HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2018 July 01.

Published in final edited form as:

Neuroimage. 2017 July 01; 154: 174–187. doi:10.1016/j.neuroimage.2017.03.020.

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 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], and **Theodore D. Satterthwaite**[a,*]

[a]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA

[b]Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, New York, 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]Department of Clinical Neuroscience and Medical Psychology, Heinrich-Heine University Düsseldorf

[f]Institute of Neuroscience and Medicine (INM-1), Research Center Jülich

[g]Department of Bioengineering, University of Pennsylvania, Philadelphia PA, USA

[h]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia PA, USA

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 young adults. 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-

*Corresponding author: sattertt@mail.med.upenn.edu (Theodore D. Satterthwaite).

offs among methods. First, methods that include global signal regression minimize the relationship between connectivity and motion, but unmask 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 proposed methods, and suggest that different confound regression strategies may be appropriate in the context of specific scientific goals.

Keywords

fMRI; functional connectivity; artifact; confound; motion; noise

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 (Di-Martino 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 reevaluation of numerous published findings.

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. (2015b,a)). 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 (wm-Local; 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 and Fox, 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, recent work has 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, such as 36-parameter confound regression + censoring, using a broad range of benchmark measures.

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. Participants' ages ranged from 8 to 22 years. 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 $> 0{:}2mm$, or > 20 volumes with framewise relative RMS displacement $> 0{:}25mm$. 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 3T 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 × 240mm; resolution 1mm 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 × 192mm; resolution 3mm 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 coils for each scanning pulse sequence. During these sessions, feedback regarding head movement was provided using the MoTrack motion tracking system (Psychology Software Tools, Inc, Sharpsburg, PA). Motion feedback was only given during the mock scanning session. In order to further minimize motion, prior to data acquisition subjects' heads were stabilized in the head coil using one foam pad over each ear and a third over the top of the head. During the resting-state scan, a fixation cross was displayed as images were acquired. Subjects were instructed to stay awake, keep their eyes open, fixate on the displayed crosshair, and remain still.

Structural image processing

A study-specific template was generated from a sample of 120 PNC subjects balanced across sex, race, and age bins using the build Template Parallel procedure in ANTs (Avants et al., 2011a). Study-specific tissue priors were created using a multi-atlas segmentation procedure (Wang et al., 2014). Next, each subject's high-resolution structural image was processed using the ANTs Cortical Thickness Pipeline (Tustison et al., 2014). Following bias field correction (Tustison et al., 2010), each structural image was diffeomorphically registered to the study-specific PNC template using the top-performing SyN deformation (Klein et al., 2009). Study-specific tissue priors were used to guide brain extraction and segmentation of the subject's structural image (Avants et al., 2011b).

BOLD time series processing

Task-free functional images were processed using the XCP Engine (Ciric et al., In Preparation), which was configured to support the 14 pipelines evaluated in this study (see Figure 1). Each pipeline was based on de-noising strategies previously described in the neuroimaging literature. A number of preprocessing procedures were included across all de-noising pipelines. Common elements of preprocessing included (1) correction for distortions induced by magnetic field inhomogeneities using FSL's FUGUE utility, (2) removal of the 4 initial volumes of each acquisition, (3) realignment of all volumes to a selected reference volume using MCFLIRT (Jenkinson et al., 2002), (4) demeaning and removal of any linear or quadratic trends, (5) co-registration of functional data to the high-resolution structural image using boundary-based registration (Greve and Fischl, 2009), and (6) temporal filtering using a first-order Butterworth filter with a passband between 0.01 and 0.08 Hz. We did not apply slice timing correction during preprocessing, as recent data suggest that the interpolation that occurs may artificially reduce motion estimates (Power et al., under review). These preliminary processing stages were then followed by the confound regression procedures described below. In order to prevent frequency-dependent mismatch during confound regression (Hallquist et al., 2013), all regressors were band-pass filtered to retain the same frequency range as the data. As in our prior work (Satterthwaite et al., 2012, 2013), the primary summary metric of subject motion used was the mean relative RMS (root-mean-squared) displacement calculated during time series realignment using MCFLIRT.

Overview of confound regression strategies

The primary objective of the current study was to evaluate the performance of common de-noising strategies. We selected 14 de-noising models, labelled **1**–**14** below, for evaluation (Figure 1). Models **1**–**5** used nuisance parameters derived from 6 movement estimates and 3 physiological time series, as well as their temporal derivatives and quadratic expansions.

- **Model 1.** (2P) Used only the 2 physiological time series: mean signal in WM and mean signal in CSF, and functioned as a base model for comparison to other more complex confound regression models.
- **Model 2.** (6P) Used only the 6 motion estimates derived from MCFLIRT realignment as explanatory variables.
- **Model 3.** (9P) Combined the 6 motion estimates and 2 physiological time series with global signal regression. This model has been widely applied to functional connectivity studies (Fox et al., 2005, 2009).
- **Model 4.** (24P) Expansion of model **2** that includes 6 motion parameters, 6 temporal derivatives, 6 quadratic terms, and 6 quadratic expansions of the derivatives of motion estimates for a total 24 regressors (Friston et al., 1996).
- **Model 5.** (36P) Similar expansion of model **3**: 9 regressors, their derivatives, quadratic terms, and squares of derivatives (Satterthwaite et al., 2013).

Models **6**–**8** further expanded upon this maximal 36P strategy by incorporating censoring approaches.

- **Model 6.** (36P+despike) Included 36 regressors as well as despiking (Cox, 1996).
- **Model 7.** (36P+spkreg) Included 36 regressors as well as spike regression, as in Satterthwaite et al. (2013).
- **Model 8.** (36P+scrub) Included 36 regressors as well as motion scrubbing, as in Power et al. (2014a).

Models **9** and **10** adapted variants of the PCA-based *CompCor* approach.

- **Model 9.** (aCompCor) Used 5 principal components each from the WM and CSF, in addition to motion estimates and their temporal derivatives (Muschelli et al., 2014).
- **Model 10.** (tCompCor) Used 6 principal components from high-variance voxels (Behzadi et al., 2007).

Models **11** and **12** comparatively evaluated the efficacy of local and global-mean tissue-class regressors.

- **Model 11.** (wmLocal) Used a voxelwise, localised WM regressor in addition to motion estimates and their temporal derivatives and despiking (Jo et al., 2013).

- **Model 12.** (wmMean) Identical to model **11** except that it used the mean signal across the WM instead of a voxelwise, localised WM regressor (Jo et al., 2013). Models **13** and **14** evaluated subhect-specific ICA de-noising.
- **Model 13.** (ICA-AROMA) Used a recently developed ICA-based procedure for removal of motion-related variance from BOLD data, together with mean WM and CSF regressors (Pruim et al., 2015a,b).
- **Model 14.** (AROMA+GSR) Combined ICA-AROMA as implemented in model **13** with global signal regression, in a procedure somewhat analogous to Burgess et al. (2016).

We explicitly limited our scope to models that did not require training a classifier, and did not evaluate confound regression strategies that require extensive parameter optimization (Salimi-Khorshidi et al., 2014; Griffanti et al., 2014; Patel et al., 2014). Furthermore, in order to constrain the parameter space, we did not examine unpublished combinations of de-noising approaches.

Confound regression using realignment parameters

Time series of six realignment parameters (three translational and three rotational) for each subject were returned by MCFLIRT as part of time series realignment (motion correction). Additionally, the temporal derivative, quadratic terms, and quadratic of the temporal derivative of each of the realignment parameters were calculated, yielding 24 realignment regressors in total. The original six realignment parameters were included in confound regression models **2** and **3**. Models **9** and **11** included 12 realignment regressors – the 6 realignment parameters and their temporal derivatives – while the full set of 24 expanded realignment regressors were included as part of confound regression models **4**–**8**.

Global signal regression

The mean global signal was computed by averaging across all voxelwise time series located within a subject-specific mask covering the entire brain. The global signal was included in model **3**, while the expanded models **5**–**8** included 4 global signal regressors: the global signal, its derivative, its square, and the derivative of its square.

Tissue class regressors

Mean white matter (WM) and cerebrospinal fluid (CSF) signals were computed by averaging within masks derived from the segmentation of each subject's structural image; these masks were eroded using AFNI's 3dmask tool (Cox, 1996) to prevent inclusion of gray matter signal via partial-volume effects. The WM mask was eroded at the 2-voxel level, while the CSF mask was eroded at the 1-voxel level. More liberal erosion often resulted in empty masks in our data. Temporal derivatives, quadratic terms, and squares of the derivative were computed as above. Two tissue class regressors (WM and CSF) were included in models **3** and **12**, whereas their expansions (8 regressors) were included in models **4**–**8**.

Local white matter regression

Model **11** used a local WM regressor (Jo et al., 2013). This was computed in AFNI using 3dLocalstat (Cox, 1996). Unlike the regressors described above, which were voxel-invariant, the value of the local WM regressor was computed separately at each voxel. For each voxel, a sphere of radius 45mm was first centered on that voxel; this sphere defined that voxel's local neighborhood. Next, this spherical neighborhood was intersected with an eroded WM mask to produce a local WM mask, which included only the fraction of the WM that was also in the voxel's neighborhood. The mean signal within this new local WM mask was then used to model the local WM signal at the voxel (Jo et al., 2013). This process was repeated at every voxel in order to generate the local WM regressor. This local WM regressor was included in model **11** along with realignment parameters and their derivatives (12 total); this model also included voxelwise de-spiking. In order to evaluate the efficacy of this local regressor in comparison with a more typical mean tissue regressor, model **12** was identical to model **11** with a mean WM regressor substituting for the local WM regressor.

CompCor

Principal component analysis (PCA) can be used to model noise in BOLD time series (Behzadi et al., 2007; Muschelli et al., 2014). Broadly, the use of PCA-derived regressors to model noise is called component-based correction (CompCor). Numerous variants of CompCor have been developed; here, our focus will be on anatomical CompCor (aCompCor, model **9**) and temporal CompCor (tCompCor, model **10**). In aCompCor, a PCA is performed within an anatomically defined tissue class of interest. We extracted 5 components for WM and CSF each, yielding 10 compcor components (Muschelli et al., 2014). As part of model **9**, as in Muschelli et al. (2014), these 10 aCompCor components were combined with 12 re-alignment parameters (raw and temporal derivative). In tCompCor, the temporal variance of the BOLD signal is first computed at each voxel. Subsequently, a mask is generated from high-variance voxels, and principal components are extracted from the time series at these voxels. In confound regression model **10**, tCompcor was implemented using ANTs, with 6 tCompCor components used as confound regressors for each participant.

ICA-AROMA

ICA-AROMA (automatic removal of motion artifact) is a recently-introduced, widely-used method for de-noising using single-subject ICA (Pruim et al., 2015a,b); we evaluated ICA-AROMA in confound regression models **13** and **14**. In contrast to other ICA based methods (e.g., ICA-FIX: Salimi-Khorshidi et al. (2014)), it does not require dataset-specific training data. The input to ICA-AROMA is a voxelwise time series that has been smoothed at 6mm FWHM using a Gaussian kernel. After decomposing this time series using FSL's MELODIC (with model order estimated using the Laplace approximation) (Beckmann et al., 2005), ICA-AROMA uses four features to determine whether each component corresponds to signal or noise. The first 2 features are spatial characteristics of the signal source: (1) the fraction of the source that falls within a CSF compartment and (2) the fraction of the source that falls along the edge or periphery of the brain. The remaining features are derived from the time series of the source: (3) its maximal robust correlation with time series derived from

realignment parameters and (4) its high-frequency spectral content. ICA-AROMA includes two de-noising steps. The first de-noising step occurs immediately after classification. All component time series (signal and noise) are included as predictors in the linear model, and the residual BOLD time series is obtained via partial regression of only the noise time series. A second confound regression step occurs after temporal filtering, wherein mean signals from WM and CSF (models **13** and **14**) and the global signal (model **14** only) were regressed from the data.

Temporal censoring: de-spiking, spike regression, and scrubbing

In addition to regression of nuisance time series, a number of 'temporal censoring' approaches were used to identify motion-contaminated volumes in the BOLD time series and reduce their impact on further analysis. These approaches included despiking, spike regression, and scrubbing. Despiking is a procedure that identifies outliers in the intensity of each voxel's detrended BOLD time series and then interpolates over these outliers. Despiking was implemented in AFNI using the 3dDespike utility (Cox, 1996) as part of confound regression model **6**.

Unlike despiking, which identifies outliers on a voxel-wise basis, spike regression and scrubbing censor complete volumes based on metrics of subject movement defined *a priori*. For spike regression, as in Satterthwaite et al. (2013), volumes were flagged for spike regression if their volume-to-volume RMS displacement exceeded 0.25mm. Next, as part of confound regression model **7**, *k* 'spike' regressors were included as predictor variables in the de-noising model, where *k* equalled the number of volumes flagged (Satterthwaite et al., 2013). For each flagged time point, a unit impulse function that had a value of 1 at that time point and 0 elsewhere was included as a spike regressor.

For scrubbing, the framewise displacement (FD) (Power et al., 2012) was computed at each time point as the sum of the absolute values of the derivatives of translational and rotational motion estimates. If framewise displacement (FD) at any point in time exceeded 0.2mm, then that time point was flagged for scrubbing. It should be noted that the conversion of FD to RMS displacement is approximately 2:1, and thus the published criterion for scrubbing has a lower threshold for flagging high-motion volumes than does spike regression. Scrubbing of BOLD data was performed iteratively (Power et al., 2014a) as part of confound regression model **8**. At any stage where a linear model was applied to the data (for instance, during detrending procedures), high-motion epochs were temporally masked out of the model so as not to influence fit. During temporal filtering, a frequency transform was used to generate surrogate data with the same phase and spectral characteristics as the unflagged data. This surrogate data was used to interpolate over flagged epochs prior to application of the filter. During confound regression, flagged timepoints were excised from the time series so as not to contribute to the model fit. For scrubbing (but not spike regression) if fewer than five contiguous volumes had unscrubbed data, these volumes were scrubbed and interpolated as well.

Overview of outcome measures

We evaluated each de-noising pipeline according to four benchmarks. Residual *QC-FC correlations* and *distance-dependence* provided a metric of each pipeline's efficacy, while *loss of temporal DOF* provided an estimate of each pipeline's efficiency. Finally, the *modularity quality* provided an estimate of network identifiability after de-noising.

Relationship between mean relative RMS displacement and functional connectivity (QC-FC correlations)

In order to estimate the residual relationship between subject movement and connectivity after de-noising, we computed *QC-FC* correlations (quality control / functional connectivity) (Power et al., 2015; Satterthwaite et al., 2012, 2013; Power et al., 2012). While other metrics have been used in prior reports, including FD-DVARS correlations, we favor *QC-FC* as the most useful metric of interest as it directly quantifies the relationship between motion and the primary outcome of interest (rather than two quality metrics, as in FD-DVARS). For an extended discussion of the rationale for this measure, see Power et al. (2015).

We evaluated *QC-FC* relationships within two commonly-used whole-brain networks, the first consisting of spherical nodes distributed across the brain (Power et al., 2011) and the second comprising an areal parcellation of the cerebral cortex (Gordon et al., 2016). For each network, the mean time series for each node was calculated from the denoised residual data, and the pairwise Pearson correlation coefficient between node time series was used as the network edge weight (Biswal et al., 1995). For each edge, we then computed the correlation between the weight of that edge and the mean relative RMS motion. To eliminate the potential influence of demographic factors, *QC-FC* relationships were calculated as partial correlations that accounted for participant age and sex. We thus obtained, for each de-noising pipeline, a distribution of *QC-FC* correlations. This distribution was used to obtain two measures of the pipeline's ability to mitigate motion artifact, including: 1) the number of edges significantly related to motion, which was computed after using the false discovery rate (FDR; Benjamini and Hochberg (1995)) to account for multiple comparisons; and 2) the median absolute value of all *QC-FC* correlations. All graphs were generated using ggplot2 in R version 3.2.3 (Wickham, 2009); brain renderings were prepared in BrainNet Viewer (Xia et al., 2013).

Distance-dependent effects of motion

Early work on motion artifact demonstrated that in-scanner motion can bias connectivity estimates between two nodes in a manner that is related to the distance between those nodes (Satterthwaite et al., 2012; Power et al., 2012). Under certain processing conditions, subject movement enhances short-distance connections while reducing long-distance connections. To determine the residual distance-dependence of subject movement, we first used the center of mass of each node to obtain a distance matrix D where entry D_{ij} indicates the Euclidean distance between the centers of mass of nodes i and j. We then obtained the correlation between the distance separating each pair of nodes and the *QC-FC* correlation of the edge connecting those nodes; this correlation served as an estimate of the distance-dependence of motion artifact.

Network modularity

Including additional regressors in a confound model has the potential to remove real signal in addition to motion-related noise. In order to evaluate this possibility, we computed modularity quality (Q), an explicit quantification of the degree to which there are structured sub-networks in a given network, in this case the de-noised connectome. Prior work has demonstrated a relationship between Q and subject motion (Satterthwaite et al., 2012) that is mitigated by participant-level de-noising approaches (Satterthwaite et al., 2013). If confound regression and censoring were removing real signal in addition to motion-related noise, we expect that Q would decline. To determine Q, community detection was performed on each subject's de-noised network using the Louvain heuristic (Blondel et al., 2008), which partitions the connectome into sub-networks in a manner that maximizes the value of Q. As functional connectomes included positive and negative weights, we used a version of the Louvain algorithm that accommodates signed data (Rubinov and Sporns, 2010), and did not threshold connectivity matrices. Because this approach to community detection is degenerate, a consensus partition was obtained over 100 Louvain optimizations (Lancichinetti and Fortunato, 2012). Finally, the modularity quality of the resultant consensus partition was estimated according to an established null model (Girvan and Newman, 2002); the mean of Q values across subjects provided an estimate of the sub-network definition still present in the connectome after de-noising. Furthermore, we also assessed whether (and to what extent) Q was correlated with motion; to do this, we computed the Pearson correlation coefficient between subjects' Q values and their motion estimates. These relationships were computed as partial correlations that accounted for participant age and sex.

Additional degrees of freedom lost in confound regression

Confound regressors and censoring both reduce the temporal degrees of freedom (DOF) in data. This loss in temporal DOF may introduce bias if it is variable across subjects. While removal of temporal DOF reduces the number of observations that sample the connectome, the current analysis is not biased by the concatenation of temporally discontinuous time series (i.e., after censoring) because the current analysis (1) uses a time-invariant measure of connectivity that is not dependent on any temporal autocorrelation structure (i.e., Pearson correlation) and (2) applies any procedures that are dependent upon temporal autocorrelation structure (e.g., temporal filtering) prior to concatenation. De-noising strategies ideally limit the loss of temporal DOF, for instance by including fewer, more efficacious regressors. In the present study, we assessed the number of temporal DOF lost in each confound regression approach.

As in previous work (Pruim et al., 2015a), we assumed that each time series regressed out and each volume excised from the data constituted a single temporal DOF. Consequently, the loss of temporal DOF was estimated as the sum of the number of regressors in each confound model and the number of volumes flagged for excision under that model. It should be emphasized that the values thus obtained are imperfect estimates. First, because functional MR time series typically exhibit temporal autocorrelation, the actual loss in DOF will be less than the estimated loss in DOF. Accordingly, censoring adjacent volumes does not remove the same number of DOF as does censoring volumes separated in time.

Furthermore, a temporal bandpass filter was uniformly applied to all data prior to confound regression; this filtering procedure would itself have removed additional temporal DOF and elevated the autocorrelation of the data. Because this filter was uniform across all de-noising strategies, it was not considered when estimating the loss of additional temporal DOF in each model.

Results

Heterogeneity in confound regression performance

Confound regression strategies typically remove some, but not all, of the artifactual variance that head motion introduces into the BOLD signal. The motion-related artifact that survives de-noising can be quantified to provide a metric of pipeline performance. Here, our primary benchmark of confound regression efficacy was the residual relationship between brain connectivity and subject motion, or the *QC-FC correlation*. We measured *QC-FC* correlations using two metrics: the percentage of network connections where a significant relationship with motion was present (Figure 2), and the absolute median correlation (AMC) between connection strength and head movement across all connections (Figure 3).

No preprocessing strategy was completely effective in abolishing the relationship between head movement and connectivity. However, different approaches exhibited widely varying degrees of efficacy. The top four confound regression strategies included 36 parameters, comprising an expansion of GSR, tissue-specific regressors (WM, CSF), and realignment parameters. Beyond this base 36-parameter model, all censoring techniques provided provided some additional benefit, reducing the number of edges that were significantly related to motion to less than 7%. Convergent results were present across both *QC-FC* measures (% edges, AMC) and networks (Power, Gordon) that were evaluated. The top-performing method, GSR+spike regression, yielded $<1\%$ of edges that were significantly related to motion.

In contrast, many pipelines performed relatively poorly, leaving a majority of network edges with a residual relationship with motion. Specifically, 89% of edges were impacted by motion when the least effective method was used (6 realignment parameters). The commonly used 24-parameter expansion of realignment parameters originally suggested by Friston et al. (1996) did not provide much of an improvement (88% edges). Similarly, the local WM regressor model (77% edges) and tCompCor 70% edges) also resulted in substantial residual *QC-FC* correlations. In fact, these methods performed worse than a basic 2-parameter model composed of mean WM and CSF signals (44% edges). Notably, a local WM signal (77% edges) did not provide any benefit over the mean WM signal (39% edges) according to QC-FC metrics, and in fact performed considerably worse. Finally, several methods demonstrated intermediate performance, with 1–20% of edges impacted by motion. This middle group included methods as disparate as aCompCor (13% edges), ICA-AROMA (28% edges), ICA-AROMA with GSR (10% edges), and the classic 9-parameter confound regression model which included GSR (13% edges).

Variability in distance-dependent motion artifact after confound regression

Our second benchmark quantified the distance-dependent motion artifact that was present in data processed by each pipeline (Figure 4). We observed that distance-dependence was present even under conditions where artifact magnitude was attenuated. For example, though the 36-parameter model was among the most effective in attenuating *QC-FC* relationships, its application revealed strongly distance-dependent artifact. Examination of graphs that plot *QC-FC* by Euclidean distance (see Figure 4C) revealed that this is due to effective mitigation of motion artifact for long-range but not short-range connections.

Distance-dependence was highly prominent in models that included GSR, but did not include censoring (e.g., 9-parameter and 36-parameter models). However, despite the lack of global signal in the aCompCor and tCompcor models, data returned from both of these component-based approaches revealed substantial distance-dependent artifact. Notably, inclusion of censoring consistently reduced distance-dependence, although scrubbing was more effective than spike regression or voxelwise despiking.

The top performing method according to this benchmark was ICA-AROMA, which completely abolished any distance-dependence of residual motion artifact. In other words, the motion artifact that was still present in the data after ICA-AROMA impacted all connections in a manner that was not dependent on the spatial separation between nodes. Augmenting ICA-AROMA with GSR decreased *QC-FC* correlations but exposed distance-dependent artifact, suggesting that ICA-AROMA did not completely remove long-distance motion artifact.

There was similar lack of distance-dependence in the wmLocal model, although as noted above this model did not provide effective de-noising according to *QC-FC* benchmarks. Use of a local tissue regressor revealed less distance-dependent artifact than did the whole-tissue regressor.

Confound regression strategies mitigate the impact of motion on network modularity

We next evaluated the degree to which de-noising strategies impacted sub-network identifiability, which was operationalized as the network modularity quality (*Q*; Figure 5). First and most notably, the 4 models that exhibited the poorest performance according to *QC-FC* measures (6P, 24P, wmLocal, and tCompCor) also suffered from an inability to identify structured functional sub-networks of the brain. This suggests that motion artifact impedes network identifiability. Second, the 36-parameter models did not uniformly outperform lower order models. However, addition of any of the three censoring techniques evaluated (scrubbing, spike regression, or despiking) provided an improvement over and above the 36-parameter base model. Third, both AROMA and aCompCor performed well, at levels similar to that observed for 36-parameter models with censoring. However, addition of GSR to AROMA did not improve the observed *Q* value. Fourth and finally, the top-performing model was in fact the 9P model, which had a higher mean *Q* value than all 36-parameter models and ICA-AROMA.

To ascertain whether network identifiability was systematically impacted by motion, we also evaluated the correlation between modularity quality and motion for each de-noising

approach. In general, 36-parameter models and ICA-based models most effectively decoupled modular structure from subject motion, while models with high residual *QC-FC* correlations likewise left high correlations between motion and modularity (Figure 6). Notably, the 9P model that displayed the highest mean Q value nonetheless retained a substantial relationship with motion, as expected from the edgewise *QC-FC* analyses. Furthermore, addition of GSR to AROMA successfully reduced the relationship between motion and modularity.

Effective preprocessing strategies use many additional degrees of freedom

Perhaps unsurprisingly, the preprocessing strategies that consistently reduced both *QC-FC* correlations and distance-dependence were also among the costliest in terms of loss of temporal degrees of freedom (Figure 7). By definition, the 36-parameter models included a high fixed number of regressors. Furthermore, models that additionally included censoring resulted in a substantial additional loss of data that varied across subjects. (Because the 36P +despike model censors data in a spatially adaptive manner, the DOF loss varied by voxel. Because of this spatial variability, the DOF loss is not explicitly estimated for this model.) ICA-AROMA also had a variable loss of DOF, but of a lower magnitude than censoring or high-parameter confound regression.

Discussion

In response to rapid evolution of confound regression strategies available for the mitigation of motion artifact, in this report we evaluated 14 commonly used pipelines. Results indicate that there is substantial heterogeneity in the performance of these confound regression techniques across all measures evaluated. The context, implications, and limitations of these findings are discussed below.

Confound regression techniques have substantial performance variability

We evaluated confound regression strategies according to four intuitive benchmarks that were selected to capture different domains of effectiveness. These included *QC-FC* associations, distance-dependence of motion artifact, modularity quality and its association with motion, and additional degrees of freedom lost in confound regression. Across each benchmark, there was a striking heterogeneity in pipeline performance. While no model completely abolished motion-related variance, 36-parameter models with censoring (Satterthwaite et al., 2013; Power et al., 2014a) performed well across a range of benchmarks, as did ICA-AROMA with GSR.

Notably, in terms of limiting *QC-FC* relationships, the top six confound regression approaches all included GSR. This effect was consistent in both networks we evaluated. The effectiveness of GSR is most likely due to the nature of motion artifact itself: in-scanner head motion tends to induce widespread reductions in signal intensity across the entire brain parenchyma (see Satterthwaite et al. (2013), Figure 4). As discussed in detail elsewhere (Power et al., 2016) this effect is highly reproducible across datasets, and is effectively captured by time series regression of the global signal.

For studies of individual difference where motion may be a substantial confounding factor, our results clearly support the use of models that utilize GSR. Across nearly every benchmark, models without GSR underperformed relative to similar models that included GSR. Whereas *QC-FC* correlations were relatively zero-centered for the case of GSR-based models, most models that omitted GSR – but particularly those that included realignment parameters alone – exhibited distributions that were shifted strongly to the right. Furthermore, augmentation of GSR-based models using either scrubbing or ICA-AROMA yielded better performance than RP-based models across all benchmarks. Thus, while our results cannot advocate for any single model in all scenarios, they suggest that (1) GSR is likely to be the single most effcacious strategy for de-noising and (2) RP-based models are comparably in-effective at de-noising. These results are convergent with recent data regarding the role of motion and physiological artifact in global signal (Power et al., 2016).

Beyond GSR, a second strategy that clearly minimizes *QC-FC* relationships is temporal censoring. We evaluated three censoring variants, including scrubbing, spike regression, and de-spiking. Compared to spike regression and de-spiking, scrubbing appears to be more effective in removing distance-dependent artifact in this dataset. This is most likely due to the explicit tension between data quality and data quantity: because of the lower threshold for scrubbing than spike regression (due to differences in FD vs. RMS measures of motion; see Figure 9C in Yan et al. (2013a)), more low-quality data was excised during scrubbing. Furthermore, scrubbing includes a criterion to not leave isolated epochs (<5 volumes) of un-scrubbed data. Consequently, this leads to clear differences in the additional degrees of freedom lost by each method. In contrast to spike regression and scrubbing, which eliminate high motion volumes completely, time series de-spiking identifies and interpolates large changes in signal intensity on a voxelwise basis (Cox, 1996). This allows for spatial adaptivity (see Patel et al. (2014)) but also renders quantification of data loss and comparisons with volume-based censoring techniques more difficult.

Loss of temporal degrees of freedom should be interpreted with caution because temporal DOF may correspond in part or in whole to artifactual sources rather than signal of interest. For example, although removal of volumes contaminated by motion via censoring results in reduced temporal DOF, it improves network identification relative to the same model without any censoring. This can be explained by considering that the temporal DOF in each time series can be either primarily signal or primarily noise DOF; the DOF that censoring removes are primarily noise, and their removal thus increases the overall signal-to-noise ratio of the time series.

Another DOF-related concern about motion censoring is the potential for variable loss of temporal DOF to bias group-level analyses. One proposed solution involves excising a uniform number of volumes from all subject time series, regardless of the number of motion-contaminated frames. However, if motion results in more noise DOF, then the number of useful signal DOF is variable from the start. The extent to which each de-noising strategy removes temporal DOF (through either confound regression or volume excision) should thus not be considered in isolation, but in concert with the ability of that strategy to identify meaningful signal in the data, for instance as evidenced by the network identifiability benchmark.

Critically, while both GSR and censoring appeared effective in minimizing *QC-FC* relationships, they exhibited opposite effects on distance-dependence. While censoring techniques appear to consistently reduce the presence of distance-dependence, GSR is associated with increased distance-dependence. Thus, commonly used models that include GSR (9-parameter, 36-parameter) have among the greatest distance-dependence of the models we evaluated. However, it should be emphasized that the distance-dependence associated with GSR is not the result of worsening associations with motion in certain connections. Rather, the distance-dependence seen with GSR stems from differential de-noising efficacy, whereby motion artifact is more effectively minimized for long-distance connections than for short-range connections. Certain models such as the local WM regression approach (Jo et al., 2013) thus have minimal distance-dependence, but this is a consequence of lack of efficacy across all distances. In contrast, ICA-AROMA (Pruim et al., 2015b,a) reduced motion to a moderate degree over all connection distances, resulting in almost no distance-dependence. However, while clearly an improvement over some other methods, data processed using ICA-AROMA was noisier than other methods which included GSR or censoring, and resulting networks contained a substantial number of edges impacted by motion. As suggested by the work of Burgess et al. (2016), adding GSR to ICA-AROMA mitigates *QC-FC* relationships, but as expected exacerbates distance-dependence.

Somewhat to our surprise, benchmark results for aCompCor (Behzadi et al., 2007; Muschelli et al., 2014) were most similar to models that included GSR. Alone among models where GSR was not included, aCompCor both was relatively effective in the mitigation of residual motion (13% of edges impacted) and also exhibited substantial distance-dependence (e.g., $r = -0{:}26$). This suggests that while aCompCor does not explicitly include GSR, the practical results of its application are in fact quite similar.

Generally, the models that performed poorly in terms of residual *QC-FC* correlations also did not perform well in terms of ability to identify structured subnetworks in the connectome. This suggests that the removal of noise by more effective methods also unmasks structure. Although censoring substantially reduced the temporal degrees of freedom in the time series, it improved network identifiability in relation to the 36-parameter model alone, suggesting the possibility that the lost temporal degrees of freedom were largely contaminated and did not contain useful information regarding network topology.

Somewhat to our surprise, the very simple two-parameter model (WM, CSF) outperformed commonly used models based on re-alignment parameters alone (e.g., 6-parameter, 24-parameter), suggesting that the relative values of tissue signal regressors and realignment parameters are not equal. This is likely because the WM and CSF regressors capture, to varying extents, the global signal changes that are strongly associated with motion (Power et al., 2016).

Trade-offs of confound regression approaches: implications for investigators

The current results emphasize two clear trade-offs in the choice of confound regression strategy. First, pipelines that include global signal regression tend to be more effective at minimizing *QC-FC* relationships, but at the cost of some increase in distance-dependence.

As noted above, for minimizing *QC-FC* relationships, nearly all of the top strategies (except aCompCor) included GSR. Conversely, the two techniques that had the most substantial distance-dependence (the 9-regressor and 36-regressor methods) both included GSR. Second, censoring techniques provide a clear benefit in reducing *QC-FC* relationships and additionally tend to attenuate distance-dependence. However, by definition, removing contaminated volumes results in less data and loss of degrees of freedom.

These trade-offs suggest that a single confound regression strategy is unlikely to be optimal for every study. For example, in studies of network organization, network identifiability may be of primary interest. Somewhat to our surprise, the classic 9P model displayed the highest network modularity, and thus remains a good choice for many such studies. However, the presence of anti-correlations, altered degree distribution (Yan et al., 2013b), and distance-dependent impact of motion that occurs with GSR-based models (including 9P) may make models without GSR more appealing for certain studies of network organization. In these cases, ICA-AROMA appears to be an excellent choice, as it has high network identifiability and low distance dependence.

In contrast, for studies of group or individual differences, minimizing *QC-FC* relationships is likely to be of paramount importance so as to limit the possibility that inference is driven by artifactual signals. This concern is particularly relevant for studies of brain development or clinical sub-groups where motion is systematically related to the subject-level variable of interest (e.g., age, disease status). For such studies, models that include GSR tend to perform best, including 36P+censoring models and ICA-AROMA + GSR. Our results accord with Burgess et al. (2016) and suggest that ICA-based de-noising alone without GSR does not provide maximal control of motion artifact. Co-varying for motion at the group level is unlikely to be a panacea for such studies when inadequate subject-level time series de-noising is employed, as prior work (Power et al., 2014a) has suggested that motion effects at the group level may potentially both be nonlinear and vary across sub-samples in a manner that is difficult to predict. However, aggressive volume censoring may be problematic in datasets with relatively brief acquisitions. In datasets where long time series are acquired, such as multi-band acquisitions (Feinberg et al., 2010) and intensive acquisitions of single subjects (Laumann et al., 2015), loss of temporal degrees of freedom is less likely to be a major concern. The 36-parameter models without volume censoring offer uniformity, as does randomly or systematically censoring additional volumes until all subjects retain approximately the same degrees of freedom.

Limitations

Several limitations of the current approach should be noted. One of the principal challenges in evaluating the performance of de-noising approaches is the lack of a noise-free ground truth. Our primary benchmark of confound regression performance assumes that mitigation of the relationship between QC (participant motion) and FC (i.e., the imaging measurement) is desirable. To the degree that in-scanner motion itself represents a biologically informative phenotype, this approach will mistake signal for noise. Indeed, prior data suggests that this may sometimes be the case. For example, Zeng et al. (2014) found specific changes in connectivity for participants who had generally high levels of motion, even on scans where

motion was low. However, without multiple scans to allow such careful dissociation, most studies are incapable of disambiguating the large confounding effects of motion on connectivity. Second, in place of *QC-FC* relationships, one could focus on alternative benchmarks such as test-retest reliability (Zuo et al., 2014). Reliability is certainly of interest, but to the degree that motion tends to be highly correlated within individuals across scanning sessions, there is a substantial potential for the presence of consistent motion artifact across sessions to artificially inflate reliability, and diminish the biological relevance of observed results. A third and related concern is that certain de-noising methods could conceivably both minimize *QC-FC* relationships and even enhance reliability by aggressively removing both signal and noise, but in the process diminish sensitivity to meaningful individual differences. Indeed, one prior study demonstrated the association between canonical resting state networks and randomly generated confound parameters (Bright and Murphy, 2015). This concern is somewhat mitigated by prior work, which suggests that higher-order confound regressors improve the confound regression model fit (Yan et al., 2013a; Satterthwaite et al., 2013), while random regressors do not (see Figure 8 in Satterthwaite et al. (2013)). Furthermore, our results suggest that sub-network modularity is actually improved by effective de-noising. Fourth, while our evaluation included many of the most commonly used techniques, other approaches which require substantial training or parameter selection (i.e., ICA-FIX (Salimi-Khorshidi et al., 2014; Griffanti et al., 2014), wavelet de-spiking (Patel et al., 2014)) may be valuable and merit further consideration. Fifth, it is unknown whether the present results generalize to other datasets, which may have different acquisition parameters including much longer timeseries (Laumann et al., 2016) and multiband acquisition (Feinberg et al., 2010). In particular, the relatively short (6 min) scan time used in the current study is a potential limitation; replication of results in longer acquisitions would enhance confidence in the generalizability of the present results. However, the similar findings reported by Burgess et al. (2016)), which used the longer, multi-band time series data from the Human Connectome Project, suggest convergence with the present results. Furthermore, it is possible that the structure of motion artifact is population-dependent. In the present study, we examined the efficacy of de-noising strategies in a sample of young adults and adolescents. Consequently, the conclusions that we present here may not necessarily generalize to other populations, underscoring the importance of evaluating and reporting the residual confounding effects of motion in all studies of functional connectivity. While the current study was conducted in a combined sample of adolescents and young adults, it did not explore the extent to which the efficacy of each denoising strategy was age- or population dependent, which merits additional investigation in the future. Sixth and finally, it should be noted that while improvements in image acquisition (including multi-echo techniques) may not salvage existing motion-contaminated data, it is likely that they will change the methodological landscape of connectivity research moving forward (Kundu et al., 2012, 2013; Bright and Murphy, 2013).

Conclusions

Taken together, the present results underline the performance heterogeneity of recently-introduced, commonly-used confound regression methods. In selecting among these methods, investigators should be aware of the relative strengths and weaknesses of each approach, and understand how processing strategy may impact inference. Clearly, the

relative merit of each approach will vary by research question and study design. Perhaps most importantly, as has been emphasized in nearly every other study of motion artifact, the choice of confound regression strategy is often dwarfed in importance by the need to transparently report and evaluate the impact of motion in each dataset. At a minimum, this includes reporting the relationship between motion artifact and not only subject phenotypes (e.g., group, age, symptom or cognitive score) but also the functional connectivity measures being considered. In the context of such data, the distinction between observed results and the impact of motion artifact can be understood. Such transparency bolsters confidence in reported findings, but also will likely tend to emphasize the remaining challenges for de-noising going forward. Especially when considered in the context of the the rapid evolution of available techniques since 2012, there is no doubt that innovations in post-processing confound regression strategies will continue.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks to the acquisition and recruitment team, including Karthik Prabhakaran and Jeff Valdez. Thanks to Chad Jackson for data management and systems support. Thanks to Monica Calkins for phenotyping expertise. Supported by grants from the National Institute of Mental Health: R01MH107703 (TDS), R01MH107235 (RCG), and R01NS089630 (CD). The PNC was funded through NIMH RC2 grants MH089983, and MH089924 (REG). Additional support was provided by R21MH106799 (DSB & TDS), R01MH101111 (DHW), K01MH102609 (DRR), P50MH096891 (REG), R01NS085211 (RTS), and the Dowshen Program for Neuroscience. DSB acknowledges support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Army Research Laboratory and the Army Research Office through contract numbers W911NF-10-2-0022 and W911NF-14-1-0679, the National Institute of Mental Health (R01-DC-009209-11), the National Institute of Child Health and Human Development (R01HD086888-01), the Office of Naval Research, and the National Science Foundation (BCS-1441502 and PHY-1554488). Support for developing statistical analyses (RTS & TDS) was provided by a seed grant by the Center for Biomedical Computing and Image Analysis (CBICA) at Penn. Data deposition: The data reported in this paper have been deposited in database of Genotypes and Phenotypes (dbGaP), www.ncbi.nlm.nih.gov/gap (accession no. phs000607.v1.p1).

Literature cited

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage. 2011a; 54:2033– 2044. [PubMed: 20851191]

Avants BB, Tustison NJ, Wu J, Cook PA, Gee JC. An open source multivariate framework for N-tissue segmentation with evaluation on public data. Neuroinformatics. 2011b; 9:381–400. [PubMed: 21373993]

Baker JT, Holmes AJ, Masters Ga, Yeo BTT, Krienen F, Buckner RL, Öngür D. Disruption of cortical association networks in schizophrenia and psychotic bipolar disorder. JAMA psychiatry. 2014; 71:109–18. arXiv:15334406. [PubMed: 24306091]

Beckmann CF, Deluca M, Devlin JT, Smith SM. Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci. 2005; 360:1001–13. [PubMed: 16087444]

Behzadi Y, Restom K, Liau J, Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage. 2007; 37:90–101. arXiv:NIHMS150003. [PubMed: 17560126]

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995; 57:289–300.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine. 1995; 34:537– 541. [PubMed: 8524021]

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008; 10008:6. arXiv: 0803.0476.

Bright MG, Murphy K. Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. NeuroImage. 2013; 64:526–537. [PubMed: 23006803]

Bright MG, Murphy K. Is fMRI "noise" really noise? Resting state nuisance regressors remove variance with network structure. NeuroImage. 2015; 114:158–169. [PubMed: 25862264]

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: Anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences. 2008; 1124:1–38. [PubMed: 18400922]

Burgess GC, Kandala S, Nolan D, Laumann TO, Power JD, Adeyemo B, Harms MP, Petersen SE, Barch DM. Evaluation of denoising strategies to address motion-correlated magnetic resonance imaging data from the Human Connectome Project. Brain Connectivity. 2016

Chai XJ, Castañan AN, Öngür D, Whitfield-Gabrieli S. Anticorrelations in resting state networks without global signal regression. NeuroImage. 2012; 59:1420–1428. [PubMed: 21889994]

Cox R. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 1996; 29:162–173. [PubMed: 8812068]

Craddock RC, Jbabdi S, Yan CG, Vogelstein JT, Castellanos FX, Di Martino A, Kelly C, Heberlein K, Colcombe S, Milham MP. Imaging human connectomes at the macroscale. Nature Methods. 2013; 10:524–539. arXiv:NIHMS150003. [PubMed: 23722212]

Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. Consistent resting-state networks across healthy subjects. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:13848–53. [PubMed: 16945915]

DiMartino A, Fair DA, Kelly C, Satterthwaite TD, Castellanos FX, Thomason ME, Craddock RC, Luna B, Leventhal BL, Zuo XN, Milham MP. Unraveling the miswired connectome: A developmental perspective. Neuron. 2014; 83:1335–1353. [PubMed: 25233316]

Dosenbach NUF, Nardos B, Cohen AL, Fair DA, Power D, Church Ja, Nelson SM, Wig GS, Vogel AC, Lessovschlaggar CN, Barnes KA, Dubis JW, Feczko E, Coalson RS Jr, Barch JRP, Petersen DM, Schlaggar SE, BL. Prediction of individual brain maturity using fMRI. Science. 2011; 329:1358–1361.

Fair DA, Cohen AL, Dosenbach NUF, Church JA, Miezin FM, Barch DM, Raichle ME, Petersen SE, Schlaggar BL. The maturing architecture of the brain's default network. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105:4028–4032. [PubMed: 18322013]

Fair DA, Nigg JT, Iyer S, Bathula D, Mills KL, Dosenbach NUF, Schlaggar BL, Mennes M, Gutman D, Bangaru S, Buitelaar JK, Dickstein DP, Di Martino A, Kennedy DN, Kelly C, Luna B, Schweitzer JB, Velanova K, Wang YF, Mostofsky S, Castellanos FX, Milham MP. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. Frontiers in systems neuroscience. 2012; 6:80. [PubMed: 23382713]

Fair DA, Posner J, Nagel BJ, Bathula D, Dias TGC, Mills KL, Blythe MS, Giwa A, Schmitt CF, Nigg JT. Atypical default network connectivity in youth with attention-deficit/hyperactivity disorder. Biological psychiatry. 2010; 68:1084–91. [PubMed: 20728873]

Faraji-Dana Z, Tam F, Chen JJ, Graham SJ. A robust method for suppressing motion-induced coil sensitivity variations during prospective correction of head motion in fMRI. Magnetic Resonance Imaging. 2016; 34:1206–1219. [PubMed: 27451407]

Feinberg DA, Moeller S, Smith SM, Auerbach EJ, Ramanna S, Glasser MF, Miller KL, Ugurbil K, Yacoub ES. Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. PLoS ONE. 2010; 5:e15710. [PubMed: 21187930]

Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci. 2007; 8:700–711. [PubMed: 17704812]

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:9673–8. arXiv:NIHMS150003. [PubMed: 15976020]

Fox MD, Zhang D, Snyder AZ, Raichle ME. The global signal and observed anticorrelated resting state brain networks. Journal of neurophysiology. 2009; 101:3270–3283. [PubMed: 19339462]

Friston K, Williams S, Howard R, Frackowiak RSJ, Turner R. Movement-Related Effects in –fMRI} Time-Series. Magnetic Resonance in Medicine. 1996; 35:346–355. [PubMed: 8699946]

Girvan M, Newman MEJ. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:7821–7826. arXiv: 0112110. [PubMed: 12060727]

Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and evaluation of a cortical area parcellation from resting-state correlations. Cerebral Cortex. 2016; 26:288–303. [PubMed: 25316338]

Gotts SJ, Saad ZS, Jo HJ, Wallace GL, Cox RW, Martin A. The perils of global signal regression for group comparisons: a case study of Autism Spectrum Disorders. Frontiers in Human Neuroscience. 2013; 7:356. [PubMed: 23874279]

Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. NeuroImage. 2009; 48:63– 72. [PubMed: 19573611]

Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, Zsoldos E, Ebmeier KP, Filippini N, Mackay CE, Moeller S, Xu J, Yacoub E, Baselli G, Ugurbil K, Miller KL, Smith SM. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. NeuroImage. 2014; 95:232–247. [PubMed: 24657355]

Gur RE, Kaltman D, Melhem ER, Ruparel K, Prabhakaran K, Riley M, Yodh E, Hakonarson H, Satterthwaite T, Gur RC. Incidental findings in youths volunteering for brain MRI research. American Journal of Neuroradiology. 2013; 34:2021–2025. [PubMed: 23811972]

Hahamy A, Calhoun V, Pearlson G, Harel M, Stern N, Attar F, Malach R, Salomon R. Save the global: global signal connectivity as a tool for studying clinical populations with functional magnetic resonance imaging. Brain connectivity. 2014; 4:395–403. [PubMed: 24923194]

Hallquist MN, Hwang K, Luna B. The nuisance of nuisance regression: Spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. NeuroImage. 2013; 82:208–225. arXiv:NIHMS150003. [PubMed: 23747457]

He H, Liu TT. NeuroImage A geometric view of global signal confounds in resting-state functional MRI. NeuroImage. 2012; 59:2339–2348. [PubMed: 21982929]

Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage. 2002; 17:825–841. arXiv:arXiv: 1011.1669v3. [PubMed: 12377157]

Jo HJ, Gotts SJ, Reynolds RC, Bandettini PA, Martin A, Cox RW, Saad ZS. Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state FMRI. Journal of Applied Mathematics. 2013; 2013

Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage. 2009; 46:786–802. [PubMed: 19195496]

Kundu P, Brenowitz ND, Voon V, Worbe Y, Vértes PE, Inati SJ, Saad ZS, Bandettini PA, Bullmore ET. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. Proceedings of the National Academy of Sciences of the Unites States of America. 2013; 110:16187–16192. arXiv:arXiv:1408.1149.

Kundu P, Inati SJ, Evans JW, Luh WM, Bandettini PA. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. NeuroImage. 2012; 60:1759–1770. [PubMed: 22209809]

Lancichinetti A, Fortunato S. Consensus clustering in complex networks. Scientific reports. 2012; 2:336. arXiv:1203.6093. [PubMed: 22468223]

Laumann TO, Gordon EM, Adeyemo B, Snyder AZ, Joo SJ, Chen MY, Gilmore AW, McDermott KB, Nelson SM, Dosenbach NUF, Schlaggar BL, Mumford JA, Poldrack RA, Petersen SE. Functional system and areal organization of a highly sampled individual human brain. Neuron. 2015; 87:658–671.

Laumann TO, Snyder AZ, Mitra A, Gordon M, Gratton C, Adeyemo B, Gilmore AW, Nelson SM, Berg JJ, Greene DJ, Mccarthy JE, Tagliazucchi E, Laufs H, Schlaggar BL, Dosenbach NUF, Petersen SE. On the Stability of BOLD fMRI Correlations. Cerebral Cortex. 2016:1–14.

Macey PM, Macey KE, Kumar R, Harper RM. A method for removal of global effects from fMRI time series. NeuroImage. 2004; 22:360–366. [PubMed: 15110027]

Merikangas KR, He JP, Brody D, Fisher P, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 20012004 NHANES. Pediatrics. 2010; 125:75–81. [PubMed: 20008426]

Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? NeuroImage. 2009; 44:893–905. arXiv:NIHMS150003. [PubMed: 18976716]

Murphy K, Fox MD. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. NeuroImage. in press.

Muschelli J, Nebel MB, Caffo BS, Barber AD, Pekar JJ, Mostofsky SH. Reduction of motion-related artifacts in resting state fMRI using aCompCor. NeuroImage. 2014; 96:22–35. [PubMed: 24657780]

Patel AX, Kundu P, Rubinov M, Jones PS, Vértes PE, Ersche KD, Suckling J, Bullmore ET. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. NeuroImage. 2014; 95:287–304. [PubMed: 24657353]

Patriat R, Reynolds RC, Birn RM. NeuroImage An improved model of motion-related signal changes in fMRI. NeuroImage. 2017; 144:74–82. [PubMed: 27570108]

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage. 2012; 59:2142–2154. arXiv:NIHMS150003. [PubMed: 22019881]

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE. Functional network organization of the human brain. Neuron. 2011; 72:665–678. [PubMed: 22099467]

Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage. 2014a; 84:320–341. arXiv:NIHMS150003. [PubMed: 23994314]

Power JD, Plitt M, Laumann TO, Martin A. Sources and implications of whole-brain fMRI signals in humans. NeuroImage. 2016

Power JD, Schlaggar BL, Petersen SE. Studying brain organization via spontaneous fMRI signal. Neuron. 2014b; 84:681–696. arXiv:NIHMS150003. [PubMed: 25459408]

Power JD, Schlaggar BL, Petersen SE. Recent progress and outstanding issues in motion correction in resting state fMRI. NeuroImage. 2015; 105:536–551. arXiv:NIHMS150003. [PubMed: 25462692]

Pruim RHR, Mennes M, Buitelaar JK, Beckmann CF. Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. NeuroImage. 2015a; 112:278–287. [PubMed: 25770990]

Pruim RHR, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage. 2015b; 112:267–277. [PubMed: 25770991]

Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. NeuroImage. 2010; 52:1059– 1069. [PubMed: 19819337]

Saad ZS, Gotts SJ, Murphy K, Chen G, Jo HJ, Martin A, Cox RW. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain connectivity. 2012; 2:25–32. [PubMed: 22432927]

Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. NeuroImage. 2014; 90:449–468. arXiv:NIHMS150003. [PubMed: 24389422]

Satterthwaite TD, Connolly JJ, Ruparel K, Calkins ME, Jackson C, Elliott MA, Roalf DR, Hopsona R, Prabhakaran K, Behr M, Qiu H, Mentch FD, Chiavacci R, Sleiman PMA, Gur RC, Hakonarson H, Gur RE. The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. NeuroImage. 2016; 124:1115–1119. [PubMed: 25840117]

Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, Hakonarson H, Gur RC, Gur RE, Wolf DH. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. NeuroImage. 2013; 64:240–256. arXiv:NIHMS150003. [PubMed: 22926292]

Satterthwaite TD, Elliott MA, Ruparel K, Loughead J, Prabhakaran K, Calkins ME, Hopson R, Jackson C, Keefe J, Riley M, Mentch FD, Sleiman P, Verma R, Davatzikos C, Hakonarson H, Gur RC, Gur RE. Neuroimaging of the Philadelphia Neurodevelopmental Cohort. NeuroImage. 2014; 86:544–553. [PubMed: 23921101]

Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H, Gur RC, Gur RE. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. NeuroImage. 2012; 60:623–632. arXiv:NIHMS150003. [PubMed: 22233733]

Scheinost D, Papademetris X, Constable RT. NeuroImage The impact of image smoothness on intrinsic functional connectivity and head motion confounds. NeuroImage. 2014; 95:13–21. [PubMed: 24657356]

Smith SM, Vidaurre D, Beckmann CF, Glasser MF, Jenkinson M, Miller KL, Nichols TE, Robinson EC, Salimi-Khorshidi G, Woolrich MW, Barch DM, Ugurbil K, Van Essen DC. Functional connectomics from resting-state fMRI. Trends in Cognitive Sciences. 2013; 17:666–682. [PubMed: 24238796]

Spisák T, Jakab A, Kis SA, Opposits G, Aranyi C, Berenyi E, Emri M. Voxel-Wise Motion Artifacts in Population-Level Whole- Brain Connectivity Analysis of Resting-State fMRI. PLoS ONE. 2014:9.

Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: Improved N3 bias correction. IEEE Transactions on Medical Imaging. 2010; 29:1310–1320. [PubMed: 20378467]

Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, Kandel BM, van Strien N, Stone JR, Gee JC, Avants BB. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. NeuroImage. 2014; 99:166–179. [PubMed: 24879923]

Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. Journal of neurophysiology. 2010; 103:297–321. [PubMed: 19889849]

Van Dijk KRA, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity MRI. NeuroImage. 2012; 59:431–438. arXiv:NIHMS150003. [PubMed: 21810475]

Wang H, Cao Y, Syedamahmood T. Multi-atlas segmentation with learning-based label fusion. Machine Learning in Medical Imaging. 2014; 35:256–263.

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag; New York: 2009. http://ggplot2.org

Xia M, Wang J, He Y. BrainNet Viewer: A network visualization tool for human brain connectomics. PLoS ONE. 2013:8.

Yan CG, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, Li Q, Zuo XN, Castellanos FX, Milham MP. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. NeuroImage. 2013a; 76:183–201. arXiv:NIHMS150003. [PubMed: 23499792]

Yan CG, Craddock RC, He Y, Milham MP. Addressing head motion dependencies for small-world topologies in functional connectomics. Frontiers in human neuroscience. 2013b; 7:910. [PubMed: 24421764]

Yan CG, Craddock RC, Zuo XN, Zang YF, Milham MP. Standardizing the intrinsic brain: Towards robust measurement of inter-individual variation in 1000 functional connectomes. NeuroImage. 2013c; 80:246–262. arXiv:NIHMS150003. [PubMed: 23631983]

Yang GJ, Murray JD, Repovs G, Cole MW, Savic A, Glasser MF, Pittenger C, Krystal JH, Wang XJ, Pearlson GD, Glahn DC, Anticevic A. Altered global brain signal in schizophrenia. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:7438–43. [PubMed: 24799682]

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zollei L, Polimeni JR, Fischl B, Liu H, Buckner RL. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology. 2011; 106:1125–1165. [PubMed: 21653723]

Zeng LL, Wang D, Fox MD, Sabuncu M, Hu D, Ge M, Buckner RL, Liu H. Neurobiological basis of head motion in brain imaging. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:6058–62. [PubMed: 24711399]

Zuo XN, Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J, Breitner JCS, Buckner RL, Calhoun VD, Castellanos FX, Chen A, Chen B, Chen J, Chen X, Colcombe SJ, Courtney W, Craddock RC, Di Martino A, Dong HM, Fu X, Gong Q, Gorgolewski KJ, Han Y, He Y, He Y, Ho E, Holmes A, Hou XH, Huckins J, Jiang T, Jiang Y, Kelley W, Kelly C, King M, LaConte SM, Lainhart JE, Lei X, Li HJ, Li K, Li K, Lin Q, Liu D, Liu J, Liu X, Liu Y, Lu G, Lu J, Luna B, Luo J, Lurie D, Mao Y, Margulies DS, Mayer AR, Meindl T, Meyerand ME, Nan W, Nielsen JA, O'Connor D, Paulsen D, Prabhakaran V, Qi Z, Qiu J, Shao C, Shehzad Z, Tang W, Villringer A, Wang H, Wang K, Wei D, Wei GX, Weng XC, Wu X, Xu T, Yang N, Yang Z, Zang YF, Zhang L, Zhang Q, Zhang Z, Zhang Z, Zhao K, Zhen Z, Zhou Y, Zhu XT, Milham MP. An open science resource for establishing reliability and reproducibility in functional connectomics. Scientific data. 2014; 1:140049. [PubMed: 25977800]

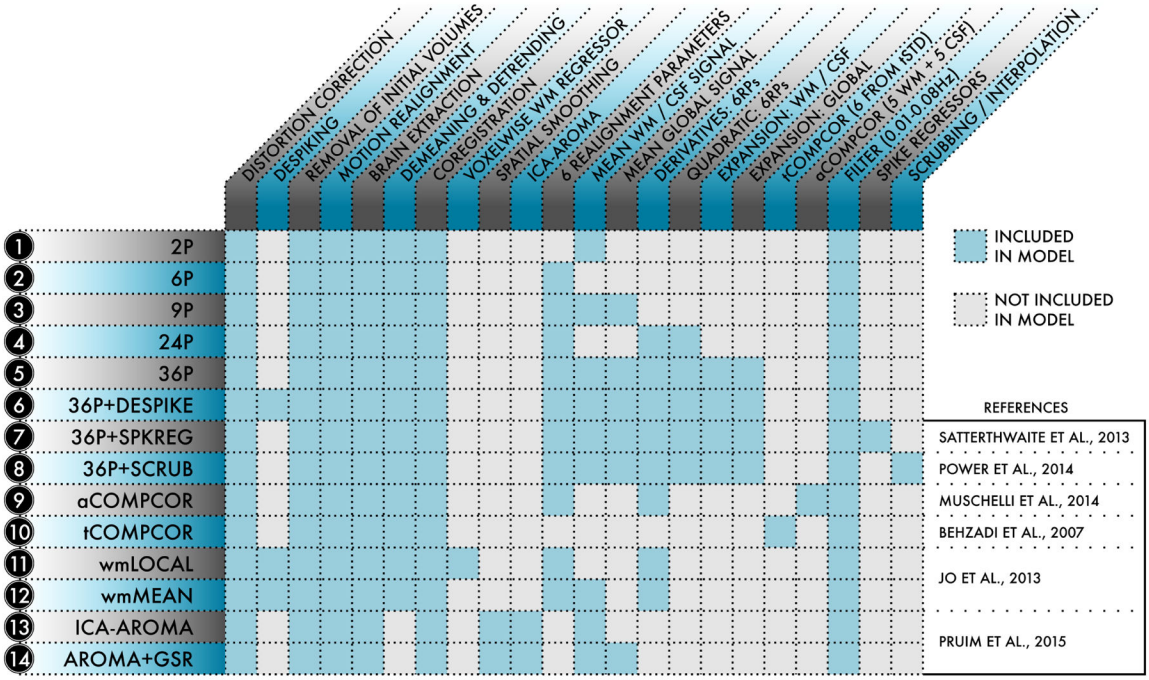

Figure 1. Schematic of the 14 de-noising models evaluated in the present study

For each of the 14 models indexed at left, the table details what processing procedures and confound regressors were included in the model. De-noising models were selected from the functional connectivity literature and represented a range of strategies.

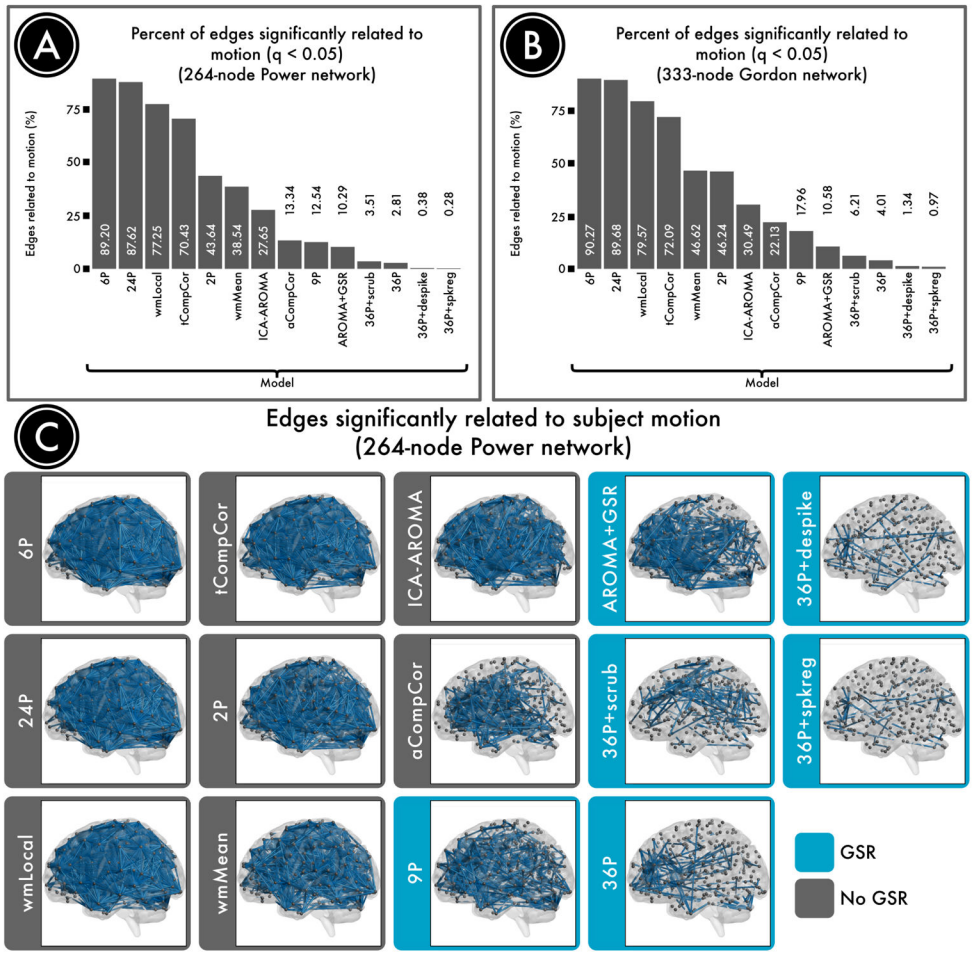

Figure 2. Number of edges significantly related to motion after de-noising

Successful de-noising strategies reduced the relationship between connectivity and motion. The number of edges (network connections) for which this relationship persists provides evidence of a pipeline's efficacy. ***A,*** The percentage of edges significantly related to motion in a 264-node network defined by Power et al. (2011). Fewer significant edges is indicative of better performance. ***B,*** The percentage of edges significantly related to motion in a second, 333-node network defined by Gordon et al. (2016). ***C,*** Renderings of significant edges with QC-FC correlations of at least 0.2 for each de-noising strategy, ranked according to efficacy. Strategies that include regression of the mean global signal are framed in blue and consistently ranked as the best performers.

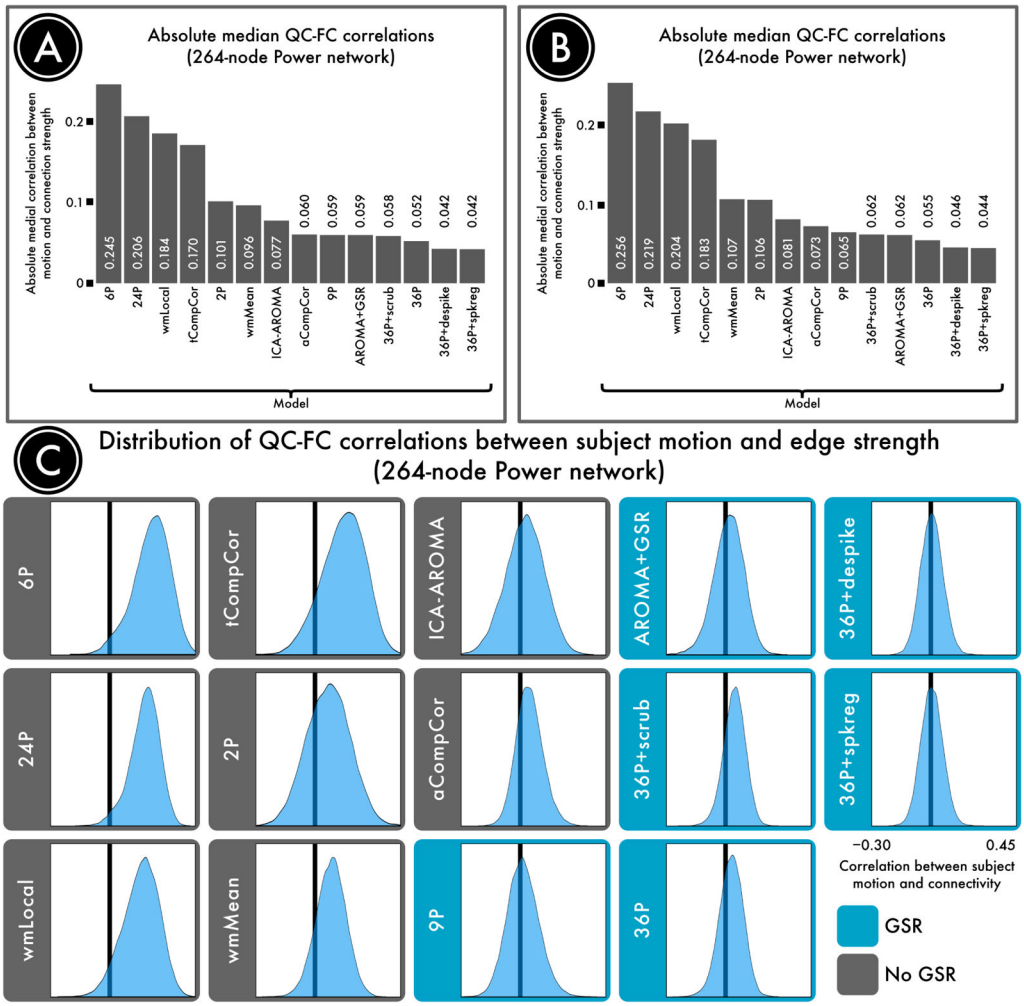

Figure 3. Residual QC-FC correlations after de-noising

The absolute median *QC-FC* correlation is another measure of the relationship between connectivity and motion. ***A,*** The absolute median correlation between functional connectivity and motion in a 264-node network defined by Power et al. (2011). A lower absolute median correlation indicates better performance. ***B,*** The absolute median correlation between functional connectivity and motion in a second, 333-node network defined by Gordon et al. (2016). ***C,*** Distributions of all edgewise *QC-FC* correlations after each de-noising strategy, ranked according to efficacy. Results largely recapitulated those reported in Figure 2, with GSR-based approaches (blue frame) collectively exhibiting the best performance. Whereas approaches that included more regressors generally yielded a narrower distribution, those approaches that included GSR tended to shift the distribution's center toward 0.

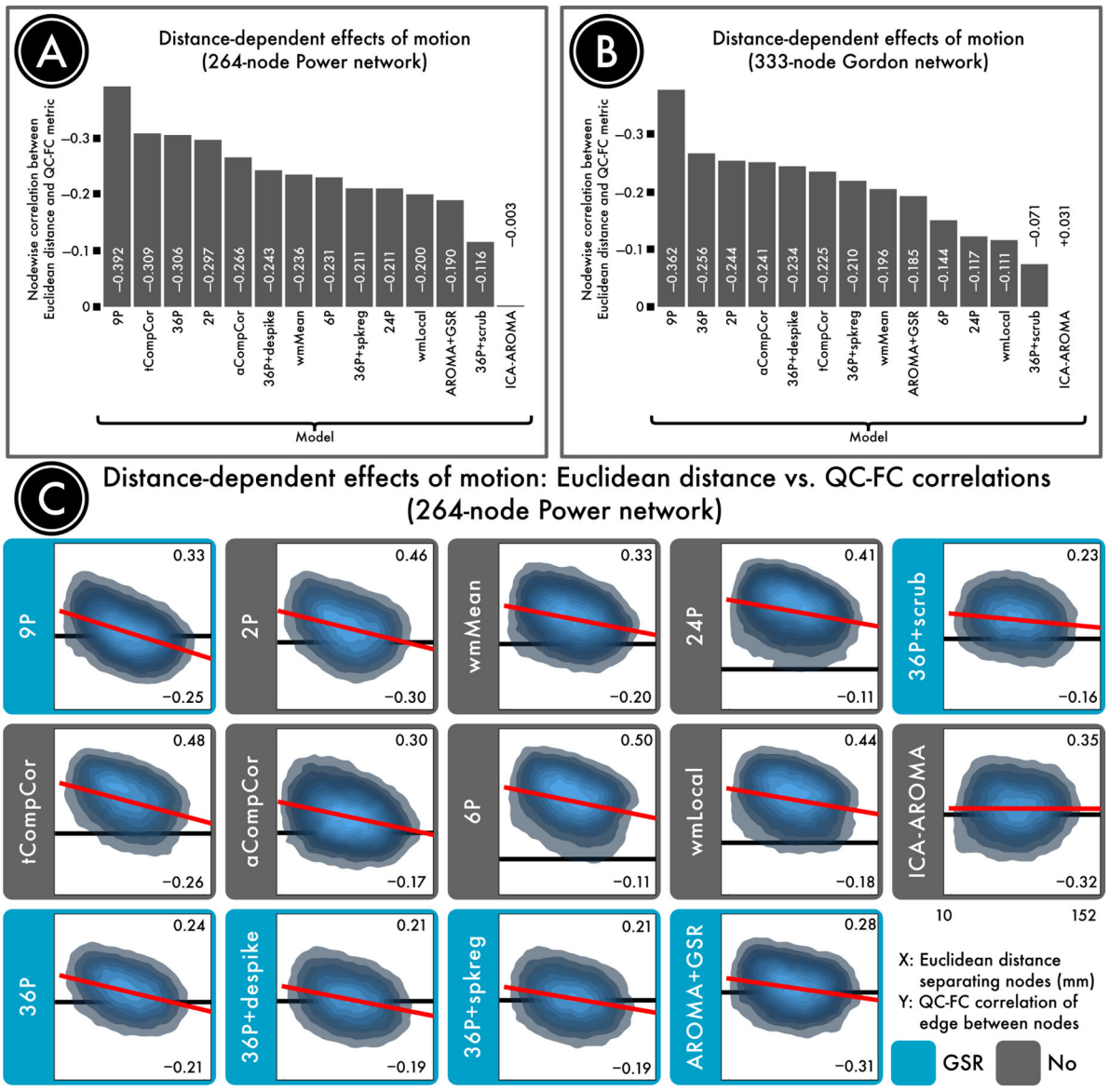

Figure 4. Distance-dependence of motion artifact after de-noising

The magnitude of motion artifact varies with the Euclidean distance separating a pair of nodes, with closer nodes generally exhibiting greater impact of motion on connectivity. ***A,*** The residual distance-dependence of motion artifact in a 264-node network defined by Power et al. (2011) following confound regression. ***B,*** The residual distance-dependence of motion artifact in a second, 333-node network defined by Gordon et al. (2016). ***C,*** Density plots indicating the relationship between the Euclidean distance separating each pair of nodes (x-axis) and the *QC-FC* correlation of the edge connecting those nodes (y-axis). The overall trend lines for each de-noising strategy, from which distance-dependence is computed, are indicated in red. For each plot, the ordinate is rescaled to the data; thus, the ordinate does not reflect the width of the distribution of QC-FC correlations. (The same data is plotted to a common ordinate in Supplemental Figure 1.) The best performing models either excised high-motion volumes (36-parameter + scrubbing) or used more localized

regressors (ICA-AROMA and wmLocal). In general, approaches that made use of GSR without censoring resulted in substantial distance-dependence. This effect was driven by differential efficacy of de-noising, with effective de-noising for long range connections but not short-range connections.

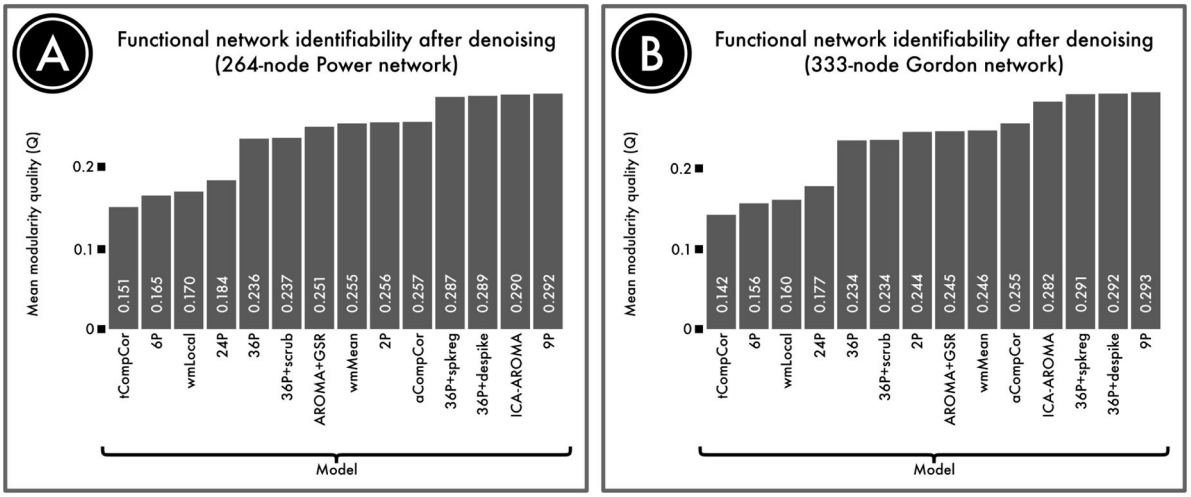

Figure 5. Identifiability of network structure after de-noising

Although de-noising approaches remove motion artifact from BOLD time series, it is possible that they also remove signal of interest. We quantified the retention of signal of interest as the modularity quality of the de-noised connectome. ***A,*** The modularity quality in a 264-node network defined by Power et al. (2011) following confound regression. ***B,*** The modularity quality in a second, 333-node network defined by Gordon et al. (2016). ICA-, GSR-, and tissue class-based models performed relatively well, while models that included realignment parameters alone did not remove enough noise to accurately identify network structure.

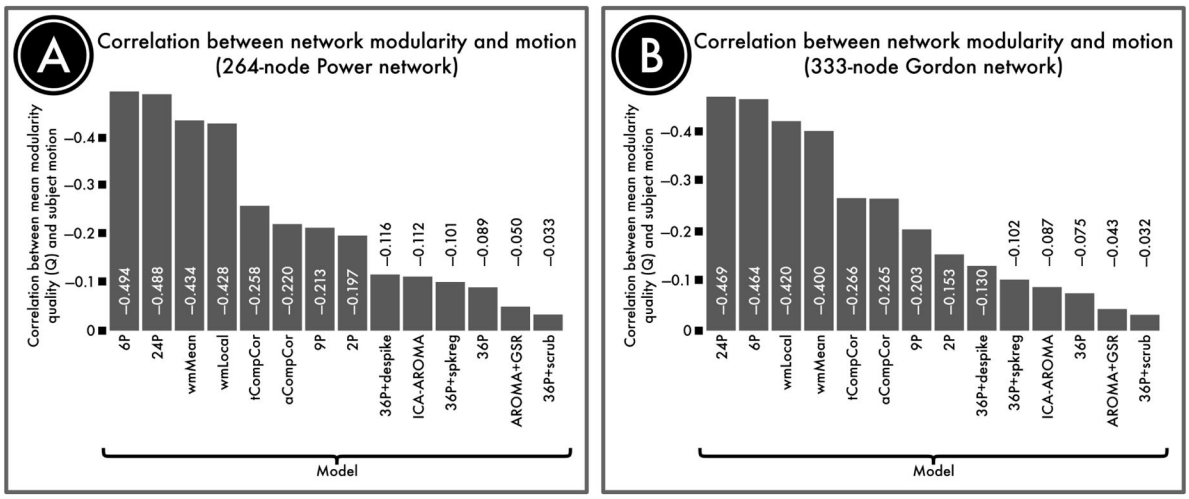

Figure 6. Correlation between subject motion and modularity quality

Motion affects network modularity to varying degrees for different de-noising approaches. We quantified the retention of signal of interest as the modularity quality of the de-noised connectome. ***A,*** The correlation between subject motion and modularity quality in a 264-node network defined by Power et al. (2011) following confound regression. ***B,*** The correlation between subject motion and modularity quality in a second, 333-node network defined by Gordon et al. (2016). In general, GSR- and ICA-based methods most effectively decoupled network structure from artifact.

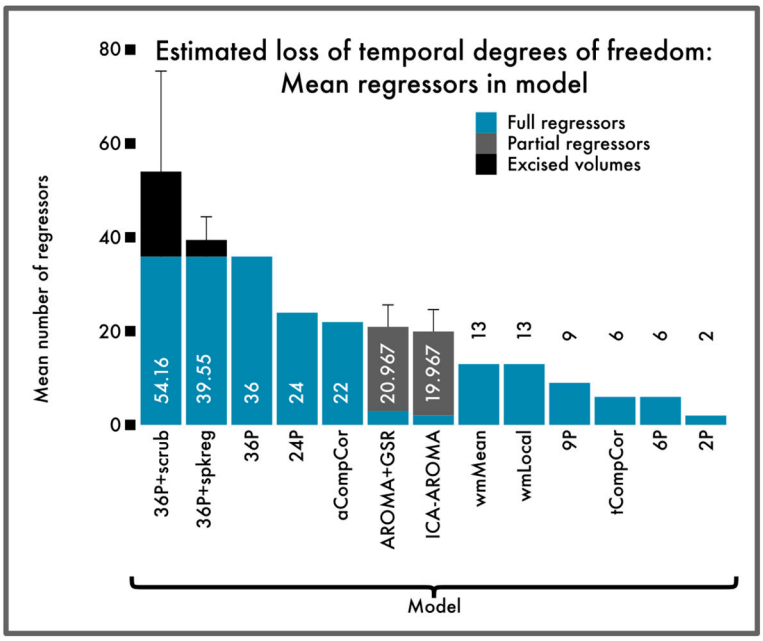

Figure 7. Estimated loss of temporal degrees of freedom for each pipeline evaluated

Bars indicate mean number of additional regressors per confound model; error bars indicate standard deviation for models where the number of confound regressors varies by subject. High-parameter models and framewise censoring performed well overall on other benchmarks, but were also costliest in terms of temporal degrees of freedom. Despite this cost, augmenting a high-parameter model with censoring improved signal detection (see Figure 5), suggesting that the lost degrees of freedom corresponded largely to noise. Because the 36P+despike model censors data in a spatially adaptive manner, the DOF loss in this case varied by voxel, and is not displayed.