NeuroImage 117 (2015) 67–79

Contents lists available at ScienceDirect

NeuroImage

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

Optimization of rs-fMRI Pre-processing for Enhanced Signal-Noise Separation, Test-Retest Reliability, and Group Discrimination

William R. Shirer [a,*,1], Heidi Jiang [a,b,*,1], Collin M. Price [a], Bernard Ng [a,c], Michael D. Greicius [a]

[a] Functional Imaging in Neuropsychiatric Disorders (FIND) Lab, Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, USA
[b] Interdepartmental Program in Neuroscience, Northwestern University, Chicago, USA
[c] Parietal Team, Neurospin, INRIA Saclay, France

ARTICLE INFO

Article history:
Received 20 December 2013
Accepted 7 May 2015
Available online 15 May 2015

Keywords:
Resting-state
Preprocessing
Reliability
Noise
Temporal filter
Global signal

ABSTRACT

Resting-state functional magnetic resonance imaging (rs-fMRI) has become an increasingly important tool in mapping the functional networks of the brain. This tool has been used to examine network changes induced by cognitive and emotional states, neurological traits, and neuropsychiatric disorders. However, noise that remains in the rs-fMRI data after preprocessing has limited the reliability of individual-subject results, wherein scanner artifacts, subject movements, and other noise sources induce non-neural temporal correlations in the blood oxygen level-dependent (BOLD) timeseries. Numerous preprocessing methods have been proposed to isolate and remove these confounds; however, the field has not coalesced around a standard preprocessing pipeline. In comparisons, these preprocessing methods are often assessed with only a single metric of rs-fMRI data quality, such as reliability, without considering other aspects in tandem, such as signal-to-noise ratio and group discriminability. The present study seeks to identify the data preprocessing pipeline that optimizes rs-fMRI data across multiple outcome measures. Specifically, we aim to minimize the noise in the data and maximize result reliability, while retaining the unique features that characterize distinct groups. We examine how these metrics are influenced by bandpass filter selection and noise regression in four datasets, totaling 181 rs-fMRI scans and 38 subject-driven memory scans. Additionally, we perform two different rs-fMRI analysis methods – dual regression and region-of-interest based functional connectivity – and highlight the preprocessing parameters that optimize both approaches. Our results expand upon previous reports of individual-scan reliability, and demonstrate that preprocessing parameter selection can significantly change the noisiness, reliability, and heterogeneity of rs-fMRI data. The application of our findings to rs-fMRI data analysis should improve the validity and reliability of rs-fMRI results, both at the individual-subject level and the group level.

Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) has been an important tool in mapping the functional networks of the brain, and in examining network changes induced by cognitive and emotional states, traits, and disorders. Many findings related to resting-state networks, including the spatial distribution of functional networks and their cognitive correlates (Biswal et al., 1995; Hampson et al., 2002; Greicius et al., 2003; Kiviniemi et al., 2003; Beckmann et al., 2005; Bellec et al., 2006; Damoiseaux et al., 2006; Seeley et al., 2007; Kiviniemi et al., 2009; Smith et al., 2009), changes in network connectivity induced by cognitive states (Richiardi et al., 2010; Shirer et al., 2012), and group differences in network connectivity due to brain disorders (Greicius et al., 2004, 2007; Hedden et al., 2009; Seeley et al., 2009; Sheline et al., 2010; Zhou et al., 2010; Johnson et al., 2013), are well-replicated at the group level. However, reliability at the individual level remains rather limited (Damoiseaux and Greicius, 2009; Honey et al., 2009). Previous studies that have evaluated the consistency of resting-state networks in individuals have mainly used the intra-class correlation coefficient (ICC) as a measure of test-retest reliability, which is scaled from 0 to 1—where 0 represents no reliability, and 1 represents perfect reliability. These studies have reported values on the order of 0.5 (Shehzad et al., 2009) to 0.6 (Zuo et al., 2010; Thomason et al., 2011; Guo et al., 2012), which reflect only modest test-retest reliability, thereby diminishing confidence in group contrasts, undermining generalizability of classification approaches, and confounding interpretation of subject-level results. The low test-retest reliability of rs-fMRI data is largely attributable to noise that remains in the data after preprocessing, wherein scanner artifacts, subject movements, and other noise sources induce non-neural temporal correlations in the blood oxygen level-dependent (BOLD) signal. Numerous preprocessing methods have been proposed to isolate and remove these confounds (Fox et al., 2009; Murphy et al., 2009; Weissenbacher

* Corresponding authors.
E-mail addresses: shirer@stanford.edu (W.R. Shirer), hhjiang@u.northwestern.edu (H. Jiang).
[1] Authors contributed equally to this work.

http://dx.doi.org/10.1016/j.neuroimage.2015.05.015
1053-8119/

et al., 2009; Braun et al., 2012; Friston et al., 1996; Power et al., 2012, 2013, 2014; Satterthwaite et al., 2013; Siegel et al., 2014; Yan et al., 2013a).

The present study seeks to comprehensively and simultaneously compare various preprocessing strategies, and assess their effects on multiple outcome measures: signal-noise separation, test-retest reliability, and group discriminability. Previous studies attempting to optimize rs-fMRI data have focused exclusively on one of these metrics without consideration as to how other aspects of data quality are affected. In particular, studies examining test-retest reliability of rs-fMRI data do not quantify how much noise remains in the data, or how maximizing this measure affects the ability to distinguish different groups of subjects (Shehzad et al., 2009; Zuo et al., 2010; Thomason et al., 2011; Guo et al., 2012; Yan et al., 2013b).

We focus our analyses on the effects of common preprocessing steps, such as global signal regression (GS) (Weissenbacher et al., 2009; Shirer et al., 2012); removal of cerebrospinal fluid (CSF) and white matter (WM) confounds (Shirer et al., 2012); noise regression of motion parameters estimated during motion correction (Friston et al., 1996; Power et al., 2012, 2013; Satterthwaite et al., 2013; Yan et al., 2013a); and temporal filtering at various frequency bands reported in the literature (Achard et al., 2006; Ko et al., 2011; Guo et al., 2012). To maximize generalizability of our results, we elect not to examine preprocessing techniques that are unavailable at some imaging centers, such as removal of physiological noise using RETROICOR or RVHRCOR (Chang and Glover, 2009).

We analyze the rs-fMRI data using two measures of functional connectivity (FC): Pearson's correlation between regions of interest (ROI) pairs, and ICA (dual regression, or DR). We then assess the quality of the outcome data based on three metrics. First, we propose a novel method of estimating signal-noise separation (SNS) in rs-fMRI data. Here, we calculate the functional connectivity between brain areas within well-established networks, and contrast this with spurious connectivity between these areas and non-neural regions outside the brain. This procedure is performed for each preprocessing permutation to identify the preprocessing parameters that produce the greatest ratio of network connectivity to noise. Second, we assess the impact of preprocessing parameters on the test-retest reliability (TRT) of whole-brain connectivity. We do so by measuring the TRT of network connectivity produced by each preprocessing pipeline, using the same intra-class correlation (ICC) metric as reported in previous rs-fMRI test-retest reliability studies (Shehzad et al., 2009; Zuo et al., 2010; Thomason et al., 2011; Guo et al., 2012). The effects of preprocessing on SNS and TRT are first evaluated on data collected by us, and then examined on an independent dataset that has been widely used in previous test-retest reliability studies for comparison (Shehzad et al., 2009; Zuo et al., 2010; Thomason et al., 2011; Guo et al., 2012). Finally, we assess the clinical relevance of different preprocessing pipelines by examining group discriminability (GD), wherein we compare different pipelines' accuracy in classification of Alzheimer's disease (AD) patients and controls. To ensure that the preprocessing strategies most successful for classifying disease states are not simply due to non-neural artifacts (e.g. differences in the amount of movement between groups), we also explore the effect of different preprocessing pipelines on the ability to classify different cognitive states within individual subjects.

Materials and methods

Subjects

Stanford University data

Data were compiled from three datasets, resulting in a total of thirty-eight healthy right-handed subjects (21 females) aged 18-78, and fifteen subjects with mild AD (8 females) aged 51-86 who participated in this study. The first Stanford healthy control dataset (S-HC1) contained twenty-three healthy right-handed subjects (15 females) aged 56-76. The second Stanford healthy control dataset (S-HC2) contained fifteen healthy right-handed subjects (6 females) aged 22-46. The third dataset, S-AD, contained fifteen subjects with very mild to mild Alzheimer's disease (8 females). The subjects with dementia were recruited from the Stanford Center for Memory Disorders, met NINCDS-ADRDA criteria (McKhann et al., 1984), and were scanned at a relatively early stage of their illness (mini-mental state examination average score: 23.9 ± 3.5). The experimental protocols for all datasets were approved by the Institutional Review Board of Stanford University.

New York University data

The New York University (NYU) healthy control dataset (N-HC) contained twenty-five healthy right-handed subjects (15 females) aged 21-49. This data has been published and made available through the 1000 Functional Connectomes Project (Biswal et al., 2010; http://www.nitrc.org/projects/fcon_1000/). Data were collected according to protocols approved by the institutional review boards of NYU and the NYU School of Medicine.

Imaging methods

Stanford University data

Functional and structural images for S-HC1, S-HC2, and S-AD were acquired on a 3 Tesla (T) General Electric scanner using an 8-channel head coil. To reduce blurring and signal loss arising from field inhomogeneities, an automated high-order shimming method based on spiral acquisitions was used (Kim et al., 2002). Thirty-one axial slices (4 mm thick, 0.5 mm skip) covering the whole brain were imaged using a T2*-weighted gradient-echo spiral pulse sequence (time repetition = 2000 ms, time echo = 30 ms, flip angle = 80°, and 1 interleave) (Glover and Lai, 1998; Glover and Law, 2001). The field of view was 220 x 220 mm^2, and the matrix size was 64x64, giving an in-plane spatial resolution of 3.4375 mm. For resting state scans, 240 contiguous functional volumes were collected in S-HC1, S-HC2, and S-AD. Subjects were instructed to close their eyes, let their minds wander, and try not to focus on any one thing. For each subject in S-HC1 and S-AD, two resting-state scans were collected within the same scan session. For each subject in S-HC2, two resting-state scans were collected in two different sessions, two to six weeks apart from one another. An additional memory-state scan containing 150 contiguous volumes was collected for S-HC1 and S-AD during the same session as the resting-state scans (Fig. 1). For this task, subjects were instructed to close their eyes and go over their day in great detail for the duration of the scan. Memory scans from S-AD were excluded, since we are unable to ensure that they performed the task. For spatial normalization and localization, a high-resolution T1-weighted magnetization prepared gradient echo sequence was also obtained (3D FSPGR, TR = 8.508 ms; TE = 3.384 ms; flip angle = 15°; 162 slices, FOV = 220 mm^2).

New York University data

Functional and structural images for N-HC were acquired on a 3 T Siemens Allegra scanner. Each scan consisted of 197 contiguous EPI functional volumes (TR = 2000 ms; TE = 25 ms; flip angle = 90°, 39 slices, matrix = 64×64; FOV = 192 mm^2; acquisition voxel size = $3 \times 3 \times 3$ mm). For spatial normalization and localization, a high-resolution T1-weighted magnetization prepared gradient echo sequence was also obtained (MPRAGE, TR = 2500 ms; TE = 4.35 ms; TI = 900 ms; flip angle = 8°; 176 slices, FOV = 256 mm^2). Subjects in N-HC were instructed to remain still with their eyes open during the scan. Three resting-state scans were collected for each subject; scans 2 and 3 were conducted in a single scan session, 45 min apart, and were 5–16 months (mean 11 ± 4) after scan 1 (Fig. 1).