NeuroImage 154 (2017) 174–187

Contents lists available at ScienceDirect

NeuroImage

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

Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity

Rastko Ciric[a], Daniel H. Wolf[a], Jonathan D. Power[b], David R. Roalf[a], Graham L. Baum[a], Kosha Ruparel[a], Russell T. Shinohara[c], Mark A. Elliott[d], Simon B. Eickhoff[e,f], Christos Davatzikos[d], Ruben C. Gur[a,d], Raquel E. Gur[a,d], Danielle S. Bassett[g,h], Theodore D. Satterthwaite[a,*]

[a] Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[b] Department of Psychiatry, Weill Cornell Medical College, NY, NY, USA
[c] Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[d] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[e] Institute for Systems Neuroscience, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany
[f] Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Germany
[g] Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA
[h] Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA

ARTICLE INFO

Keywords:
fMRI
Functional connectivity
Artifact
Confound
Motion
Noise

ABSTRACT

Since initial reports regarding the impact of motion artifact on measures of functional connectivity, there has been a proliferation of participant-level confound regression methods to limit its impact. However, many of the most commonly used techniques have not been systematically evaluated using a broad range of outcome measures. Here, we provide a systematic evaluation of 14 participant-level confound regression methods in 393 youths. Specifically, we compare methods according to four benchmarks, including the residual relationship between motion and connectivity, distance-dependent effects of motion on connectivity, network identifiability, and additional degrees of freedom lost in confound regression. Our results delineate two clear trade-offs among methods. First, methods that include global signal regression minimize the relationship between connectivity and motion, but result in distance-dependent artifact. In contrast, censoring methods mitigate both motion artifact and distance-dependence, but use additional degrees of freedom. Importantly, less effective de-noising methods are also unable to identify modular network structure in the connectome. Taken together, these results emphasize the heterogeneous efficacy of existing methods, and suggest that different confound regression strategies may be appropriate in the context of specific scientific goals.

Introduction

Resting-state (intrinsic) functional connectivity (rsfc-MRI) has evolved to become one of the most common brain imaging modalities (Craddock et al., 2013; Fox and Raichle, 2007; Power et al., 2014b; Smith et al., 2013; Van Dijk et al., 2010), and has been critical for understanding fundamental properties of brain organization (Damoiseaux et al., 2006; Fox et al., 2005; Power et al., 2011; Yeo et al., 2011), brain development over the lifespan (DiMartino et al., 2014; Dosenbach et al., 2011; Fair et al., 2008), and abnormalities associated with diverse clinical conditions (Baker et al., 2014; Buckner et al., 2008; Fair et al., 2010). rsfc-MRI has numerous advantages, including ease of acquisition and suitability for a wide and expanding array of analysis techniques. However, despite knowledge that in-scanner motion can influence measures of activation from task-related fMRI (Friston et al., 1996), the impact of in-scanner motion on measures of functional connectivity was not explored for 16 years after its initial discovery (Biswal et al., 1995). However, since the near-simultaneous publication of three independent reports in early 2012 (Van Dijk et al., 2012; Power et al., 2012; Satterthwaite et al., 2012), it has been increasingly recognized that motion can have a large impact on rsfc-MRI measurements, and can systematically bias inference. This bias is particularly problematic in developmental or clinical populations where motion is correlated with the independent variable of interest (age, diagnosis) (Satterthwaite et al., 2012; Fair et al., 2012), and has resulted in the re-evaluation of numerous published findings.

* Corresponding author.
E-mail address: sattertt@mail.med.upenn.edu (T.D. Satterthwaite).

http://dx.doi.org/10.1016/j.neuroimage.2017.03.020
Accepted 10 March 2017
Available online 14 March 2017

In response to this challenge, there has been a recent proliferation of participant-level confound regression and censoring methods aimed at mitigating the impact of motion on functional connectivity (Yan et al., 2013a; Power et al., 2015). These methods can be broadly grouped into several categories. First, high-parameter confound regression strategies use expansions of realignment parameters or tissue-compartment signals, often including derivative and quadratic regressors (Friston et al., 1996; Satterthwaite et al., 2013; Yan et al., 2013a). Second, principal component analysis (PCA) based methods (CompCor; Behzadi et al., 2007; Muschelli et al., 2014) find the primary directions of variation within high-noise areas defined by anatomy (e.g., aCompCor) or temporal variance (tCompCor). Third, whole-brain independent component analysis (ICA; Beckmann et al., 2005) of single-subject time series has increasingly been used for de-noising, with noise components selected either by a trained classifier (ICA-FIX; Griffanti et al., 2014; Salimi-Khorshidi et al., 2014) or using *a priori* heuristics (ICA-AROMA; Pruim et al., 2015a, 2015b). Fourth, temporal censoring techniques identify and remove (or de-weight) specific volumes contaminated by motion artifact, often followed by interpolation. These techniques include scrubbing (Power et al., 2012, 2014a, 2015), spike regression (Satterthwaite et al., 2013), and de-spiking (Jo et al., 2013; Patel et al., 2014). Censoring techniques have been reported to attenuate motion artifact, but at the cost of a shorter time series and variably reduced degrees of freedom. Fifth, one recent report emphasized the relative merits of spatially-tailored confound regression using local white matter signals (wmLocal; Jo et al., 2013). Finally, the inclusion of global signal regression (GSR) (Macey et al., 2004) in confound regression models remains a source of controversy (Fox et al., 2009; Murphy et al., 2009; Chai et al., 2012; Saad et al., 2012; Yan et al., 2013b; Murphy, in press). While several studies have suggested its utility in de-noising (Fox et al., 2009; Power et al., 2015; Satterthwaite et al., 2013; Yan et al., 2013a), other studies have emphasized the risk of removing a valuable signal (Yang et al., 2014; Hahamy et al., 2014), potentially biasing group differences (Gotts et al., 2013; Saad et al., 2012), or exacerbating distance-dependent motion artifact. Distance-dependent artifact (Power et al., 2012; Satterthwaite et al., 2012) manifests as increased connectivity in short-range connections, and reduced connectivity in long-range connections, which has the potential to impact measures of network topology (Yan et al., 2013b).

Substantial additional work has moved beyond use of realignment parameters and timeseries signal as regressors. Specifically, recent work has suggested that techniques such as MotSim may potentially track more signal variance related to motion (Patriat et al., 2017). Furthermore, while initial work suggested that voxel-wise motion regressors were not advantageous, work by Spisák et al. (2014) suggests that such information can be successfully utilized. Additionally, one paper evaluated the impact of motion on timeseries smoothness (Scheinost et al., 2014), and suggested that uniform smoothing may ameliorate artifact. Finally, recent work has proposed geometric techniques for correcting motion artifact (e.g., median angle correction) (He and Liu, 2012) and investigated prospective correction techniques (Faraji-Dana et al., 2016).

This recent proliferation of de-noising techniques has prompted excitement but also sowed confusion. Unsurprisingly, new de-noising pipelines have often tended to emphasize outcome measures that suggest their relative superiority. As a result, investigators often anecdotally report substantial uncertainty regarding which pipeline should be used. Such uncertainty has been exacerbated by the lack of common outcome measures used across studies, which has hampered direct comparison among pipelines. While one review paper has summarized recent developments in this rapidly-evolving sub-field (Power et al., 2015), systematic evaluation of de-noising pipelines according to a range of benchmarks remains lacking.

Several prior papers have compared some of these confound regression strategies on selected benchmark measures. For example, Yan and colleagues evaluated a range of de-noising strategies based on realignment parameters (e.g., 6P, 12P, 24P), scrubbing, and GSR (Yan et al., 2013c). Subsequently, Pruim et al. (2015a) compared ICA-AROMA to the 24-parameter model, scrubbing, and aCompCor, among other techniques. Building on such work, Burgess et al. (2016) examined the relative added value of mean grayordinate time series regression, which is similar to GSR, as an addition to ICA-based de-noising (ICA-FIX). However, prior work has not directly evaluated several of the most commonly implemented de-noising methods, which combine high-parameter confound regression, GSR, and censoring.

Accordingly, in this report we compare 14 of the most commonly used confound regression strategies in a large (N=393) dataset of adolescents and young adults. Pipelines evaluated include standard techniques, high-parameter confound regression, PCA-based techniques such as aCompCor and tCompcor, ICA-based approaches such as ICA-AROMA, spatially-tailored local white matter regression, and three different censoring techniques (spike regression, de-spiking, and scrubbing); GSR is included in many pipelines as well. It should be emphasized that this is not a comprehensive evaluation of all artifact-control strategies in use, and that models evaluated were limited to a subset of those commonly used at present. Critically, we compare these pipelines according to four intuitive benchmarks, including the residual relationship between functional connectivity and subject motion, the degree of distance-dependent artifact, the identifiability of network structure after de-noising, and the loss of temporal degrees of freedom. As described below, results underscore the relative strengths and weaknesses among these methods, and reveal clear trade-offs among commonly used confound regression approaches.

Materials and methods

Participants and data acquisition

The task-free BOLD data used in this study (N=393) were drawn from the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al., 2014, 2016) on the basis of age, health, and data quality. All participants selected for evaluation were ages 8-22, were free from medical conditions that could impact brain function (Merikangas et al., 2010), lacked gross structural brain abnormalities (Gur et al., 2013), were not taking psychotropic medication at the time of the scan, and had high quality imaging data free of gross motion. In total, N=84 (44 females) participants were not included in this sample due to gross motion, defined as a mean relative RMS (root mean squared) displacement >0.2 mm, or >20 volumes with framewise relative RMS displacement >0.25 mm. The exclusion of participants with gross in-scanner motion allowed us to evaluate the utility of confound regression strategies for the mitigation of artifact due to micro-movements.

Structural and functional subject data were acquired on a 3 T Siemens Tim Trio scanner with a 32-channel head coil (Erlangen, Germany), as previously described (Satterthwaite et al., 2014, 2016). High-resolution structural images were acquired in order to facilitate alignment of individual subject images into a common space. Structural images were acquired using a magnetization-prepared, rapid-acquisition gradient-echo (MPRAGE) T1-weighted sequence (T_R = 1810 ms; T_E = 3.51 ms; FoV = 180 × 240 mm; resolution 1 mm isotropic). Approximately 6 minutes of task-free functional data were acquired for each subject using a blood oxygen level-dependent (BOLD-weighted) sequence (T_R = 3000 ms; T_E = 32 ms; FoV = 192 × 192 mm; resolution 3 mm isotropic; 124 spatial volumes). Prior to scanning, in order to acclimate subjects to the MRI environment and to help subjects learn to remain still during the actual scanning session, a mock scanning session was conducted using a decommissioned MRI scanner and head coil. Mock scanning was accompanied by acoustic recordings of the noise produced by gradient