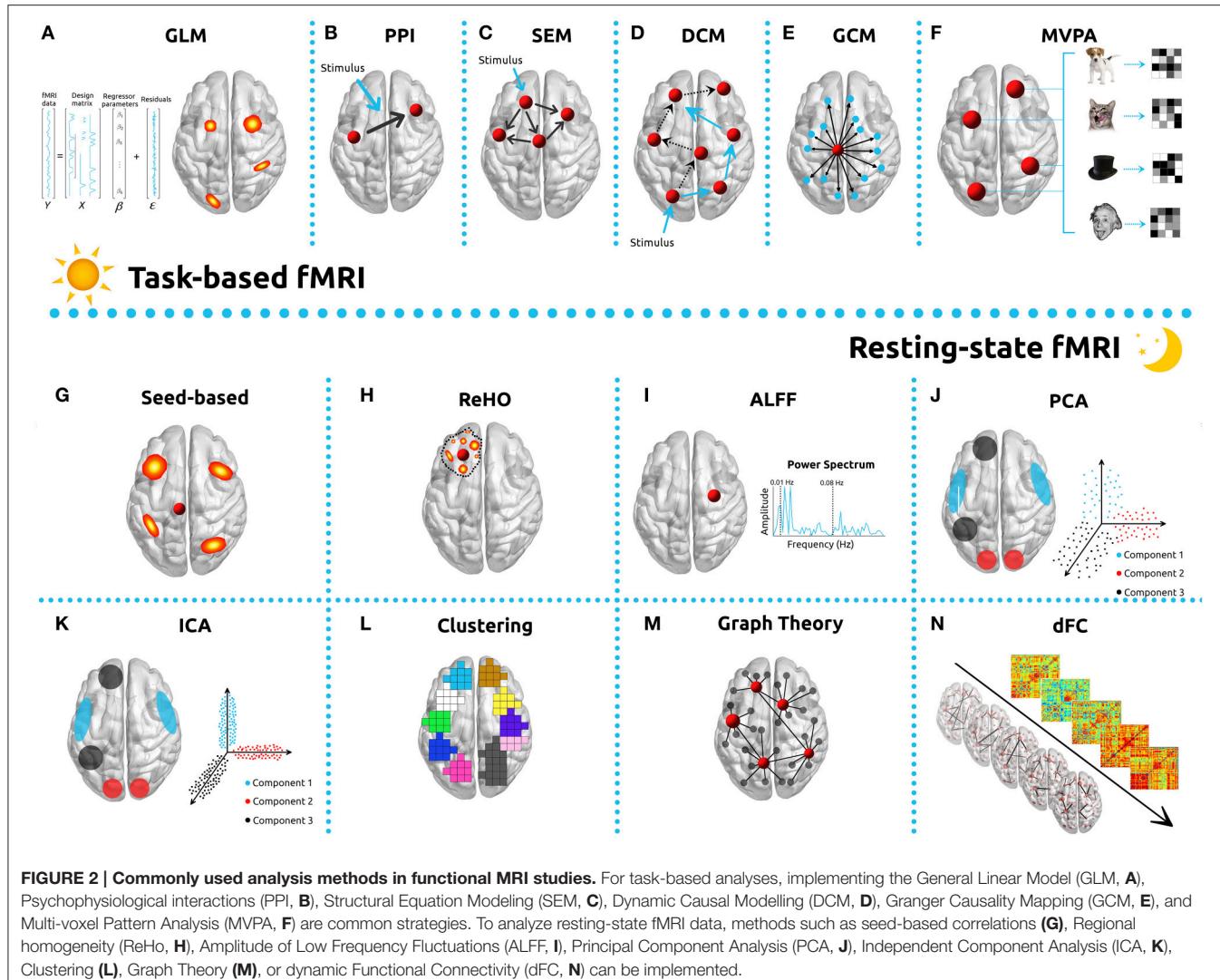
Effective quality control is of fundamental importance in the optimization of data usability reliability and reproducibility. Software tools have been developed in order to implement quality control procedures complementary to the ones already mentioned, such as BIRN QA ([http://www.nitrc.org/projects/bxh\\_xcede\\_tools/](http://www.nitrc.org/projects/bxh_xcede_tools/)), NYU CBI Data Quality tool (<http://cbi.nyu.edu/software/dataQuality.php>), and the CANLAB Diagnostic Tools (<http://wagerlab.colorado.edu/tools>).

## ANALYSIS METHODS

The next stage in the fMRI workflow is the selection of the most suitable method to extract the relevant functional information. There are many fMRI analysis methods and software tools for both task-based (**Figure 1O**) and rs-fMRI (**Figure 1P**). Thus, choosing the one most suitable for a specific study may be a complex, often confusing and time-consuming task. In order to assist with this choice, we herein present a table with the most commonly used software tools for the analysis of task-based and rs-fMRI data (**Table 4**). Some existing reviews have already explored fMRI analysis methods (van den Heuvel and Hulshoff Pol, 2010; Lohmann et al., 2013; Smith et al., 2013; Sporns, 2014; Haynes, 2015; Zhan and Yu, 2015; Pauli et al., 2016). In the following sections we distinguish between task-based and resting-state fMRI analysis according to the prominence use of each method, nevertheless, some are suitable for both fMRI acquisitions. Other distinctions could be performed, namely between methods suitable for localization and for connectivity approaches. The appropriateness application of each method will also be discussed below.

### Typical Task-Based Analyses Methods

The most employed method in the analysis of task-based fMRI is Statistical Parametric Mapping (SPM), which is based on the GLM (**Figure 2A**) (Friston et al., 1994a; Kiebel and Holmes, 2003; Poline and Brett, 2012). GLM's popularity is based on its straightforward implementation, interpretability and computability. It incorporates most data modeling structures



**FIGURE 2 | Commonly used analysis methods in functional MRI studies.** For task-based analyses, implementing the General Linear Model (GLM, **A**), Psychophysiological interactions (PPI, **B**), Structural Equation Modeling (SEM, **C**), Dynamic Causal Modelling (DCM, **D**), Granger Causality Mapping (GCM, **E**), and Multi-voxel Pattern Analysis (MVPA, **F**) are common strategies. To analyze resting-state fMRI data, methods such as seed-based correlations (**G**), Regional homogeneity (ReHo, **H**), Amplitude of Low Frequency Fluctuations (ALFF, **I**), Principal Component Analysis (PCA, **J**), Independent Component Analysis (ICA, **K**), Clustering (**L**), Graph Theory (**M**), or dynamic Functional Connectivity (dFC, **N**) can be implemented.

and provides the means for minimizing/controlling the effects of confounding factors such as motion, respiratory and cardiac and HRF derivatives (Calhoun et al., 2004; Lund et al., 2006; Bright and Murphy, 2015). One common approach in the use of this technique is to convolve the stimulus onsets and durations with a canonical HRF, which results in quantifying an estimate of the expected BOLD signal for any condition of interest. These estimates are then defined, along with intrinsic confounding factors (e.g., motion parameters), as the independent variables of the GLM. Each voxel time-series is then set as the dependent variable. The result of this process is to generate a test statistic for each voxel in the brain, which makes possible the creation of a parametric map (SPM). The process is performed separately for each subject and is commonly designated as first-level analysis. The GLM can be used very generally, ranging from the simplest subtraction method to parametric correlations with behavior, and also serves as the reference for several methods used to estimate connectivity. The main criticism of the GLM is based upon the intrinsic assumptions which must be made related to parametric

testing in general, and the GLM in particular, and which are not usually verified nor are they tested (Monti, 2011). Despite these reservations, GLM analysis remains extremely popular for fMRI.

Task based connectivity analysis is being performed with increasing frequency and its results are quite sensitive to the choice of analysis tool. For it to be used appropriately, it is necessary to distinguish undirected associations between brain regions (functional connectivity—FC) from directed and causal relationships (effective connectivity) (Horwitz et al., 2005; Friston, 2011). Functional connectivity will be discussed in greater detail below, since the methods involved are more widely used for rs-fMRI, although some of the same principles apply also to task-based analysis. Also closely related to the GLM, and concerned with effective connectivity, psychophysiological interaction (PPI) is a method used to quantify how task-specific FC between a particular brain ROI (source/seed) and the rest of the brain voxels are affected by psychophysiological variables (Figure 2B) (Friston et al., 1997; O'Reilly et al., 2012).

Some caveats when using PPI analyses are the hemodynamic deconvolution, the low power, and the difficulties intrinsic to event-related designs (Gitelman et al., 2003; O'Reilly et al., 2012).

Similar to PPI analysis in that it explores how the experimental context affects connectivity between a group of regions, the structural equation model (SEM) is used to assess the effective connectivity based on an *a priori* model of causality (**Figure 2C**) (McLntosh and Gonzalez-Lima, 1994; Büchel and Friston, 1997; Kline, 2011). It starts with the definition of a set of ROIs, and then tries to determine the connection strength between those ROIs that best fit the model. SEM allows the investigation of several brain regions simultaneously, and incorporates prior anatomical and functional knowledge to determine causal relationships, but assumes that the interactions are linear and, (similar to PPI) it cannot take into account the dynamic changes of the BOLD signal (Tomarken and Waller, 2005). Most often used for task-based fMRI, SEM has seen application with rs-fMRI (James et al., 2009).

DCM allows estimation of the effective connectivity (model states) between brain regions by determining hemodynamic response (model output) as a function of specified external experimental variables (model input) (**Figure 2D**). One of the primary characteristics of DCM is that it allows exploration of the brain as a dynamic system, accounting for changes in populations of neurons, and is able to build non-linear models of interacting regions (Friston et al., 2003; Penny et al., 2004; Stephan et al., 2008; Friston, 2009). DCM is a reliable and potentially a more biologically realistic method for fMRI in that it deals with function at the neuronal level. It does require pre-specified models and based on its non-linearity and complexity, involves the estimation of many parameters (using Bayesian estimation) and thus considerably more processing time. Each region ultimately is characterized by a single parameter (neuronal activity) (Friston et al., 2003; Frässle et al., 2015). DCM is primarily used for task-based fMRI but can also be applied to rs-fMRI analyses (Friston et al., 2014a; Razi et al., 2015; Rigoux and Daunizeau, 2015).

Another method which may be used to investigate effective connectivity is Granger Causality Mapping (GCM) (**Figure 2E**). The process is based upon determining temporal precedence in neural time-series and infers causality from time-lagged correlation (Goebel et al., 2003; Friston et al., 2013; Seth et al., 2015). GCM does not require the specification of an *a priori* model, but does have significant limitations imposed by inherent latency differences in the HRF across different brain regions, low-sampling rates and noise (Wen et al., 2013). It has been applied both to task-based (Anderson et al., 2015) and rs-fMRI (Liao et al., 2011).

Enjoying increasing popularity, Multivoxel Pattern Analysis (MVPA) uses pattern-classification algorithms (classifiers) (Haynes, 2015) in the attempt to delineate different mental states, as well as to correlate the patterns with specific perceptual, cognitive, or disease states (**Figure 2F**) (Norman et al., 2006; Mahmoudi et al., 2012; Premi et al., 2016). In contrast to the standard GLM approach (focus on patterns of activity of individual voxels), MVPA incorporates the signal from

the distributed activity or connectivity across multiple voxels simultaneously, allowing to infer mental states from patterns of distributed neural activity and the formulation of proper reverse inferences (Poldrack et al., 2009). Furthermore, enables a greater sensitivity and specificity, as well as the possibility to test hypothesis with designs that cannot be implemented in mass-univariate methods implemented with the standard GLM approach (Etzel et al., 2013). Another difference between the approaches relies on the fact that while *t*-tests model the complete set of time points, a classification trains on a subset of data (Coutanche, 2013). MVPA analyses are typically implemented using a “decoding” approach, which is based on the use of classifiers, such as neural networks (Polyn et al., 2005; Nickl-Jockschat et al., 2015), support vector machines (Meier et al., 2012; Måansson et al., 2015), and linear discriminant analysis (Cox and Savoy, 2003; Mandelkow et al., 2016), as a mean to differentiate between different classes or groups of individuals. Despite its popularity in the neuroimaging field, the “decoding” approach has some limitations, particularly related with the different results obtained with different parameters and/or algorithms. An alternative approach, the “searchlight” mapping, performs multivariate analysis on a spherical “searchlight” centered on each voxel in turn, resulting in a statistical map of local multivariate effects (Allefeld and Haynes, 2014), which can be interpreted similarly to a GLM statistics output map (Kriegeskorte et al., 2006). MVPA analyses can be applied both to task-based and rs-fMRI and, with their high sensitivity and effective use of spatial information, allow pattern detection of increasingly complex scenarios. On the other hand, the use of complex and specific classifiers may make it difficult to generalize the results of employing this technique (Dosenbach et al., 2010; Cole et al., 2013).

## Typical Resting State Analyses Methods

Historically, the first method applied to rs-fMRI was seed-based correlational analysis (**Figure 2G**) (Biswal et al., 1995). The method is based on the activity in an *a priori* defined ROI (the seed region) which may be a volume or a single voxel, which is compared to that in all other voxels in the brain (Lee et al., 2013). Seed-based analyses are characterized by simple implementation and statistics and are straightforward to interpret, but do require an *a priori* selection of ROI. Such selection can be optimized using the data itself (Golestani and Goodyear, 2011). This form of analysis is widely used for rs-fMRI (each RSN can be extracted from a specific associated ROI), but can additionally be applied to fMRI tasks (Schurz et al., 2015) and to PPI analysis, which is in principle a seed-based analysis.

Regional Homogeneity analysis (ReHo) (**Figure 2H**) uses Kendall's coefficient of concordance to measure the synchronization between the time-series of each voxel and that of its nearest neighbors (based on a pre-defined ROI) (Zang et al., 2004). The ReHo method is easy to implement and interpret, and is normally applied to rs-fMRI determinations (Zang et al., 2007; Pedersen et al., 2015). The Amplitude of Low-Frequency Fluctuations (ALFF) and more recently, the fractional ALFF (fALFF, which has reduced sensitivity

to physiological noise), measures signal magnitude on a voxel by voxel basis (**Figure 2I**) (Zou et al., 2008). ReHo and (f)ALFF both are methods which reflect properties of local spontaneous activity and, because they manifest different properties of the BOLD signal (synchronization and amplitude), they are usually implemented as complementary analyses.

In order to overcome the limitations of model-based analyses, exploratory data-driven methods, which require neither prior information nor a previously defined model, have been applied to fMRI. The three primary techniques are Principal Component Analysis (PCA), Independent Component Analysis (ICA), and clustering. PCA is a method built on finding a set of orthogonal axes (identified as principal components) that can maximize the explained variance of data and separate the relevant information from the noise (**Figure 2J**) (Wold et al., 1987; Viviani et al., 2005; Abdi and Williams, 2010; Smith et al., 2014). The efficacy of PCA is strongly dependent on assumptions of linearity, orthogonality of principal components, and high SNR. It can be applied both to task-based (Nomi et al., 2008) and rs-fMRI (Zhong et al., 2009). The method most frequently used for studies of rs-fMRI FC is ICA (an extension of PCA) (**Figure 2K**) (Jutten and Herault, 1991). This processing technique separates individual elements into their underlying components, and models the fMRI data set as a constant number of spatially or temporally independent components, which then are linearly mixed (Kiviniemi et al., 2003; Beckmann, 2012). For fMRI, ICA maps are normally generated using spatial ICA methods (spatially independent components) however temporal ICA also can be implemented and is used primarily for task fMRI. Limitations to the use of the technique in the temporal domain are its high computational demands and necessity of relying on fewer data points than studies considering spatial components (Calhoun et al., 2001). ICA generates a set of spatial maps and corresponding time-courses. The selection of components of interest is not trivial (in the absence of an *a priori* hypothesis) and is usually performed by visual inspection or correlation with a predefined RSN template. While straightforward to implement in single-subject analyses, group ICA analyses are more complex and require choosing between several different workflows and algorithm definitions (Beckmann and Smith, 2004; Calhoun et al., 2009; Schöpf et al., 2010; Du et al., 2016). ICA methods also have been used extensively in rs-fMRI studies (Beckmann et al., 2005; Soares et al., 2016), task-based fMRI (Calhoun et al., 2008), and for artifact removal (Perlberg et al., 2007; Feis et al., 2015; Pruijm et al., 2015).

The use of clustering methods constitutes a different approach based on mathematical algorithms that groups data into subsets (clusters) such that parameters of the same cluster are more similar to one another than they are to those of different clusters (**Figure 2L**). Similarly to PCA and ICA, clustering is a totally data-driven approach that enables, for example, the grouping of brain voxels with similar connectivity in the same cluster. The main difference relies on the fact that ICA assumes that there are spatially independent regions that form a network through a shared fMRI time-course, while clustering does not

rely on assumptions and simply groups voxels with similar time-courses. Clustering methods have been successfully implemented both with rs-fMRI (Mezer et al., 2009; Lee et al., 2012) and task-based fMRI (Goutte et al., 1999; Heller et al., 2006). The major challenges associated are the requirements that the spatial reproducibility of networks be optimized across subjects and that individual network homogeneity be maximized (Shams et al., 2015). Clustering can be implemented using hierarchical techniques (Cordes et al., 2002), partitional clustering (such as k-means) (Fadili et al., 2000), spectral clustering approaches (Craddock et al., 2012), or sparse geostatistical analysis (Ye et al., 2011). Despite serving similar purposes as ICA, clustering methods were shown to outperform ICA for classification purposes (Meyer-Baese et al., 2004).

An increasingly prominent and powerful tool for the study of functional brain networks is graph theory. These methods model the brain as a network comprised of nodes (voxels or regions) and edges (connections between nodes, e.g., time-series correlations). This enables the establishment of functional interactions between every possible brain region, constituting an extension of the seed-based analysis where all possible seeds are explored, also known as the functional connectome. This whole-brain network is mathematically modeled as graph and, consequently, graph-theory metrics can be used to study the topological properties of such network (**Figure 2M**). Properties such as clustering-coefficient, characteristic path length, centrality, efficiency, modularity, among others, provide insights about functional integration, segregation, resilience or organization of the network as whole or of its individual nodes (Reijneveld et al., 2007; Stam and Reijneveld, 2007; Bullmore and Sporns, 2009). The approach has been used extensively with rs-fMRI (Wang et al., 2010; Ye et al., 2015; Marques et al., 2016) and, to a lesser extent, for task-based fMRI (Cao et al., 2014), where it has been described as sometimes difficult to implement and interpret (Fornito et al., 2013). Another approach, which is somewhat more straightforward to implement, is to characterize the edges of the graph, rather than to consider the topological properties of the entire network.

In contrast to most rs-fMRI strategies, which are based on the assumption of stationarity, dynamic functional connectivity (dFC) addresses the temporal component (fluctuations) of spontaneous BOLD signals (**Figure 2N**). Dynamic FC analysis has the potential to clarify the constant changes in patterns of neural activity and may be a more appropriate choice for the analysis of rs-fMRI studies (Bassett et al., 2011; Cabral et al., 2011; Madhyastha et al., 2015; Kaiser et al., 2016). The technique can be implemented using the sliding window correlations approach (most common) (Hindriks et al., 2016), time-frequency analysis (Chang and Glover, 2010), single-volume co-activation patterns (Liu et al., 2013), repeating sequences of BOLD activity (Pan et al., 2013), or through phase synchronization (Glerean et al., 2012). A number of limitations associated with the approach include the initial steps of sliding-window specification and specificity of pre-processing, as well as its sensitivity to physiological noise and complexity of the attendant statistical analysis (Hutchison et al., 2013; Leonardi and Van De Ville, 2015; Tagliazucchi and Laufs, 2015).

## STATISTICAL ANALYSES

In a single fMRI experiment, images made up of roughly 100,000 voxels are acquired from hundreds to thousands of times, resulting in a massive data set which has a complex spatial and temporal structure (**Figure 1Q**).

### Group-Level Analyses

In order to make inferences at the group-level (i.e., second-level), the analyses of fMRI data most widely used are performed within the GLM framework. In general terms, the GLM approach models the time series of the fMRI signal as a linear combination of different signal components, in order to test whether the activity in a defined brain region is systematically associated with a particular condition of interest (Lindquist, 2008). The GLM is expressed as:

$$Y = X * \beta + \epsilon$$

where  $Y$  is the observed BOLD response,  $X$  corresponds to the design matrix,  $\beta$  is related with the parameter estimates and  $\epsilon$  is the error. Hypothesis testing in the GLM framework include a set of parametric approaches, comprising the familiar *T*-Tests (independent and paired), Multiple Regression and ANalysis Of VAriance (ANOVA) (Friston et al., 2007). Commonly, the research question leads to more complex experimental designs which involve both within-subjects (e.g., condition A vs. B) and between-subjects (e.g., control vs. experimental group) factors. More than one within-subjects factor or the analysis of between-subjects factors, cannot be performed with the traditional tools in a single model. Even though most allow the parametrization of such models, the results can be invalid due to the inherent inability of the tools to incorporate all the factors into a single model (Chen et al., 2014). An alternative approach, the GLM Flex tool (Harvard Aging Brain Study, Martinos Center, MGH, Charlestown, MA, [http://mrtools.mgh.harvard.edu/index.php/GLM\\_Flex](http://mrtools.mgh.harvard.edu/index.php/GLM_Flex)) was developed. The tool can handle multiple within- and between-subjects' factors, while also modeling all the possible interactions between factors within the same model. Parametric tests are popular due to their simplicity and ease of application. However, these tests make some strong assumptions that are minimally met, or not met at all, in fMRI data sets (e.g., assumption of normality). As a result, it is often more appropriate to use non-parametric tests. Such tests estimate the null distribution from the data itself. The most common non-parametric tests used in fMRI analysis are permutation (randomization) tests. Tools that implement such tests include randomize from FSL (Winkler et al., 2014) and SnPM (Nichols and Holmes, 2002).

### Statistical Significance

As in all standard statistical inference, the evaluation of fMRI data requires the establishment of a criterion for statistical significance. In early fMRI studies, the commonly-used standard for statistical significance was an uncorrected  $p$ -value of 0.001 at each voxel, a value that is 50 times more restrictive than that typically used in scientific research (Lieberman and Cunningham, 2009). In a typical fMRI experiment, more than

100,000 statistical tests may be performed (one test per voxel). Because this number of determinations is so great, a  $p < 0.001$  would likely produce up to 100 voxels which would be erroneously identified as significant. Such a false-positive rate would clearly be unacceptable, so a variety of methods have been proposed to cope with the multiple comparisons issue. They can be divided into two main categories: voxel-based thresholding, including the family-wise error rate (FWER) and the false-discovery rate (FDR); and cluster-extent based thresholding (Forman et al., 1995). A widely used method for voxel-based thresholding consists of using the FWER in combination with Random Field Theory (RFT). The technique is implemented by estimating the smoothness of the image, expressed in the number of resels (image resolution element), since the neighboring voxels share statistical dependency. Although it can be thought of as roughly similar to the Bonferroni correction, FWER control using the RFT approach has a number of unique attributes and limitations due to the inherent smoothness of fMRI data. While enabling a great control over type I error, it often is over-conservative and may prevent true results from being detected (Hayasaka and Nichols, 2004). The FDR approach, another popular technique to control false-positives in neuroimaging studies (Genovese et al., 2002), considers the proportion of false positives in all the rejected tests. FDR control is less stringent than FWER and usually results in increased power. Because this approach is applied to  $p$ -values (rather than to the test statistics themselves) it can be used with any valid statistical test, but is highly dependent on the sample size. The most widely used FDR approach to functional imaging data is the Benjamini-Hochberg (BH) procedure, which assumes independence between tests (Benjamini and Hochberg, 1995). Statistical tests in fMRI are known to be dependent, however, so concern has been raised regarding its applicability (Chumbley and Friston, 2009; Chumbley et al., 2010). Most common software tools, specifically AFNI, FSL, and SPM, implement this type of correction method.

A significant problem associated with conservative approaches is the increased probability of committing type II errors (failure to detect true effects), particularly evident with small samples (Nichols and Hayasaka, 2003). It has also been postulated that this approach may favor the extraction of more obvious effects (such as sensorimotor processes), associated with signals of large magnitude. While failing to capture more subtle phenomena (such as complex cognitive and affective processes) often associated with signals of low amplitude (Lieberman and Cunningham, 2009), cluster-extent based thresholding has been put forward in order to address some of these shortcomings. It detects significant clusters based on the number of contiguous voxels that surpass a pre-determined primary threshold (Friston et al., 1994b). The main rationale for its use is that adjacent voxels are more likely to be involved in the same neuronal processes and thus are not independent (Smith and Nichols, 2009). The net result is that instead of estimating the false positive probability of each voxel, this approach estimates the false positive probability of the region as a whole (Woo et al., 2014). The cluster size is determined from the sampling distribution of the largest null cluster size under the null hypothesis of no signal. The

reasoning behind this correction is based on the observation that false-positives are randomly distributed and thus are not likely to occur in contiguous groups of voxels (Woo et al., 2014). Cluster-extent approaches also are associated with reduced spatial specificity, describing the likelihood of finding a cluster of a given size or greater under the null hypothesis. The implication is that the larger the cluster, the less spatially specific the inference, though this is often an overlooked aspect of functional imaging (Woo et al., 2014). The most well-known cluster-size estimation methods are based on RFT as implemented in SPM, or on Monte Carlo simulations such as AlphaSim distributed with AFNI and with the REST toolbox. All these methods require the definition of an arbitrary primary cluster-defining threshold. An alternative method, termed threshold-free cluster enhancement (TFCE), was developed in order to eliminate the need for the definition of the primary threshold and is implemented in FSL (Smith and Nichols, 2009), CAT toolbox (<http://dbm.neuro.uni-jena.de/cat/>), and MatlabTFCE (<https://github.com/markallenthornton/MatlabTFCE>).

Yet another method of performing fMRI statistical analyses is through the use of specified ROIs. Analyses of this type are usually performed when the researcher has some *a priori* hypothesis regarding a specific brain region, which renders the previously discussed corrections for multiple comparisons too restrictive (see Poldrack, 2007 for other rationales). Generally, ROI analyses lead to an increased sensitivity (signal is average across groups of voxels) but a false sense of specificity of a given activation (activity patterns in regions outside the ROIs are masked out). The simplest approach consists of averaging the estimates over the voxels from the ROI and then performing the statistical testing with the averaged estimate. An alternative method, commonly named Small-Volume Correction (SVC), consists of restricting the voxel-wise analysis to the voxels inside the ROI, thus reducing the number of tests required to account for multiple comparisons corrections. Most software tools, such as SPM, FSL, and AFNI, contain routines for ROI-based analysis. The Marsbar tool (<http://marsbar.sourceforge.net>) for SPM was specifically developed for this purpose.

## Effect Sizes

Contrary to the standard practice in other research areas, effect estimates (i.e., the effects' magnitude) are usually not provided in most neuroimaging reports. A recent publication highlighted that the statistic value does not provide information regarding the actual significance of the findings, serving rather as auxiliary evidence for the existence of the targeted effect. On the contrary, the effect estimate provides a clear picture of the property of interest and, consequently should be the focus of the investigation. For this reason, the absence or misreporting of effect-sizes has direct implication on the reliability and interpretability of fMRI findings (Chen et al., 2016). Taking this into consideration, it is strongly recommended that effect size maps/images are made available. With this practice, the whole range of effects and not only significant findings can be used to compare and properly aggregate effect sizes across different studies/research centers, and also allowing the use of power analysis in future studies (Poldrack et al., 2008).

## Meta-Analysis

The number of fMRI publications continues to grow exponentially, but the results are often not consistent across studies (Radua and Mataix-Cols, 2012). Therefore, the meta-analysis of functional imaging studies may be essential for the continued development of new hypothesis about the neural mechanisms of cognition, emotion, and social processes (Wager et al., 2007). Individual studies generally provide evidence about brain activity rather than mental states, whether meta-analyses can help to identify consistently activated regions related to the same psychological state (Wager et al., 2007). Neuroimaging meta-analysis pools statistically significant results and offers the potential to improve predictive power, to build analytic tools and models, and to detect emergent properties of neural systems through large-scale data mining and computational modeling (Fox P. T. et al., 2014). The methods work by counting the number of activation peaks in each local brain area and comparing the observed number of peaks to a null-hypothesis distribution in order to establish a criterion for significance. Functional MRI meta-analysis can be performed using either full statistical parametric maps—image-based meta-analysis (IBMA)—or the coordinates of significant findings—coordinate-based meta-analysis (CBMA). Whereas, IBMA captures consistent patterns of brain activation across studies, even though these patterns are not identified as significant in individual studies, neuroimaging studies rarely provide full statistical parametric maps, which preclude these analyses. Thus, the majority of analysis aggregating neuroimaging results relies on CBMA, in which each eligible study included, reports using standard atlas or template based, 3-dimensional locations of peak activations. As a result, CBMA only aggregate results that are reported as significant across studies, and fail to capture individually non-significant, but consistent findings across different studies. A number of different algorithms have been developed for CMBA analyses, including the Activation Likelihood Estimation (ALE) (Eickhoff et al., 2012), Kernel Density Analysis (KDA) (Wager et al., 2004), Multi-level Kernel Density Analysis (MKDA) (Wager et al., 2007), and the Effect-size Signed-Differential-Mapping (ES-SDM) (Radua et al., 2012).

## MULTIMODAL STUDIES

Collecting multimodal brain data using different neuroimaging methods has become increasingly popular and is definitely a future trend, which provides an opportunity to develop a more global description of brain structure and function (Figure 1R). A number of different modalities and techniques have been used to complement fMRI analysis, either simultaneously or separated in time, and have been reviewed elsewhere (Biessmann et al., 2011; Uludag and Roebroek, 2014; Liu et al., 2015a; Garcés et al., 2016). One particularly powerful approach to better understanding the brain is to model it as a network of functional connections between every possible region. The connectomic paradigm provides the investigator with an effective framework with which to study how dynamic changes in function are

related to structural change, and how both are connected with brain states. Several extensive studies and worldwide projects [e.g., Human Connectome Project (Van Essen et al., 2013), Developing Human Connectome Project (<http://www.developingconnectome.org/>), Baby Connectome Project (<http://www.fnih.org/what-we-do/current-research-programs/baby-connectome>), or MyConnectome project (<http://myconnectome.org/wp/>)] are currently under way and have been enhancing multimodal approaches by combining fMRI data with structural information (e.g., diffusion data, volumetric data, cortical thickness, and voxel based morphometry) (Labudda et al., 2012; Crossley et al., 2014; Horn et al., 2014; Frank et al., 2016). Another approach employing complementary methodology is the combination of fMRI with the recording of brain electrical activity (electrophysiological response) using either electroencephalography (EEG) or magnetoencephalography (MEG) (Bledowski et al., 2004; Vaudano et al., 2012; Tewarie et al., 2015). Both techniques add improved temporal resolution to the very good spatial resolution of fMRI (Huster et al., 2012; Hall et al., 2014; Jorge et al., 2014). Positron emission tomography (PET) and single-photon emission computerized tomography (SPECT) both have a long history of providing fundamental information regarding brain metabolism. Though lacking the time resolution of fMRI, they complement that methodology by having the ability to study such parameters as neurotransmitter-receptor interactions and local glucose metabolism for longer periods in time (minutes) (Price, 2012; Sander et al., 2013; Tousseyen et al., 2015). It is possible to perform fMRI and functional near-infrared spectroscopy (fNIRS) simultaneously, and such a multimodal approach may be used to improve the temporal resolution of the former, thus allowing better correlation of the BOLD signal with local hemodynamic changes (Steinbrink et al., 2006; Sato et al., 2013). Inducing small direct currents in the brain using transcranial magnetic stimulation (TMS) or transcranial direct current stimulation (tDCS), make possible relatively focal excitation or inhibition and, when performed concurrently with fMRI, allows the study of functional interactions (Ruff et al., 2009; Peters et al., 2013; Weber et al., 2014; Leitão et al., 2015). The rapid growth of multimodal neuroimaging techniques has triggered the parallel development of computing methods and workflows capable of analyzing the resultant complex data sets (for review Liu et al., 2015b), and has led to the development of several tools dedicated to this type of study (Casanova et al., 2007; McFarquhar et al., 2016).

While the primary focus of this guide has been that of human neuroimaging, it is useful to note that many of the concepts and strategies described also can be applied to animal experimentation. The availability of ultra-high field scanners, capable of achieving very high resolution, has made feasible the application of fMRI to brains as small as that of a mouse (Jonckers et al., 2011; Schlegel et al., 2015). Other animals studied using this technique are rats (Liang et al., 2012; Henckens et al., 2015), non-human primates (Hutchison et al., 2015; Petkov et al., 2015), dogs (Andics et al., 2014; Berns et al., 2015), and cats (Brown et al., 2013; Hall et al., 2016). Translational research opportunities allow the investigator to develop animal models for studies which cannot be undertaken in patients or volunteers.

A number of technical issues which must be considered when designing protocols for animal work are: the impact of higher magnetic fields and the ability to detect functional contrasts (Ciobanu et al., 2015); the use, or not, of anesthesia or sedation and its effects on regional and global brain activity (Kalthoff et al., 2013; Schlegel et al., 2015); physiological differences between animals and humans (Kalthoff et al., 2011; Sumiyoshi et al., 2012); and the fact that relatively few reference templates and atlases are available for animals (Stoewer et al., 2012; Nie et al., 2013; Papp et al., 2014).

## REPORT AND INTERPRETATION OF RESULTS

The results reported for a typical fMRI study include such information as the peak cluster coordinates (in x, y, and z), cluster size, the multiple comparisons correction method used, the statistical score (usually T-statistics or Z-values), and the brain regions of interest labeled with reference to a standard atlas and/or visual inspection. The correct interpretation of fMRI results is never straightforward and is dependent upon factors which range widely from the technical and methodological to the conceptual and statistical issues. Because there is such great variation in the manner with which studies are performed (Lange et al., 1999; Carp, 2012a; McGonigle, 2012), it is critically important that researchers/clinicians fully describe and report the methodological details as well as results, thus allowing replication as well as the potential incorporation of the findings into meta-analytic studies (Carp, 2012b). Comprehensive guidelines for reporting an fMRI study (Poldrack et al., 2008), as well as the principles of open and reproducible research for neuroimaging (Nichols et al., 2016) have been proposed, and have been accompanied by the development of a number of databases (Van Horn and Ishai, 2007; Poldrack and Gorgolewski, 2014; Poldrack and Poline, 2015). Specific examples of such data pools include OpenfMRI (Poldrack and Gorgolewski, 2015), ConnectomeDB (Hodge et al., 2016), Neuroinformatics Database (NiDB) (Book et al., 2016), or NeuroVault.org (Gorgolewski et al., 2016b). Effective communication of the results of fMRI investigations requires that the information has been organized and described in a clear and straightforward manner, using an unambiguous ontology (formal description of all terms and syntax) (Burns and Turner, 2013; Poldrack and Yarkoni, 2016) and format (Gorgolewski et al., 2016a).

The BOLD signal itself has a number of characteristics which present challenges to the accurate interpretation of fMRI data acquired with its use (Aguirre et al., 1998). BOLD responses are known to vary with different acquisition parameters (Renwall et al., 2014) and to be highly dependent on the specific parameters of neurovascular coupling which are known to vary with age, medication and in certain pathological states (D'Esposito et al., 2003; Bangen et al., 2009; Di et al., 2013; Tsvetanov et al., 2015). In addition, the nature of the BOLD signal has been shown to be affected by a variety of chemical compounds (e.g., caffeine and alcohol) (Levin et al., 1998; Mulderink et al., 2002; Perthen et al., 2008) as well as by respiration (Birn et al., 2008) and oxygen level (Cardenas et al., 2015).

The fundamental challenge of fMRI research is to draw conclusions which are completely supported by the data and which are unbiased. The literature contains numerous examples wherein foci of static regional activation are interpreted as associated with specific cognitive functions. Such empirical conclusions, termed “reverse inference” (Poldrack, 2006) are based on the implicit assumption that when a region of activation changes as a function of the performance of a specific task, that the region whose activity has changed is responsible for the associated cognitive process. This assumption fails to take into account either brain compensatory mechanisms (Sack et al., 2005; Meade et al., 2016) or plasticity (Poldrack, 2000; Colcombe et al., 2004; Amad et al., 2016). It is now generally accepted that a more complete description of brain function must include not only the notion of causality but also recognize the relationship between interconnected regions (network properties) through the characterization both of functional specialization (specific roles played by the different regions) and integration (how the regions interact with one another) (Van Horn and Poldrack, 2009). For all these reasons, drawing significant conclusions about mental states from fMRI data is challenging at best and the use of classification and predictive models such as machine learning algorithms have increasingly been tasked for this purpose (Pereira et al., 2009; Dosenbach et al., 2010).

As stated often throughout this guide, the statistical analysis of fMRI data is a complex process and great caution must be exercised when interpreting the experimental results. Questions have been raised, for example, about whether certain studies have reported findings able to be supported by the methodology used and the data obtained. Some studies purport to find extremely high degrees of correlation between individual behavioral characteristics (including personality, emotion, and social cognition) and specific regions of increased brain activity (Vul et al., 2009). Critics have pointed out that, considering the degree of methodological imprecision both of fMRI and in the measurement of individual characteristics, that the reported results may not be robust (Vul et al., 2009). Another issue is that of circular analysis, unfortunately seen with some frequency in functional studies. The issue arises when the data first are analyzed, subsets of those data selected, and then the same subsets are re-analyzed to obtain the results (Kriegeskorte et al., 2009). An fMRI example might be to define a ROI on the very basis of a statistical mapping which highlights the voxels of which it is composed in response to a functional activation state (Kriegeskorte et al., 2009). Such “double dipping,” the use of the same data for selection and subsequent selective analysis, results in an invalid statistical inference. It violates the criterion that the test statistics must be inherently independent of the selection criteria under the null hypothesis.

## CONCLUSIONS AND FUTURE DIRECTIONS

Functional MRI currently is enjoying popularity in the study of brain function and promises to become even more prominent

in the future. A number of factors contributing to the optimism about the expanding role of fMRI in neuroscience include: greater understanding of the BOLD and other contrast mechanisms; higher resolution and increased sensitivity; the use of new, more optimized preprocessing and analytic techniques; more powerful computational models; and extensive data sharing, enabling the design of studies comprised of large numbers of participants. Strategically, functional neuroimaging appears to be moving from the description and characterization of brain states toward predictive models of function based on the whole brain network. It is hoped that such models will incorporate behavioral features, genetic factors and biomarkers and will evolve to play an increasingly prominent clinical role in diagnosis, monitoring, and treatment of central nervous system disorders. In order to contribute to future progress, this article has sought to highlight the typical challenges faced when performing fMRI studies, and to offer some practical strategies with which they may be overcome. We have provided guidelines and references for the tools most commonly used at each step of the principal fMRI pipeline. As a concluding remark, we outline a set of general recommendations that we consider to be of upmost relevance for a better transparency and reproducibility of neuroimaging studies: before the study, perform a suitable experimental planning, including a proper design, power analysis (e.g., use the reported estimates as a means to estimate the adequate sample size) and identify the specific targets and analyses to be implemented; during the study, define the adequate acquisition protocols, identify as soon as possible and prevent the potential artifacts (in order to avoid losing data), carefully check the quality of the data, perform an accurate preprocessing, analysis and statistical testing and organize all the information in a standardized way, preferably with open-source software; after having the results, discuss them with caution, report them as well as the methodological details with great detail and following the guidelines (allowing study replication) and share the full statistical maps ideally in open repositories (allowing meta-analyses and power analyses for other similar studies). It is our hope that this guide will be of assistance both to those beginning to explore the potential of functional imaging as well as those who might appreciate a source book of current practice.

## AUTHOR CONTRIBUTIONS

JMS, RM, PSM, AS (Alexandre Sousa), and PM contributed in literature search, figures, study design, and writing. EG contributed in writing. AS (Adriana Sampaio), VA, and NS contributed in study design and writing.

## ACKNOWLEDGMENTS

This article has been developed under the scope of the project NORTE-01-0145-FEDER-000013, supported by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER).