NeuroImage xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Review

Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes

Chao-Gan Yan [a,b,c], R. Cameron Craddock [a,b], Xi-Nian Zuo [d], Yu-Feng Zang [e], Michael P. Milham [a,b,*]

[a] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA
[b] Center for the Developing Brain, Child Mind Institute, New York, NY, USA
[c] The Phyllis Green and Randolph Cowen Institute for Pediatric Neuroscience, New York University Child Study Center, New York, NY, USA
[d] Key Laboratory of Behavioral Science, Laboratory for Functional Connectome and Development, Magnetic Resonance Imaging Research Center, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[e] Center for Cognition and Brain Disorders, Hangzhou Normal University, Hangzhou, China

ARTICLE INFO

Article history:
Accepted 16 April 2013
Available online xxxx

Keywords:
Functional connectomics
Standardization
Test–retest reliability
Resting-state fMRI
Data aggregation

ABSTRACT

As researchers increase their efforts to characterize variations in the functional connectome across studies and individuals, concerns about the many sources of nuisance variation present and their impact on resting state fMRI (R-fMRI) measures continue to grow. Although substantial within-site variation can exist, efforts to aggregate data across multiple sites such as the 1000 Functional Connectomes Project (FCP) and International Neuroimaging Data-sharing Initiative (INDI) datasets amplify these concerns. The present work draws upon standardization approaches commonly used in the microarray gene expression literature, and to a lesser extent recent imaging studies, and compares them with respect to their impact on relationships between common R-fMRI measures and nuisance variables (e.g., imaging site, motion), as well as phenotypic variables of interest (age, sex). Standardization approaches differed with regard to whether they were applied post-hoc vs. during pre-processing, and at the individual vs. group level; additionally they varied in whether they addressed additive effects vs. additive + multiplicative effects, and were parametric vs. non-parametric. While all standardization approaches were effective at reducing undesirable relationships with nuisance variables, post-hoc approaches were generally more effective than global signal regression (GSR). Across approaches, correction for additive effects (global mean) appeared to be more important than for multiplicative effects (global SD) for all R-fMRI measures, with the exception of amplitude of low frequency fluctuations (ALFF). Group-level post-hoc standardizations for mean-centering and variance-standardization were found to be advantageous in their ability to avoid the introduction of artifactual relationships with standardization parameters; though results between individual and group-level post-hoc approaches were highly similar overall. While post-hoc standardization procedures drastically increased test–retest (TRT) reliability for ALFF, modest reductions were observed for other measures after post-hoc standardizations—a phenomena likely attributable to the separation of voxel-wise from global differences among subjects (global mean and SD demonstrated moderate TRT reliability for these measures). Finally, the present work calls into question previous observations of increased anatomical specificity for GSR over mean centering, and draws attention to the near equivalence of global and gray matter signal regression.

Contents

* Corresponding author at: Center for the Developing Brain, Child Mind Institute, 445 Park Avenue, New York, NY 10022, USA. Fax: +1 646 625 4348.
E-mail address: Michael.Milham@childmind.org (M.P. Milham).

1053-8119/$ – see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.neuroimage.2013.04.081

Introduction

Measurement standardization represents a key challenge for the field of functional connectomics. As researchers increase their efforts to characterize variations in the functional connectome observed across studies and individuals, concerns about the many known and unknown sources of nuisance variation present and their impact on resting state fMRI (R-fMRI) measures continue to grow (Cole et al., 2010; Kelly et al., 2012). Between studies, MR acquisition methodologies are among the most commonly cited sources of measurement variation (Friedman and Glover, 2006b); yet a multitude of experimental, environmental and subject-related factors can introduce unintended variations in measurement as well (Table 1). Few, if any, of these factors are addressed in imaging studies. Finally, head-motion and physiologic parameters (cardiac or respiratory effects) are major sources of measurement variation, which can at times be related to systematic variables of interest (e.g., age, diagnostic status) (Power et al., 2012a; Satterthwaite et al., 2012; Van Dijk et al., 2012). A growing reality is that even the best efforts to standardize data acquisition and limit the number of unknowns, unwanted sources of variation in R-fMRI studies will remain.

In 2009, the publicly released 1000 Functional Connectomes Project (FCP) and International Neuroimaging Data-sharing Initiative (INDI) provided a stark portrayal of variability in imaging methodologies employed by the neuroimaging field. Comprised of R-fMRI samples independently collected at imaging sites around the world, notable variation in almost every aspect of imaging acquisition methodologies represented in these datasets while the majority of participant-related variables are not reported (and in most cases, were not systematically recorded). As expected, remarkable site-related variation is detectable in R-fMRI measures derived from the FCP/INDI datasets, raising understandable concerns about whether such data could be harmonized and analyzed. Fortunately, despite justifiable skepticism, feasibility analyses demonstrated that meaningful explorations of the aggregate dataset ($n = 1093$; 24 imaging sites) could be performed (Biswal et al., 2010). After accounting for site-related differences, discovery analyses revealed brain–behavior relationships with phenotypic variables such as sex, age, and diagnostic status, and confirmed a variety of prior hypotheses (Biswal et al., 2010; Fair et al., 2012; Tomasi and Volkow, 2010; Zuo et al., 2012). Although encouraging, the many unknown and uncontrolled factors in the FCP/INDI remain a source of concern, as they extend beyond simple site effects and can limit the utility of the datasets.

The goal of the present work is to provide a comprehensive assessment of the impact of post-acquisition standardization methodologies on common R-fMRI data analyses, using data from the original 1000 Functional Connectomes Project. Several strategies for standardization have already emerged in the field including mean division (Zang et al., 2004, 2007); Z-score standardization (Beckmann et al., 2005; Buckner et al., 2009; Calhoun et al., 2001; Zou et al., 2008; Zuo et al., 2010a, 2012, 2013); and Gaussian function fit normalization (Lowe et al., 1998). However, these methods are not consistently used and have not been systematically compared. Additional approaches can be borrowed from the molecular genetics community, which has made noteworthy strides in dealing with unwanted variation in microarray technologies and procedures (see Quackenbush, 2002 for a review). Drawing on these two sources we identified 11 standardization approaches (Table 2) to apply to the original FCP dataset and compare with respect to their impact on commonly examined R-fMRI measures, their test–retest (TRT) reliability and phenotypic relationships (sex, age), as well as nuisance variables of interest.

Table 1
Factors can introduce unintended variations in fMRI measurement.

Category	Factor
1. Acquisition-related variations	Scanner make and model (Friedman and Glover, 2006b), sequence type (spiral vs. echo planar; single-echo vs. multi-echo) (Klarhofer et al., 2002), parallel vs. conventional acquisition (Feinberg et al., 2010; Lin et al., 2005), coil type (surface vs. volume, number of channels, orientation), repetition time, number of repetitions, flip angle, echo time, and acquisition volume (field of view, voxel size, slice thickness/gaps, slice prescription) (Friedman and Glover, 2006a)
2. Experimental-related variations	Participant instructions (Hartstra et al., 2011), eyes-open/eyes-closed (Yan et al., 2009; Yang et al., 2007), visual displays, experiment duration (Fang et al., 2007; Van Dijk et al., 2010)
3. Environment-related variations	Sound attenuation measures (Cho et al., 1998; Elliott et al., 1999), attempts to improve participant comfort during scans (e.g., music, videos) (Cullen et al., 2009), head-motion restraint techniques (e.g., vacuum pad, foam pad, bite-bar, plaster cast head holder) (Edward et al., 2000; Menon et al., 1997), room temperature and moisture (Vanhoutte et al., 2006).
4. Participant-related variations	Circadian cycle (Shannon et al., 2012), prandial (Haase et al., 2009), caffeine (Rack-Gomer et al., 2009), and nicotine status (Tanabe et al., 2011), sleepiness/arousal (Horovitz et al., 2008), sleep deprivation (Samann et al., 2010), scanner anxiety (de Bie et al., 2010), and menstrual cycle status (for women) (Protopopescu et al., 2005)

Please cite this article as: Yan, C.-G., et al., Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes, NeuroImage (2013), http://dx.doi.org/10.1016/j.neuroimage.2013.04.081