NeuroImage 124 (2016) 704–713

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Challenges in the reproducibility of clinical studies with resting state fMRI: An example in early Parkinson's disease

Ludovica Griffanti [a], Michal Rolinski [b,c], Konrad Szewczyk-Krolikowski [b,c], Ricarda A. Menke [a], Nicola Filippini [a,d], Giovanna Zamboni [a], Mark Jenkinson [a], Michele T.M. Hu [b,c], Clare E. Mackay [a,b,d,*]

[a] Centre for the functional MRI of the Brain (FMRIB), University of Oxford, Oxford, UK
[b] Oxford Parkinson's Disease Centre (OPDC), University of Oxford, Oxford, UK
[c] Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[d] Department of Psychiatry, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 21 May 2015
Accepted 9 September 2015
Available online 16 September 2015

Keywords:
Resting state functional magnetic resonance imaging (rfMRI)
Functional connectivity
Artefact removal
Dual regression
Basal ganglia network
Parkinson's disease

ABSTRACT

Resting state fMRI (rfMRI) is gaining in popularity, being easy to acquire and with promising clinical applications. However, rfMRI studies, especially those involving clinical groups, still lack reproducibility, largely due to the different analysis settings. This is particularly important for the development of imaging biomarkers. The aim of this work was to evaluate the reproducibility of our recent study regarding the functional connectivity of the basal ganglia network in early Parkinson's disease (PD) (Szewczyk-Krolikowski et al., 2014). In particular, we systematically analysed the influence of two rfMRI analysis steps on the results: the individual cleaning (artefact removal) of fMRI data and the choice of the set of independent components (template) used for dual regression. Our experience suggests that the use of a cleaning approach based on single-subject independent component analysis, which removes non neural-related sources of inter-individual variability, can help to increase the reproducibility of clinical findings. A template generated using an independent set of healthy controls is recommended for studies where the aim is to detect differences from a "healthy" brain, rather than an "average" template, derived from an equal number of patients and controls. While, exploratory analyses (e.g. testing multiple resting state networks) should be used to formulate new hypotheses, careful validation is necessary before promising findings can be translated into useful biomarkers.

Introduction

Resting state functional MRI (rfMRI) has been shown to be a promising tool for exploring brain functions and assessing their alteration in neurodegenerative conditions (Barkhof et al., 2014). Over the last decade, several resting state networks (RSNs) have been identified (Beckmann and Smith, 2004; Smith et al., 2009) and associated with specific brain functions through the comparison with results obtained from task-based fMRI experiments (Smith et al., 2009; Zamboni et al., 2013). Moreover, rfMRI has been shown to be stable across subjects (Smith et al., 2009; Zuo and Xing, 2014), easy to acquire, and as it is not dependent on task performance, functional connectivity (FC) of the RSNs can be evaluated in impaired subjects. Therefore, rfMRI has become a common technique in clinical research studies. With observed alterations of RSNs now reported in subjects with clinical symptoms and increased at-risk of developing pathology (Barkhof et al., 2014; Filippini et al., 2009; Sole-Padulles et al., 2013), rfMRI may have a vital role in the development of novel imaging biomarkers.

Despite the importance of obtaining reliable and stable results that may be later used as biomarkers, rfMRI studies, especially those involving clinical groups, still lack reproducibility. In fact, even when reproducibility tests are performed, they are usually performed on healthy controls, and issues may only become apparent when dealing with patient groups. For example, logistical difficulties may arise from subjecting patients to long or multiple scanning sessions. Moreover, in clinical studies, images are typically acquired using clinical scanners. This may result in poorer data quality, leading to suboptimal processing steps, such as registration and artefact removal. Importantly, the most reproducible networks (default, control and attention networks—see Zuo and Xing, 2014) may not necessarily be the ones that are of the greatest clinical importance. For example, although only recently described (Robinson et al., 2009) and, therefore, not studied in great detail, the basal ganglia network (BGN) has recently been shown to be affected in early PD (Szewczyk-Krolikowski et al., 2014).

In addition to the paucity of within-group test–retest reliability (Zuo and Xing, 2014), the lack of reproducibility between studies may be due

* Corresponding author at: Department of Psychiatry, University of Oxford Warneford Hospital Oxford, OX3 7JX. Fax: +44 1865 793101.
E-mail address: clare.mackay@psych.ox.ac.uk (C.E. Mackay).

http://dx.doi.org/10.1016/j.neuroimage.2015.09.021

to the different analysis settings, with a major contributor being the many permutations in analysis pipelines. In a fast moving field of rfMRI, there is continual development and refinement of methodology. Several studies evaluated the impact of analysis methods on the reproducibility and reliability of RSNs (Franco et al., 2013; Zuo et al., 2010; Zuo and Xing, 2014). Specifically, it has been demonstrated that independent component analysis (ICA), and in particular group-ICA followed by dual regression, rather than single-subject ICA and template matching (Zuo et al., 2010), is more stable than seed-based analysis (Zuo and Xing, 2014). However, even within these guidelines, there are several analytical details that can influence the results and make comparisons difficult.

In light of these observations, we endeavoured to evaluate the reproducibility of our recent study of functional connectivity within the BGN of patients with early PD (Szewczyk-Krolikowski et al., 2014). The difference observed in the BGN connectivity was substantial in both magnitude and extent and therefore provides a good test-bed. In particular, we systematically analysed the influence of two rfMRI analysis steps: the individual cleaning (artefact removal) of fMRI data and the choice of a RSNs template (a set of independent components) within the framework of dual-regression ICA. The aim of this work was to establish how strongly the settings of these steps affected the observed results. We hoped to aid interpretations and comparisons across studies and contribute to the translational pipeline for reliable imaging clinical biomarkers.

Materials and methods

Participants

Fifty-nine patients with PD (mean age = 63.2 ± 10.9 years, F:M = 25:34) and thirty age- and gender-matched healthy controls (HC) (mean age = 62.8 ± 7.2, F:M = 14:16) were recruited from the Oxford Parkinson's Disease Centre (OPDC) cohort (Rolinski et al., 2014). This sample includes the cohort described in Szewczyk-Krolikowski et al. (2014). Patients included in the PD group met the UK PD Society Brain Bank Criteria for clinically probable idiopathic PD (Hughes et al., 1992), having predominantly akinetic-rigid parkinsonism with minimal tremor. Patients taking dopaminergic medications were scanned in a clinically defined "off-state," a minimum of 12 hours after the withdrawal of their relevant medications. Subjects included in the HC group had no family history of parkinsonism and were recruited largely from the spouses and friends of the PD participants. All participants underwent a detailed clinical assessment (Szewczyk-Krolikowski et al., 2014). Both groups only included subjects classified as cognitively healthy, as defined by a Mini-Mental State Examination (MMSE) >26 (Folstein et al., 1975) and no subjective complaint of memory problems.

Each subject gave written consent to participate in the study, which was conducted with the approval of the local NHS ethics committee and in compliance with national legislation and the Declaration of Helsinki.

Neuroimaging data acquisition and preprocessing

Scanning was performed at the Oxford Centre for Clinical Magnetic Resonance Research (OCMR) using a 3 T Trio Siemens MRI scanner (Erlangen, Germany) equipped with a 12-channel head coil. The protocol included 1) high-resolution T1-weighted images (MPRAGE, resolution $1 \times 1 \times 1$ mm^3, TE/TR = 4.7 ms/2040 ms, 192 axial slices, 6 minutes); 2) rfMRI images (EPI, resolution $3 \times 3 \times 3.5$ mm^3, TE/TR = 28 ms/2000 ms, 34 axial slices per volume, covering both hemispheres with incomplete coverage of the cerebellum, 180 volumes in 6 minutes, eyes open); 3) field map images, to account for distortions caused by field inhomogeneities (GRE, resolution $3 \times 3 \times 3.5$ mm^3, TR = 488 ms, TE = 5.19 ms and 7.65 ms).

The analysis of resting state fMRI data was performed using FSL software package (Jenkinson et al., 2012). Firstly, images were motion corrected with MCFLIRT; from this operation, the six rigid-body parameter time series were extracted for each subject (to be used for subsequent cleaning) and the mean relative displacement was calculated to ensure that the two groups were matched in terms of average amount of head motion (HC: 0.14 ± 0.09 mm; PD: 0.12 ± 0.05 mm, $p = 0.23$). Following preprocessing steps included brain extraction, unwarping using fieldmap data, spatial smoothing using a Gaussian kernel of FWHM of 6 mm, and high-pass temporal filtering of 150 s. Single-subject probabilistic independent component analysis (ICA) was then performed with MELODIC tool (Beckmann and Smith, 2004) with automated dimensionality estimation to be used for ICA-based artefact removal.

T1-weighted images were brain-extracted and used as anatomical references for fMRI. Tissue segmentation was also performed with FAST (Zhang et al., 2001) and the grey matter (GM) images were registered to the MNI 152 standard space using non-linear registration with FNIRT and used to generate voxel-wise confound regressors for fMRI statistical analyses.

Reproducibility analyses of resting state fMRI data

Analyses overview

In this work, we aimed to systematically analyse the influence of two rfMRI analysis steps: (1) the individual cleaning (artefact removal) of fMRI data and (2) the choice of the set of independent components used as input for dual regression (from now on referred as *template*).

The impact of artefact removal was tested on a subsample of 19 HC and 19 PD (matched for age, sex, and head motion) of our cohort, specifically the same subjects used in Szewczyk-Krolikowski et al. (2014), comparing six cleaning options (see Section 2.3.2. for details). The rationale for using this subsample for this first analysis is that we judged it to be sufficiently large to test differences among the different approaches, while limiting the manual intervention (in terms of both expertise and time) required for hand-labelling the single-subject components (used as gold standard cleaning method). Firstly, we tested the effect of cleaning on the temporal signal-to-noise ratio, which should be higher with better cleaning. Subsequently, we calculated spatial correlations between the subject-specific BGN maps (derived with dual regression) obtained after each cleaning approach with respect to a gold standard (the BGN maps obtained with manual cleaning, see Section *Influence of artefact removal*). A higher spatial correlation corresponds to a better cleaning approach. In order to compare the effect of cleaning on between-group discriminability, we performed a regions-of-interest (ROI) analysis and a voxel-wise analysis of the BGN. We then repeated the comparison, among the automated methods only, on the full sample (30 HC and 59 PD, which included the subsample described above) to verify that the results obtained in the subsample were consistent and reproducible with respect to sample size.

Secondly, the impact of the template used for dual regression was tested on the whole cohort of 30 HC and 59 PD, comparing six templates (see Section *Influence of template for dual regression* for details). Similarly to the analyses carried out to compare the effect of the cleaning approaches, we evaluated the impact of the template choice on between-group discriminability by performing an ROI analysis and a voxel-wise analysis of the BGN, also quantifying the level of similarity/overlap among the results of the voxel-wise analyses.

Additionally, to ensure that our results were not influenced by the sample composition, we randomly split the full sample 100 times into two group pairs of PD patients and HC, repeated the analyses with different cleaning methods and the templates, and calculated the reproducibility across groups' composition. The detailed methods and results relative to this analysis are described in the supplementary material.

Influence of artefact removal

To remove the effect of motion, non-neural physiology, scanner artefacts, and other confounds, we applied a number of different cleaning