Mitigating head motion artefact in functional connectivity MRI

Rastko Ciric[1], **Adon FG Rosen**[1], **Guray Erus**[2], **Philip A Cook**[3], **Danielle S Bassett**[4], **Christos Davatzikos**[2,3], **Daniel H Wolf**[1,2], **Theodore D Satterthwaite**[1,2,*]

[1] Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

[2] Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, USA

[3] Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

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

Abstract

Participant motion during functional magnetic resonance image (fMRI) acquisition produces spurious signal fluctuations that can confound measures of functional connectivity. Without mitigation, motion artefact can bias statistical inferences about relationships between connectivity and individual differences. To counteract motion artefact, this protocol describes the implementation of a high-performance denoising strategy that combines a set of model features, including physiological signals, motion estimates, and mathematical expansions, to target both widespread and focal effects of subject movement. This method can reduce motion-related variance to near zero in studies of functional connectivity, providing up to a hundredfold improvement over minimal processing approaches in large data sets. Image denoising requires 40 minutes to 4 hours of computing per image, depending on model specifications and data dimensionality. The protocol additionally includes instructions for assessing the performance of a denoising strategy. Associated software implements all denoising and diagnostic procedures using a combination of established image processing libraries (FSL, AFNI, and ANTs) and new pipeline software (the XCP system, downloadable from Github: http://github.com/PennBBL/xcpEngine).

Keywords

fMRI; artefact; software; noise; functional connectivity; motion

---

INTRODUCTION

Functional connectivity analysis has yielded important insights into the intrinsic organisation of the human brain[1,2]. Whereas task fMRI examines the functional responses to specific cognitive, perceptual, or motor manipulations, functional connectivity MRI can provide detailed maps of major functional networks without requiring use of a task[3–5]. Prior studies have shown that important individual differences, including age across the lifespan, cognitive performance, and psychiatric diagnoses are associated with differences in functional connectivity[6–9]. However, many measures of interest in neuroimaging studies are also correlated with the movement of subjects in the scanner: for instance, children and subjects from clinical populations tend to move more when they are scanned than healthy adults[10]. In-scanner motion induces large signal fluctuations in fMRI time series data, which can systematically alter observed patterns of functional connectivity[10–13].

If data processing fails to account for motion-related signal variance, artefact can easily confound inference. For example, initial reports that brain development is associated with strengthening of long-range connections and weakening of short-range connections have been shown to be dramatically inflated by the presence of motion artefact in younger children[8,14]. Accordingly, investigators have developed numerous *denoising* strategies to mitigate the influence of motion on functional connectivity, and thus disentangle motion effects from effects of interest (Box 1). However, not all of these strategies are equally efficacious. Furthermore, the rapid proliferation of denoising approaches has introduced uncertainty among researchers as to which approach has greatest utility for their data.

While the development of denoising strategies will continue into the foreseeable future, an emerging corpus of evidence suggests that the most successful strategies share a number of common elements that target particular features of motion artefact (Box 2). Here, we describe a validated, high-performance protocol for removal of motion artefact from functional MRI data. We also provide a theoretical and historical context for our protocol, its development, and its utility. Finally, we introduce a software package that implements common denoising protocols and that provides tools for assessing the efficacy of denoising.

Development of the protocol

Although the susceptibility of fMRI to motion has been acknowledged since 1996, it was not until 2012 that the influence of motion artefact on functional connectivity was systematically investigated[15]. Three studies published by independent groups not only reported the systematic contamination of functional connectivity measures by motion artefact, but also introduced measures to control for this artefact[10–12]. By and large, most approaches to denoising functional connectivity data have expanded on a family of strategies that had previously been used to denoise functional activation data: *confound regression*[15].

Confound regression aims to separate signal from noise by modelling artefactual processes as time series. These time series might include, for instance, frame-to-frame estimates of how much the head has moved, or signals from noise-prone tissue compartments such as the ventricles. These "artefact" time series are fit to the observed BOLD time series using a general linear model (the confound model). The residuals of the fit, corresponding to the

variance that the confound model cannot explain, can then be analysed as "cleaned" data, while the BOLD variance explained by the confound model can be discarded from further analyses.

Confound regression remains the most prevalent method for removing head motion artefact in functional connectivity MRI. However, the performance of any confound regression strategy will depend on the constituent time series of its confound model. A more *effective* model will explain more artefactual variance in the data (and thus remove more artefact), while a more *efficient* model will minimise false positives (and thus not mistakenly identify signal of interest as artefactual in origin). The constitution of the optimal confound model has been hotly debated in the field, with numerous studies proposing a new, ostensibly better model every year[16–21] (summarised in Box 1).

More recently, several studies have attempted to cope with the proliferation of confound models by establishing a standard set of benchmarks to evaluate their efficacy and efficiency (Tables 1 and 2). These benchmarking studies have yielded remarkably convergent results[22–25]. Taken together, these studies indicate that models that rely exclusively on estimates of frame-to-frame head movement fail to correct for motion artefact. In contrast, models that remove the time course of the average BOLD signal across the entire brain[26,27] ("global signal regression", or GSR) and/or time series obtained via signal decomposition techniques[16,19–21] (principal component analysis [PCA] or independent component analysis [ICA]) exhibit markedly improved performance. Such models can be profitably augmented with *temporal censoring* operations (e.g., spike regression, scrubbing)[11,17,18] that specifically remove the influence of volumes corrupted by artefact. Most top-performing strategies combine several techniques in order to mitigate the influence of both local and global features of motion artefact[28] (Box 2). Here, we provide a protocol for implementation of a high-performance denoising strategy, while acknowledging that different strategies have specific strengths and weaknesses. Ultimately, the specific choice of confound regression strategy should be dictated by the hypothesis of interest.

Comparison with other methods

Despite its ubiquity, confound regression is not the only means of correcting for head motion artefact in functional connectivity MRI. Two additional families of motion correction approaches that have gained traction are prospective and group-level motion correction strategies. Prospective approaches operate during data acquisition, for instance by using behavioural interventions to pre-emptively minimise head movement[29], by leveraging MR sequences designed to facilitate separation of signal from noise[30–32], or by ensuring that some minimal number of low-noise snapshots of the brain are available for each participant[33]. While minimization of motion during acquisition is critical for all studies, novel techniques for aiding prospective motion correction cannot be applied to the vast majority of datasets that have already been acquired. Furthermore, even in data sets where prospective approaches have been applied, it is likely that data quality can be further improved via retrospective denoising approaches. For instance, multi-echo acquisition techniques support improved identification and removal of focal motion artefact, but denoising can be optimised by combining the advantages offered by this prospective

approach with global artefact removal using retrospective denoising strategies such as GSR[34].

In contrast, group-level retrospective approaches are applied after image processing, during group-level analysis. They often take the form of a nuisance covariate that is included in explanatory models alongside variables of interest. Importantly, correction at the group level is not equivalent to participant-level correction[35], nor is it mutually exclusive. On the one hand, it can reduce the sensitivity of analyses to any connections that are both related to the variable of interest and susceptible to motion artefact[36]. For example, including both age and motion in a regression model as part of a study of brain development will by definition remove any shared variance. On the other hand, inclusion of a single motion covariate in a group-level linear model is insufficient to address either nonlinear effects or interactive effects of motion with variables of interest such as age. In summary, while group-level approaches can often supplement participant-level confound regression, they may be inadequate as a stand-alone approach. In general, best practices minimize motion during acquisition, use a high performance confound regression model during participant-level de-noising, and evaluate the impact of motion in group-level statistical models.

Applications of the protocol

Functional connectivity analysis is a routine component of many neuroimaging studies, and a growing number of large studies are organised around the explicit aim of mapping the functional connectome in humans[2,37]. Any MRI study of the brain's functional connectivity can reasonably be expected to include some form of motion denoising as part of image processing. However, denoising is most critical in cases where the scientific hypothesis is focused around individual differences, especially age, clinical diagnosis, symptom burden, or cognition. The strengths of the protocol we present below have been demonstrated most recently in a pair of head-to-head evaluations of motion correction strategies according to standardised benchmarks[22,23].

Limitations

The implementation of confound regression strategies has not been without controversy. One of the most heated controversies in functional neuroimaging has concerned the appropriateness of using the global mean signal in confound regression[27,38–43]. A primary concern raised regarding GSR is the possibility that the mean global signal might have neural origin, or that it could reflect processes of interest[44,45]. However, accumulating evidence suggests that the global signal is in fact dominated by signals of non-neural origin: the content of motion-related and respiratory noise in the global signal is demonstrably high across datasets and scanning platforms[13]. Furthermore, a major limitation of alternative approaches based on spatially specific nuisance models (including ICA), is their inability to remove widespread, globally distributed artefact that is the primary result of motion ("Type 2" artefact, in Box 2)[24,46].

Although GSR is a highly effective denoising strategy, data processed using GSR necessarily include negatively correlated time series[27,39]. Whereas the distribution of connection strengths in data processed without GSR typically has a positive centre and

negative skewness, GSR-based processing often results in a distribution that is zero-centred and relatively symmetrical[39]. Without artefact-free data available to provide a ground truth, it is unknown whether negative correlations in the connectome reflect a biologically meaningful process or an artefact of processing. However, even when data are processed without GSR, lower-motion subjects tend to exhibit a greater number of negative edges and a distribution of connection strengths that is closer to zero-centred in comparison with higher-motion subjects[18]. Finally, it should be noted that GSR tends to increase the distance-dependence of motion artefact, especially when it is not supplemented with either temporal censoring or spatially specific nuisance regression[17] (see Box 2). However, it should be emphasized that this relationship is due to differential de-noising efficacy for long-range over short-range connections. Many alternative models that do not include GSR have less distance-dependence, but this insensitivity is often due to ineffective performance at connections of all lengths.

In addition to GSR, another highly effective technique is temporal censoring. However, denoising strategies that include temporal censoring can be costly in terms of lost temporal degrees of freedom (tDOF); they are effective but not always efficient. As a result, the number of frames censored – and thus the cost in tDOF – will vary across subjects, leading to the possibility that group differences are introduced because of the resultant sampling differences. Furthermore, temporal censoring clearly disrupts the temporal dynamics of the timeseries. However, objections to temporal censoring should be tempered by the observation that the signal being removed in censoring may be of questionable value in the first place. Thus, motion-corrupted data do not have the same as tDOF or temporal dynamics as artefact-free time series. Moreover, investigators can randomly censor non-corrupted data in order to balance the tDOF between groups, albeit at the cost of losing true observations.

Another open problem in motion mitigation is the quality and nature of frame-to-frame motion estimates. Typically, subject movement is estimated post hoc, by aligning each frame to a reference and using the alignment matrix to estimate that frame's translational and rotational displacement from the reference. Real-time, optical measures provide an alternative option that may better estimate true head motion. Although such measures are under active development, they have yet to be adopted broadly, in part because the requisite technology is at present unavailable at most MRI centres. Furthermore, although the motion estimates obtained via post hoc realignment may be imperfect, prior work indicates convergence between these estimates and more direct optical measures[47].

Overview of the procedure

In this protocol, we provide a step-by-step guide to implementing a high-performance de-noising pipeline (Figure 1). In addition to the confound regression protocol itself, an equally important dimension of motion artefact correction is transparent reporting of data quality indicators (Tables 1 and 2). Such indicators provide a measure of the artefact initially present in each image and across the entire sample, and also of the residual artefact after images have been de-noised. These measures can be leveraged as proxies for denoising performance. Thus, we additionally describe how to quantify and report (A) the presence of artefact in a functional brain image and (B) the efficacy and efficiency of the motion

correction protocol for an individual subject and for the sample as a whole. Code for all stages of the protocol is freely and publicly available online (https://github.com/PennBBL/ xcpEngine).

Each step of the protocol both describes the objective of that step and includes a specific implementation in code. A variety of software libraries are available for image processing, and in many cases, alternatives are available to the specific implementation that we provide. A description of the general objective of each processing step is included so that investigators can select the best tool tailored to their own specific needs.

The denoising protocol can be implemented step-by-step by following the procedure outlined here, but we also provide a software package that flexibly implements the protocol in an automated manner. The eXtensible Connectivity Pipeline (or "XCP") provides a platform for the processing of neuroimaging data that packages common processing steps into configurable modules, and then combines modules into processing streams, with each stream corresponding to an imaging modality or an analytic objective. Each module uses common tools, enforces consistent output directory structures, and utilises identical naming conventions. Metadata and provenance of any image derivatives are internally tracked, allowing the pipeline to process each derivative in the most appropriate way.

The XCP system is powered by AFNI, ANTs, and FSL[48–50], but also introduces multiple standalone image processing utilities. XCP supports processing streams for a number of sequences and analytic modalities, including structural morphometry, task based fMRI, and arterial-spin labelled MRI. However, in this protocol we focus on applications to functional connectivity MRI. At present, XCP includes default configurations for three high-performance confound regression strategies: an anatomical CompCor model based on principal component analysis[19], an ICA-AROMA model based on independent component analysis[20], and a high-parameter model based on noisy tissue compartments and motion estimates[17]. All streams include GSR[26], and all can optionally incorporate censoring[11,17,18,51]. (A version of each stream that does not incorporate GSR is also available to facilitate assessing the robustness of results to alternative processing schemes, but we do not recommend these alternatives for most use cases.)

When processing data with XCP Engine, a *design file* and a *cohort file* are required. The design file parametrises the processing stream, while the cohort file characterises the input sample. A configuration script and GitHub repository are available for easy configuration of the design file (https://github.com/PennBBL/xcpConfig). A brief overview of cohort file specifications is included in Box 3, as is an example. More extensive documentation on the design and cohort files, including examples, format specifications, and instructions for cohort file setup are available online (https://pipedocs.github.io/config/index.html). XCP can be run either from a local installation or using the publicly available Image Processing Portal (IPP, https://ipp.cbica.upenn.edu) from the Center for Biomedical Image Computing and Analytics (CBICA) at the University of Pennsylvania, which provides the scientific community with the resources of a high-performance computing cluster for processing data with XCP, and also provides automatic access to future extensions of XCP functionality.

Functional connectivity processing generally requires that each participant's anatomical data be processed first, and in the protocol detailed here we assume that the user has already executed an anatomical pipeline to completion. Outputs of anatomical processing should minimally include (1) a high-resolution anatomical image with non-brain tissue stripped and with any intensity bias artefact removed; (2) a segmentation of the anatomical image into grey matter, white matter, and cerebrospinal fluid; and (3) a set of transforms that map between coordinates in the subject's anatomical image and coordinates in a standard reference space that is the same for all subjects, such as the MNI template. The segmentation of the structural image is required to isolate nuisance signals from white matter and cerebrospinal fluid, and it can additionally be used to improve co-registration quality. Transforms to template space (also called "image normalisation" or "registration") are a prerequisite for generating functionally homologous connectomes across subjects; they are used to move a standard brain parcellation into each subject's coordinate space so that regional signals can be computed and correlations between them can be estimated. XCP requires that normalisations to template space use warps generated by the top-performing registration procedures included in Advanced Normalization Tools (ANTs)[49,52]. ANTs transforms are saved as an affine matrix and a NIfTI-formatted vector field. These transforms and all other anatomical prerequisites can be easily obtained using the ANTs Cortical Thickness pipeline (ANTsCT; https://github.com/ANTsX/ANTs)[53]. As an alternative to the ANTsCT pipeline, XCP supports several different levels of anatomical processing to obtain the requisite anatomical derivatives; these options are detailed in the online documentation.

Experimental design

The design of a confound model is subject to a number of considerations. Principal among these are the research hypothesis under consideration and the nature of the data set being processed[54]. Because motion can confound studies of individual or group differences (especially in clinical and developmental samples), a confound model's efficacy in reducing the motion-connectivity relationship is a primary concern in these studies. For basic research in connectomics, the integrity of network topology might be of equal importance; in this case, the distance-dependence of residual artefact could be a critical indicator.

Although top-performing models can be expected to generalise reasonably well across data sets, there will be some variability in their performance. For example, it should be noted that these processing streams were developed and benchmarked on fMRI data with standard temporal resolution (i.e., TRs of 2–3s), and their performance on multiband fMRI data with high temporal resolution has not been fully evaluated. As such, if performance benchmarking is a feasible option, then evaluating several motion correction strategies prior to processing the data set can inform the decision as to which model to use. Otherwise, the set of existing benchmarking studies suggest first that top-performing models include GSR. Second, framewise censoring approaches provide a highly effective complement to GSR that can mitigate the distance-dependent profile of motion artefact but that can also incur considerable costs in the data's degrees of freedom and autocorrelation structure. Third, although they might not be as effective as censoring, signal decomposition strategies such as ICA-AROMA and aCompCor provide an alternative mechanism for targeting spatially focal

artefact that can be missed when applying GSR alone[22–25]. Importantly, denoising elements are often combined to produce an effective confound model (Boxes 1 and 2).

To evaluate whether denoising is successful, investigators can examine measures computed at both subject and sample levels (Tables 1 and 2). At the subject level, we suggest assessing the outcome of the denoising procedure by visualising the entire data set with a voxelwise carpet plot[55] (Figure 2). This plot can qualitatively illuminate frame-to-frame relationships between movement and the BOLD signal and can be used to identify BOLD signal fluctuations that co-occur with subject movement. At the sample level, the residual correlation between motion and functional connectivity across subjects,[10,11,17,28] or *quality control—functional connectivity* (QC-FC) correlation, provides a quantitative measure of denoising success (Figure 3). Strong QC-FC correlations indicate that functional connectivity was substantially impacted by motion. Beyond QC-FC relationships, it is useful to measure QC-FC distance dependence, or the extent to which QC-FC correlations are dependent on the Euclidean distance between the centres of mass of any pair of regions[10,11,17]. QC-FC distance dependence provides an informative diagnostic of the spatial profile of residual artefact (Box 2). In addition to QC-FC and distance dependence, another potentially useful standard for the assessment of denoising methods is test-retest reliability. The Consortium for Reliability and Reproducibility (CoRR[56]) hosts a public repository of data sets with repeated measures that provide an ideal resource for studies assessing test-retest reliability. Investigators using this metric are cautioned that motion artefact itself exhibits test-retest reliability[25,56].

Another consideration is the decision to include or exclude censoring in the confound model. Censoring is a denoising approach whereby all variance in a motion-contaminated frame is completely removed from the time series, either via deletion of the frame or via spike regression[17,18,51] (Box 1). Modern censoring approaches are designed to minimise the influence of censored frames by incorporating the censoring procedure into any intermediate stages of the analysis, such as de-trending and temporal filtering, leaving the autocorrelation structure of the data intact for as long as possible[18,57,58]. Two large-scale benchmarking studies have ranked censoring in combination with GSR as one of the most effective families of retrospective participant-level motion correction strategies[22,23].

However, a frequent argument against the use of censoring is that its application results in a large and variable loss of temporal degrees of freedom in the data (see Limitations)[18,42]. In general, we advocate censoring except in cases where (1) the input time series is short enough that removal of any additional frames could result in insufficient degrees of freedom, or (2) the analytic objective requires the autocorrelation structure of the data to remain intact. For instance, sliding window connectivity analysis depends on the autocorrelation of adjoining frames, and investigators using this technique in combination with censoring may need to account for any disruptions that censoring incurs to the autocorrelation structure. Although our research group has historically favoured a model that combines censoring (either framewise or spatially adaptive despiking) with GSR, tissue-based regressors, motion estimates, and temporal and quadratic expansions, current indications suggest that models that combine GSR with either ICA-AROMA[20] or anatomical CompCor[16,19] are very reasonable alternatives[22,23].

Level of expertise needed to implement the protocol

Designing a motion de-noising pipeline *de novo* is a demanding programing task, requiring competence in time series analysis and image processing, and a considerable familiarity with the range of software suites available for processing of brain images. In contrast, executing the automated processing stream provided via XCP is substantially more accessible. To use XCP, users should have attained a minimal level of competence with command line usage, and should have access to a computing environment where AFNI, ANTs and FSL can be installed. We anticipate that most users of existing fMRI tools will be able to use this software successfully. Additionally, familiarity with statistical modelling and an understanding of functional connectivity is required to design and interpret QC-FC models.

MATERIALS

EQUIPMENT

Software

- A computer or virtual machine running a terminal emulator with Bash shell v4.0 or higher.

- Standard software libraries for processing of neuroimaging data:

    – FSL[50] (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation).

    – AFNI[48] (https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/background_install/install_instructs/index.html).

    – ANTs[49] (http://stnava.github.io/ANTs/).

    – C3D (http://www.itksnap.org/pmwiki/pmwiki.php?n=Downloads.C3D)

- The statistical language R, along with image processing packages available from CRAN (filter, pracma, RNifti, and optparse) (https://cran.r-project.org/). To support graphical renderings of denoising diagnostics, it is recommended that the R packages ggplot2, reshape2, svglite, and grid also be installed. To support modelling of repeated measures from each subject, the R package lme4 must be installed.

- XCP system (https://github.com/PennBBL/xcpEngine).

- (Optional) ICA-AROMA[20] software implementation (https://github.com/rhr-pruim/ICA-AROMA) and Python 2.7, with Python modules os, argparse, commands, numpy, and random.

Input data

- BOLD-weighted functional MR images of each subject's brain. These are the 4-dimensional time series to be de-noised.

    CRITICAL All images should be converted to NIfTI format prior to processing.

CRITICAL Because the results of fMRI processing are impacted by data integrity, both functional and anatomical images should be examined, either visually or using quantitative metrics[59,60], to evaluate suitability for processing.

CRITICAL The quality of image registration will be improved by correcting for magnetic field distortions incurred during functional image acquisition. If the acquired sequences support distortion correction of functional images (e.g. via FSL's FUGUE or TOPUP routines), this procedure should be performed before motion correction.

Anatomical and atlas input

- A high-resolution anatomical MR image for each subject, with skull and non-brain tissue stripped from the image.

- Anatomical segmentations of the subject's brain into grey matter, white matter, and cerebrospinal fluid.

- Transforms that map between coordinates in subject-anatomical space and coordinates in a template space that can be used as a standard across all subjects. This template space could be either the MNI152 template or a sample-specific template. For users of XCP Engine, XCP assumes that the transforms are in ITK (ANTs) format.

- If anatomical segmentations or ITK-formatted transforms between subject and template space are not available, these can be obtained by running the complete ANTs Cortical Thickness pipeline[53] for each subject. Alternatively, XCP implements several options for preparing all necessary anatomical input; refer to the online documentation for details.

- An atlas or parcellation that defines regions of interest across the brain.

  CRITICAL The atlas should ideally be designed as a functional parcellation that is suitable for connectomic analysis. Commonly used parcellations include the 264-region Power parcellation[3], the 333-region Gordon areal parcellation[61], the Yeo network parcellation[4], the multiscale Lausanne parcellations[62], data-driven functional parcellations[63], and the multiscale local-global Schaefer parcellations[64].

  CRITICAL If manually processing the data without XCP, take note of the coordinate space in which the parcellation is defined. Many atlases are available in the MNI152 coordinate space, for instance. The transforms, above, should be able to map, or "warp", the parcellation into subject-anatomical space using an appropriate warping command, such as antsApplyTransforms (with ANTs) or applywarp (with FSL). If they are not, then it will be necessary to either (1) compute new transforms from subject-anatomical space to the parcellation space and use those as the input transforms, or (2) obtain a parcellation in a space that is accessible to the existing transforms.

> **CRITICAL** If processing the data using XCP, the connectome will be computed over the Power parcellation by default. To run additional parcellations, download them from the brainspaces database on GitHub (https://github.com/brainspaces) and then run, from the XCP install directory:

```
utils/atlasMetadata -a <parcellation name> -d <downloaded
repository>
```

Equipment setup—*ANTs, AFNI, FSL, R.* Instructions for installation and setup of all software packages are available at the websites of those packages. R can typically be installed using the operating system's package manager. Required R packages can be installed from within R by calling the install.packages() command with the package name as argument. To verify successful installation of all dependencies, run the XCP command-line utility core/checkDependenciesXCP.

XCP utilities.: To run XCP utilities, define the environmental variable XCPEDIR to point to the XCP install directory:

```
export XCPEDIR=<full path to install directory>
```

Including this definition statement in the user's .bash_profile or .bashrc is recommended.

Running XCP Engine.: XCP automates all motion correction and quality assessment steps described below through its front end, xcpEngine. This automation is powered by a user-provided cohort file (subject list) and design file (which toggles the optional steps specified below). The design file can be configured by running the script xcpConfig in the XCP install directory, after defining XCPEDIR as above, and then selecting Functional connectivity. Consult the software manual for detailed documentation (https://pipedocs.github.io).

Running XCP Engine on a computing cluster, without installing dependencies.—In lieu of installing all dependencies, it is possible to complete the motion correction protocol using the Image Processing Portal (IPP, https://ipp.cbica.upenn.edu/) at the Center for Biomedical Image Computing and Analytics at UPenn, which implements pre-processing in XCP. For input to IPP, prepare a design file as described above. Upload functional input data in compressed NIfTI format (.nii.gz), and upload any processed anatomical data, first archiving and compressing the directory containing all anatomical inputs (e.g., the ANTs Cortical Thickness output directory) in .tar.gz format. If anatomical data were processed in XCP, each subject's struc subdirectory should be archived and compressed as .tar.gz.

Alternatives.—Alternative software packages are available for automation of many protocol steps. A selection of alternatives is listed here. Consult the relevant manuals for detailed documentation and installation instructions.

- The Python-based software package C-PAC[65] implements most subject-level motion correction strategies.

- The Python-based software package QAP[59], available from the Preprocessed Connectomes Project, computes many subject-level diagnostics of motion artefact.

- The script afni_proc.py bundled with AFNI provides similar functionality.

- The MATLAB-based software packages CONN[66] and Connectome Computation System[67] implement many motion correction strategies and subject-level diagnostics.

- The Python-based software package fmriprep implements robust pre-processing, co-registration, and confound modelling (Steps 2–20) with minimal user overhead for data sets in the BIDS format[68].

- The rs-fMRI repository used in a recent benchmarking study[23] (https://github.com/lindenmp/rs-fMRI) provides a MATLAB-based implementation of the complete protocol that requires some manual editing by users to run on new data sets.

Input data.—Structural and functional images should be named in a systematic manner. The BIDS standard[69] presents one recommended way to organise data in a consistent manner.

PROCEDURE

Preparing a subject list

TIMING 1—10 minutes

1. When using XCP Engine, a subject list (or "cohort file") supports iteration and uniform processing across the entire sample. An example is provided in Box 3, along with basic specifications. Detailed specifications are available in the online documentation (https://pipedocs.github.io/config/cohort.html). To prepare a subject list, associate each subject's functional time series with that subject's processed anatomical data (segmentation, transforms, and brain-extracted, bias-corrected image). For each subject, write a comma- or tab-delimited row of values to the subject list file. Each subject's values can include any information that could facilitate processing of the subject's data, including subject identifiers and input paths. However, values should be consistent across subjects – if the third column (value) for subject 1 contains the path to subject 1's functional time series, then the third column for all subjects should contain the path to those subjects' functional time series. All processing steps described below under the headings 'Minimal pre-processing', 'Co-registration', 'Confound modelling', 'Confound regression', and 'Performance assessment: subject level' should be performed separately for each subject, while 'Performance assessment: sample level' should be performed for the analytic sample as a whole by pooling outputs from all subjects.

Minimal pre-processing

> **TIMING 5—15 minutes per subject**—XCP's prestats module implements minimal pre-processing.

> 2. *Discard the initial volumes of the functional time series.* This can be achieved using fslroi, specifying the functional time series as <input>, the number of volumes to be discarded as <tmin>, and −1 (indicating retention of all remaining volumes) as <tsize>:

```
fslroi <input> <output> <tmin> <tsize>
```

> **CRITICAL STEP** The number of volumes to discard will depend on the acquisition parameters of the functional data and on the specifications of the scanner. This step is unnecessary if initial volumes have already been discarded as part of the acquisition protocol.

> 3. *Estimate framewise head motion* by computing the realignment parameters (RPs) and the overall framewise displacement (FD) of the head. Estimating the realignment parameters will require selection of an exemplar volume from the 4D time series, to serve as a reference for realignment. By default, this will be the midpoint of the time series. The realignment parameters can be computed using FSL's mcflirt, specifying the time series as input[70]:

```
mcflirt -in <input> -out <output root> -reffile <reference for
realignment> -plots -rmsrel -rmsabs -spline_final
```

> The rmsrel and rmsabs flags should be toggled in order to compute additional summary measures of the head's framewise and overall displacement. The plots flag configures the program to generate a visualisation of motion estimates, while spline_final specifies the interpolation approach. The output with suffix '.par' contains the 6 realignment parameters in 6 columns, while the output suffixed 'rel.rms' contains the framewise displacement in a single column.

> **CRITICAL STEP** By convention, the displacement of the head in the first acquired frame is defined to be 0. By default, however, mcflirt does not write out the framewise displacement for the first frame. To ensure that temporal masking (Step 8) selects the correct frames for removal, the default output from mcflirt should be padded with a leading 0.

> 4. (Optional) *Apply correction for timing of slice acquisition.* This can be achieved using FSL's slicetimer. Determine the most appropriate call to slicetimer by consulting the acquisition parameters of the data.

> **CRITICAL STEP** Interpolation steps, such as slice-time correction and despiking, affect estimates of realignment parameters[71]. To ensure that realignment parameters are computed independently of these procedures, they

should not be performed until after motion parameters have been estimated. However, proper slice-time correction requires information about the 2D slice in which each voxel was acquired. Because motion realignment re-slices the time series, slice-time correction should be performed *before* realignment but *after* estimation of realignment parameters. Accordingly, the input to slice-time correction should be the output of Step 2, and not the output of Step 3.

5. *Realign all volumes relative to a reference*. As in Step 3, mcflirt should be called:

```
mcflirt -in <input> -out <output root> -reffile <reference for
realignment> -mats -spline_final
```

Use a different <output root> from the <output root> in Step 3 to ensure that the realignment parameters are not overwritten. Because realignment parameters have already been computed, it is also possible to select a reference volume on the basis of low framewise head motion or to use the average of all volumes as the reference for this iteration of realignment. Without use of slice-time correction, Steps 3 and 5 can be combined into a single step.

6. *Classify all image voxels as brain or non-brain*, and delete non-brain values from the image. This step can be completed by passing the exemplar volume to the FSL command-line tool bet to generate a binary brain mask, and afterward multiplying the processed functional time series by the brain mask in fslmaths to produce a skull-stripped time series.

7. *Compute standardised DVARS*, an index of the frame-to-frame signal change across the brain that can be used to flag outliers[11,72,73]. Use the skull-stripped time series from Step 6 as the input to the XCP utility script dvars:

```
${XCPEDIR}/utils/dvars -i <input> -o <output>
```

dvars will produce three outputs. The output with suffix std.1D contains standardised DVARS.

8. (Optional; skip this step if censoring is not performed.) *Compute temporal masks and generate spike regressors*. A temporal mask is a binary-valued time series that is equal in length to the functional time series and that indicates whether each frame of the functional time series should be preserved or excised[11,18] (censored). Before proceeding, determine whether temporal censoring is a reasonable processing strategy for the dataset and analytic goals (see EXPERIMENTAL DESIGN). Next, select criteria and thresholds for censoring (for instance, framewise displacement > 0.2mm or standardised DVARS > 2). Temporal masks can be generated for each criterion using either 1d_tool.py in AFNI or the XCP utility tmask.R:

```
${XCPEDIR}/utils/tmask.R -s <time series> -t <threshold> -o
<output>
```

To prepare a temporal mask, use the criterion <time series> (rel.rms, the
framewise motion estimates from Step 2, and optionally the DVARS estimates
from Step 7) and the censoring <threshold> for each criterion. Define the final
temporal mask as the union of all single-criterion temporal masks. After
obtaining the final temporal mask, derive spike regressors[17,51] from this mask by
using the XCP utility tmask2spkreg.R, specifying the temporal mask as
<tmask>:

```
${XCPEDIR}/utils/tmask2spkreg.R -t <tmask> >> <output spike
regressors>
```

CRITICAL STEP Although censoring has been established as an efficacious
motion correction strategy, the exact threshold and criteria for determining
whether a frame should be censored vary from study to study. For example, in
typical single-band fMRI studies with TRs of 3s, censoring any volumes with FD
> 0.1 mm is very stringent, while censoring volumes with FD > 0.25 mm is less
stringent but still effective. Furthermore, thresholds for framewise censoring
criteria have typically been reported using units per frame (e.g., mm per frame
for FD). However, it should be emphasized that 0.25 mm FD during a 1 s frame
does not reflect the same degree of motion as does 0.25 mm FD during a 3 s
frame.

CRITICAL STEP There are several different but highly correlated ways to
define FD. Here, we refer to FD as defined by the output of mcflirt ($FD_{Jenkinson}$).
An alternative formulation of FD, FD_{Power}, can be computed by passing the
realignment parameters from Step 2 to the XCP utility fd.R; this measure can
then be used as a criterion for temporal masking. FD_{Power} and $FD_{Jenkinson}$ are
highly correlated, with FD_{Power} equal to approximately twice $FD_{Jenkinson}$.[25,54]

9. (optional) *Despike the time series*. Despiking is a spatially adaptive interpolation
    approach that identifies voxelwise signal outliers and imputes new values for
    each of those outliers. Despiking can be performed using AFNI's 3dDespike
    program:

```
3dDespike -prefix <output> -nomask -NEW <input>
```

Note that alternative de-spiking techniques exist, including wavelet-based
despiking[74]. Additionally, the relationship between despiking and temporal
censoring is not characterised; it is likely that both techniques target the same
types of noise using different identification criteria. As a result, despiking is

often used as an alternative to temporal censoring, while combining despiking and temporal censoring in the same processing stream is uncommon.

10. *Remove the mean and any linear or polynomial trends* from the functional time series using the XCP utility dmdt.R:

```
${XCPEDIR}/utils/dmdt.R -i <img> -d <order> -o <output> -m <mask> -
x <mean image> [-t <tmask>]
```

specifying the skull-stripped time series from Step 6 as <img>, the brain mask from Step 6 as <mask>, and (if applicable) the temporal mask from Step 8 as <tmask>. De-meaning and de-trending operates by fitting a polynomial basis to the input time series using a general linear model and afterward discarding the explained variance. Any frames flagged for censoring in the temporal mask are not considered in the fit[18]. The de-trend <order> refers to the highest order polynomial term included in the model, which can either be defined *a priori* or estimated using the formula

$$1 + \left\lfloor \frac{T_R * n}{150} \right\rfloor$$

where T_R is the sampling time of the functional time series in seconds and n is the number of frames sampled[48]. Toggle the -x option to generate a mean functional image, defined as the fit of the constant term of the model, and define an appropriate output path.

CRITICAL STEP Save the last output produced during minimal preprocessing, as useful diagnostic information can be obtained from it.

Co-registration

TIMING 20—45 minutes per subject—XCP's coreg module implements co-registration.

11. *Identify the white matter boundary*. Use the tissue-class segmentation to obtain a binary mask indicating whether each voxel in the anatomical image represents white matter (WM) tissue. If necessary, the tissue-specific WM mask can be generated with the XCP utility val2mask.R, using the anatomical segmentation as the reference image and selecting all values that correspond to WM as the values of interest:

```
${XCPEDIR}/utils/val2mask.R -i <input> -v <WM values> -o <output>
```

12. *Co-register the functional image to a high-resolution anatomical reference* acquired for the same subject. Use FSL's flirt to compute the co-registration, specifying either the reference volume from Step 4 or the mean image from Step

10 as the input (<source>) and the subject's anatomical image as the <reference>:

```
flirt -in <source> -ref <reference> -dof <degrees of freedom> -out
<output image> -omat <output matrix> -cost bbr -wmseg <white
matter segmentation volume>
```

Set the <degrees of freedom> to 6 (or alternatively 9 if distortion correction has not been performed), and ensure that an <output matrix> path is defined. Specify bbr (boundary-based registration) as the cost function, and use the binary WM mask from Step 11 as the <white matter segmentation volume>[75]. After the co-registration is performed, use FSL's convert_xfm to calculate the inverse of the <output matrix> (a .mat file) specifying the <output matrix> from the forward transformation as the argument to the inverse flag:

```
convert_xfm -omat <output inverse> -inverse <matrix>
```

CRITICAL STEP Because the co-registration computed here will later be necessary for computing tissue-based confounds, the quality of the co-registration should be assessed before proceeding. The most reliable way to assess co-registration quality is through expert visual inspection of all registered images. However, co-registration quality can also be quantified using spatial cross-correlation and coverage metrics. Cross-correlation can be computed using fslcc, while coverage can be computed as the percentage of the binarised structural image that also lies within the binarised co-registered image or as the Dice and Jaccard coefficients between the binarised structural and co-registered images. All quality indices can be obtained using the XCP utility maskOverlap.R:

```
${XCPEDIR}/utils/maskOverlap.R -m <co-registered functional image>
-r <anatomical image>
```

After computing these values for all subjects, they can be used to flag outliers for more thorough visual inspection.

13. (Optional; skip this step if using any program other than ANTs to warp images between coordinate spaces.) *Create ITK-compatible versions of the forward and inverse co-registration matrices.* This will enable use of the transforms with ANTs. Converting the FSL-based co-registration matrices to ITK format can be accomplished using c3d_affine_tool with the fsl2ras and oitk flags, specifying each .mat file from Step 12 as the <transform> file, providing source and reference images for each transform as for flirt, and specifying the output path using the oitk flag:

```
c3d_affine_tool -src <source> -ref <reference> <transform> -
fsl2ras -oitk <output>
```

To verify that the conversion was successful, apply the reformatted transform to the <source> image from Step 12 and evaluate whether the output is identical to the <output image> from Step 12.

Confound modelling

TIMING 1—5 minutes per subject without ICA-AROMA, 20—30 minutes per subject with ICA-AROMA XCP's confound module implements confound modelling.

14. (Optional) *Use ICA-AROMA* to decompose the pre-processed data into linearly independent signal sources, identify motion-related components, and remove them from the data[20]. Additional documentation and detailed instructions for implementation are available online along with the ICA-AROMA code (https://github.com/rhr-pruim/ICA-AROMA/blob/master/Manual.pdf).

15. *Compute the mean global signal*[26]. Use the fslmeants function, providing the functional image as the <input> and providing the whole-brain mask from Step 6 as the <mask>:

```
fslmeants -i <input> -o <output> -m <mask>
```

16. *Compute eroded white matter (WM) and cerebrospinal fluid (CSF) masks.* The XCP utility script erodespare erodes a mask such that only a user-specified percentage of the deepest tissue in the original mask remains:

```
${XCPEDIR}/utils/erodespare -i <input> -o <output> -r <retention
criterion> -v <value set>
```

It is recommended that masks be eroded to retain 5–10 percent of their original size (<retention criterion> of 5 – 10) so as to minimise partial volume effects. Use the anatomical segmentation as the <input>. For the <value set>, enter the values that correspond to the WM or CSF label in the anatomical segmentation. Finally, warp the output eroded masks into the same coordinate space as the functional time series.

CRITICAL STEP Because the WM and CSF signals mix with the grey matter (GM) signal near the GM interface, the WM and CSF masks derived from the anatomical segmentation will include influence from adjoining GM voxels[18,28]. This influence should be mitigated by eroding WM and CSF masks to exclude boundary voxels; more liberal erosion will result in WM and CSF signal estimates that are more independent of the global signal. However, erosion

should leave enough voxels in each mask in order to spatially sample the tissue class of interest.

17. *Compute nuisance signals in WM and CSF*. Several standard options are available for extracting nuisance signals from the WM and CSF.

    A.   *Compute mean signals within the white matter (WM) and cerebrospinal fluid (CSF)*, and add these to the motion model. These can be computed as in Step 15, substituting the eroded tissue-specific masks from Step 16 for the whole-brain mask.

    B.   *Use anatomical CompCor*[16,19]. CompCor decomposes the signals in the eroded WM and CSF masks using principal component analysis (PCA). Select the first several component time series from the decomposition such that 50 percent of the variance in WM and CSF compartments is explained in each case[19]. PCA can be implemented using AFNI's 3dpc utility, with the cumulative variance explained indexed in the <output>_eig.1D output file under the cumulative fraction heading, and with the principal component time series saved in an output file called either <output>_vec.1D or <output>.1D. After determining the correct number of components, re-run 3dpc, this time specifying the number of components to be included using the -pcsave option.

```
3dpc -prefix <output> -mask <eroded mask> [-pcsave <number
of PCs>] <input>
```

18. (Optional) *Prepare any custom nuisance time series*. These can include any additional nuisance time series available for the current data set, including recordings of artefactual processes such as pulse and respiration or nuisance time series defined separately for each voxel (as well as expansions of these where appropriate)[76]. If the data are being processed for a task-constrained functional connectivity analysis, the task model can be included in the confound model depending on analytic objectives[77].

19. (Optional) *Compute time series expansions*[15]. For each time series in the motion model, compute its temporal derivative, its quadratic term, and the temporal derivative of its quadratic term. Note that adding a previous time point to the confound model and adding a temporal derivative to the confound model for the purposes of confound regression are equivalent, as the temporal derivative of a discretely sampled time series can be expressed as the difference between the original time series and the backwards shifted time series (a linear combination). To ensure that the confound model is not over-specified due to collinearity, a single confound model should not include both derivatives and temporal shifts.

20. *Build the confound model into a matrix*. Concatenate the time series produced in Steps 3 and 15–19 into a model matrix, for instance using AFNI's 1dcat utility.

CRITICAL STEP Do not add the spike regressors to the confound model yet, as they should not be subjected to any temporal filtering.

CRITICAL STEP To verify that WM and CSF signals are sufficiently independent from the global signal, it is helpful to estimate the collinearity of all predictor variables. A matrix of correlations among predictor variables can be obtained using the XCP utility ts2adjmat.R, specifying the complete motion model as the input time series. (This matrix will be written in the form of a feature vector, which can be converted to a symmetric square matrix using the squareform function in SciPy, Matlab, or the R library pracma.)

Confound regression

TIMING 5—10 minutes per subject without censoring, 60 minutes per subject with censoring XCP's regress module implements confound regression.

14. (Optional; skip this step if censoring is not performed.) *Interpolate over epochs marked for censoring.* Temporal filtering (Step 22) can result in propagation of artefactual variance from frames flagged for censoring into adjoining frames[18,57,58]. Interpolation of values in contaminated frames can help make the temporal filter robust to noise in censored frames while preserving the overall autocorrelation structure of the data. Currently, the recommended approach[18] is based on the Lomb-Scargle periodogram[78]; this approach generates surrogate data based on the spectral characteristics observed in low-noise (unflagged) frames. This approach is, however, computationally intensive. Less rigorous but faster interpolation can be achieved using neighbour-based approaches, which sacrifice the data's autocorrelation structure[57,58]. The periodographic approach is currently implemented in the XCP utility interpolate.R.

15. *Apply a temporal filter to the data.* A commonly used pass-band ranges from a high-pass limit of 0.01 Hz to a low-pass limit of 0.08 Hz, above which the power spectra of high- and low-motion subjects diverged even when a high-performance confound model was used[17].

     CRITICAL STEP Ensure that the same filter is applied to both the time series data and the confound model to prevent spectral misspecification, which can result in reintroduction of artefactual variance into suppressed frequency bands during the model fit stage[79] (Step 24).

16. *Concatenate the spike regressors* computed in Step 8 into the filtered model matrix, as in Step 20.

17. *Perform confound regression.* Use a general linear model to estimate the contributions of artefactual signals to the BOLD signal at each voxel. The AFNI program 3dTproject computes the model fit and returns the residuals as <output>:

```
3dTproject -input <input> -ort <regressors> [-dsort <voxelwise>] -
prefix <output>
```

Provide the filtered time series as <input> and the filtered confound model as <ort>. Any voxelwise-defined regressors should be provided as <dsort>, and the desired output path should be provided as <output>. (It is also possible to combine Steps 22, 23, and 24 into a single step using 3dTproject. Refer to the documentation for details.)

18. (Optional; skip this step if censoring is not performed.) *Censor the data*[11,18]. Censoring can be performed using the XCP utility script censor.R, providing the temporal mask as <tmask> and the time series residuals as <img>:

```
${XCPEDIR}/utils/censor.R -i <img> -t <tmask> -o <output>
```

(Optional) Re-compute the mean framewise displacement after censoring. In Step 35, residual motion artefact is estimated as the correlation between mean framewise displacement and connectivity. Using the post-censoring framewise displacement in this computation links the outcome value more directly to the residual motion in the dataset but reduces the variance in levels of motion that is present in the pre-censoring data.

CRITICAL STEP Censoring will disrupt the autocorrelation structure of the data. Therefore, any processing steps that are sensitive to autocorrelation structure (e.g., temporal filtering or dynamic connectivity analysis) should either be performed prior to censoring or be performed in a way that accounts for disruptions in the autocorrelation structure.

19. (Optional) *Re-mean the time series* by adding the mean image computed in Step 10 to the residualised data using fslmaths.:

```
fslmaths <input> -add <mean image> <output>
```

20. (Optional) *Spatially smooth the image data.* Smoothing across tissue boundaries can be reduced by using an adaptive filter that combines smoothing and edge detection. This can be achieved using the FSL program susan, providing the processed time series as input[80]:

```
susan <input> <bt> <kernel> 3 1 1 <usan> <bt> <output>
```

Specify the mean image from Step 10 as the <usan> (the image over which tissue boundaries are detected). The <bt> (brightness thresholds) can be estimated as 75 percent of the median intensity value in the <usan>. To obtain the median intensity value, use fslstats to compute the 50th percentile, limiting the

computation to voxels inside the brain by using the brain mask with the -k parameter:

```
fslstats <usan> -k <mask> -p 50
```

Alternatively, to completely restrict smoothing across tissue boundaries, warp the anatomical segmentation into the same coordinate space as the processed image and then use this as the <usan>, setting the <bt> arbitrarily low.

Performance assessment: subject level

TIMING 5—10 minutes per subject—XCP's fcon and qcfc modules implement performance assessment.

21. *Compute the functional connectome within a selected parcellation.* To do this, first select a parcellation and warp it into the same coordinate space as the processed image. We suggest using multi-label interpolation rather than nearest neighbour (using antsApplyTransforms). Next, compute the local time series for each parcel using the XCP utility roi2ts.R, providing the parcellation as the roi and providing a list of regions as labels if one is available. Finally, obtain the adjacency matrix for the connectome using the XCP utility ts2adjmat.R, providing the local time series as the ts argument.

22. *Prepare a depth map for the anatomical segmentation.* A depth map is a voxelwise map of the brain where the value of each voxel indicates the voxel's depth within its assigned tissue class; this map will be used to create a voxelwise summary plot of the data (Figure 2). (In the depth map, the first digits of the voxel's value indicate its tissue class, while the last two digits indicate its depth in that tissue class, as a percentile. For instance, if WM is assigned a value of 3 in the anatomical segmentation, a superficial WM voxel at the 30th percentile of depth would be assigned a value of 330, while a deep WM voxel at the 98th percentile of depth would be assigned a value of 398.) This depth map can be generated using the XCP utility layerLabels, defining the input label set as the segmentation:

```
${XCPEDIR}/utils/layerLabels -i <label set> -o <output>
```

Next, warp the depth map into the same coordinate space as the minimally preprocessed image and as the final processed image using nearest-neighbour interpolation.

23. *Prepare a subject-level voxelwise summary plot* for the minimally preprocessed data and for the de-noised data[55]:

```
${XCPEDIR}/utils/voxts.R -i <img1>,<img2> -r <roi> -o <output> -t
FD:<FD>:<FD threshold>,DV:<DVARS>:<DVARS threshold>
```

Use the depth map from Step 29 as the <roi>. Provide the minimally preprocessed image as <img1> and the final, de-noised image as <img2>. Provide the path to the framewise displacement time series as <FD> and the path to the DVARS time series as <DVARS>. If the data were censored based on FD or DVARS, then provide the censoring thresholds as <FD threshold> and the DVARS threshold as <DVARS threshold>. Otherwise, select reasonable thresholds for flagging outliers. Label names can be found at ${XCPEDIR}/atlas/segmentation3 for 3-tissue class segmentations and ${XCPEDIR}/atlas/segmentation6 for 6-tissue class segmentations. Example output is presented in Figure 2.

24. *Compute the DVARS of the denoised data set* as in Step 7. Indicate that the input data are demeaned using the -d 1 argument and provide the skull-stripped mean image from Step 10 as the -b argument:

```
${XCPEDIR}/utils/dvars -i <input> -o <output root> -b <mean image>
-d 1
```

25. *Compute FD-DVARS correlations*[19] using the XCP utility featureCorrelation.R:

```
${XCPEDIR}/utils/featureCorrelation.R -i <path to framewise
displacement time series>,<path to standardised DVARS time series>
>> <output correlation>
```

Perform the calculation both for the minimally pre-processed DVARS from Step 7 and the denoised DVARS from Step 31.

26. *Estimate the loss of temporal degrees of freedom.* This loss can be approximated as the total number of terms in the confound model (columns in the model matrix from Step 23), including spike regressors (if censoring was performed) and independent components identified as noise (if ICA-AROMA was run).

Performance assessment: sample level

TIMING 20 minutes—XCP's fcon and qcfc modules implement performance assessment.

21. *Prepare a sample matrix* with P + 2 columns, where P is the number of unique identifiers (e.g., subject identifier, scan identifier, session identifier) for each subject. Place each subject in a separate row, and place each identifier in a separate column. Each column containing identifiers should have a header beginning with the string id, while the remaining 2 columns should have the headers motion and connectivity. In each subject's motion column, enter the

subject's mean framewise displacement. This can be obtained from a file with the suffix rel_mean.rms produced in Step 2. In each subject's connectivity column, enter the path to the subject's connectivity matrix, the final output produced in Step 26. Save the matrix in .csv format. An example sample matrix is presented in Box 4.

22. *Compute the distribution of residual QC-FC correlations*[22,28,54]. This step can be facilitated by using the XCP utility qcfc.R (usage examples in Box 4). The utility script computes, for each edge in the connectome, the partial correlation of motion with the strength of that edge, after controlling for the effects of any user-provided covariates. Each row in the sample matrix from Step 34 represents an observation for the correlation.

CRITICAL STEP The number of connections significantly related to motion is sensitive to the total sample size. Accordingly, it is important to ensure adequate statistical power for detection of motion-related variance. When evaluating the comparative efficacy of a novel denoising procedure, it is especially critical to use a sample that is sufficiently large to detect residual artefact. Refer to previous benchmarking studies to guide sample selection[22,23].

23. *Compute the distance-dependence of QC-FC correlations*[10,11,17]. QC-FC distance-dependence can be computed using the XCP utility qcfcDistanceDependence. This step entails (a) computing a pairwise matrix of distances between each pair of parcels and (b) computing the correlation between distance and residual QC-FC. qcfcDistanceDependence performs both of these steps using the warped parcellation from Step 28 and the QC-FC values from Step 35:

```
${XCPEDIR}/utils/qcfcDistanceDependence -a <parcellation> -q <QC-
FC values> -d <output distance matrix> -o <output distance-
dependence> -f <output figure>
```

? TROUBLESHOOTING

Troubleshooting advice can be found in Table 3.

TIMING

Step 1, preparing a subject list: 1–10 minutes.

Steps 2—10, minimal pre-processing: 5–15 minutes per subject, depending on the number of samples and spatial resolution of the input time series; potentially significantly longer for high-resolution multiband time series of long duration.

Steps 11—13, co-registration: 20—45 minutes per subject.

Steps 14—20, confound modelling: 1–5 minutes per subject, 20—30 minutes or longer if ICA-AROMA is included, scaling with the number of frames and spatial resolution of the input time series.

Steps 21—27, confound regression: 5—10 minutes per subject, 60 minutes per subject if censoring is included.

Steps 28—33, subject-level performance assessment: 5—10 minutes per subject.

Steps 34—36, sample-level performance assessment: 20 minutes.

ANTICIPATED RESULTS

The expected product of the protocol is a set of denoised functional time series and connectomes, one for each input to the processing stream. While some motion artefact is likely to be present even in images processed through a high-performance denoising stream, the residual artefact will be markedly mitigated in comparison with the artefact present in a time series that has been processed minimally or using a less effective stream (see Figure 3 for an illustration). If GSR is included in the denoising model, then denoised connectomes can be expected to feature a considerable fraction of negatively weighted connections, along with a relatively symmetric and approximately zero-centred distribution of connection weights. The diagnostics produced in the performance assessment stages of the protocol can be used to determine whether the final result of denoising is appropriate. Example results from subject-level and group-level diagnostic steps are summarised in Figures 2 and 3, respectively. While the details are beyond the scope of this protocol, it should be noted that XCP uses this de-noised data as part of additional modules for generation of seed-based connectivity maps[1], network partitions[81,82], regional homogeneity maps[83], and amplitude of low-frequency fluctuation maps[84].

Example use cases for denoising of subject data with XCP and for assessment of denoising performance are downloadable from FigShare (https://figshare.com/s/d0161bac47f98eb1830b), together with anticipated results. The provided examples include input images, a cohort file, expected outputs from subject-level denoising, and a set of 101 functional brain networks that can be used to validate the functionality of qcfc.R and qcfcDistanceDependence.R.

ACKNOWLEDGMENTS

REFERENCES

1. Biswal B, Yetkin FZ, Haughton VM & Hyde JS Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med 34, 537–541 (1995). [PubMed: 8524021]

2. Biswal BB et al. Toward discovery science of human brain function. Proc. Natl. Acad. Sci 107, 4734–4739 (2010). [PubMed: 20176931]

3. Power JD et al. Functional network organization of the human brain. Neuron 72, 665–678 (2011). [PubMed: 22099467]

4. Yeo BTT et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165 (2011). [PubMed: 21653723]

5. Raichle ME et al. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A 98, 676–682 (2001). [PubMed: 11209064]

6. Sharma A et al. Common dimensional reward deficits across mood and psychotic disorders : a connectome-wide association study. Am. J. Psychiatry 174, 657–666 (2017). [PubMed: 28135847]

7. Gu S et al. Emergence of system roles in normative neurodevelopment. Proc. Natl. Acad. Sci 201502829 (2015). doi:10.1073/pnas.1502829112

8. Dosenbach NUF et al. Prediction of individual brain maturity using fMRI. Science (80-.). 329, 1358–1361 (2011).

9. Fair DA et al. The maturing architecture of the brain's default network. Proc. Natl. Acad. Sci. U. S. A 105, 4028–4032 (2008). [PubMed: 18322013]

10. Satterthwaite TD et al. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632 (2012). [PubMed: 22233733]

11. Power JD, Barnes KA, Snyder AZ, Schlaggar BL & Petersen SE Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154 (2012). [PubMed: 22019881]

12. Van Dijk KRA, Sabuncu MR & Buckner RL The influence of head motion on intrinsic functional connectivity MRI. Neuroimage 59, 431–438 (2012). [PubMed: 21810475]

13. Power JD, Plitt M, Laumann TO & Martin A Sources and implications of whole-brain fMRI signals in humans. Neuroimage (2016). doi:10.1016/j.neuroimage.2016.09.038

14. Fair DA et al. Development of distinct control networks through segregation and integration. Proc. Natl. Acad. Sci 104, 13507–13512 (2007). [PubMed: 17679691]

15. Friston K, Williams S, Howard R, Frackowiak RSJ & Turner R Movement-Related Effects in {fMRI} Time-Series. Magn. Reson. Med 35, 346–355 (1996). [PubMed: 8699946]

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

17. Satterthwaite TD et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. Neuroimage 64, 240–256 (2013). [PubMed: 22926292]

18. Power JD et al. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84, 320–341 (2014). [PubMed: 23994314]

19. Muschelli J et al. Reduction of motion-related artifacts in resting state fMRI using aCompCor. Neuroimage 96, 22–35 (2014). [PubMed: 24657780]

20. Pruim RHR et al. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. Neuroimage 112, 267–277 (2015). [PubMed: 25770991]

21. Salimi-Khorshidi G et al. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468 (2014). [PubMed: 24389422]

22. Ciric R et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 154, 174–187 (2017). [PubMed: 28302591]

23. Parkes L, Fulcher BD, Yucel M & Fornito A An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. bioRxiv (2017).

24. Burgess GC et al. Evaluation of denoising strategies to address motion-correlated magnetic resonance imaging data from the Human Connectome Project. Brain Connect. (2016). doi:10.1089/brain.2016.0435

25. Yan CG et al. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. Neuroimage 76, 183–201 (2013). [PubMed: 23499792]

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

27. Fox MD, Zhang D, Snyder AZ & Raichle ME The global signal and observed anticorrelated resting state brain networks. J. Neurophysiol 101, 3270–3283 (2009). [PubMed: 19339462]

28. Power JD, Schlaggar BL & Petersen SE Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105, 536–551 (2015). [PubMed: 25462692]

29. Greene DJ et al. Behavioral interventions for reducing head motion during MRI scans in children. Neuroimage 171, 234–245 (2018). [PubMed: 29337280]

30. 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. Magn. Reson. Imaging 34, 1206–1219 (2016). [PubMed: 27451407]

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

32. Kundu P et al. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. Proc. Natl. Acad. Sci. Unites States Am 110, 16187–16192 (2013).

33. Dosenbach NUF et al. Real-time motion analytics during brain MRI improve data quality and reduce costs. Neuroimage 161, 80–93 (2017). [PubMed: 28803940]

34. Power JD et al. Ridding fMRI data of motion-related influences: Removal of signals with distinct spatial and physical bases in multiecho data. Proc. Natl. Acad. Sci 201720985 (2018). doi:10.1073/pnas.1720985115

35. Satterthwaite TD et al. Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. Neuroimage 83, 45–57 (2013). [PubMed: 23792981]

36. Zeng L-L et al. Neurobiological basis of head motion in brain imaging. Proc. Natl. Acad. Sci. U. S. A 111, 6058–62 (2014). [PubMed: 24711399]

37. Van Essen DC et al. The WU-Minn Human Connectome Project: An overview. Neuroimage 80, 62–79 (2013). [PubMed: 23684880]

38. Murphy K & Fox MD Towards a consensus regarding global signal regression for resting state functional connectivity MRI. Neuroimage 154, 169–173 (2017). [PubMed: 27888059]

39. 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 44, 893–905 (2009). [PubMed: 18976716]

40. Chai XJ, Castañán AN, Öngür D & Whitfield-Gabrieli S Anticorrelations in resting state networks without global signal regression. Neuroimage 59, 1420–1428 (2012). [PubMed: 21889994]

41. Saad ZS et al. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain Connect. 2, 25–32 (2012). [PubMed: 22432927]

42. Yan C-G, Craddock RC, He Y & Milham MP Addressing head motion dependencies for small-world topologies in functional connectomics. Front. Hum. Neurosci 7, 910 (2013). [PubMed: 24421764]

43. Power JD, Laumann TO, Plitt M, Martin A & Petersen SE On Global fMRI Signals and Simulations. Trends Cogn. Sci 21, 911–913 (2017). [PubMed: 28939332]

44. Yang GJ et al. Altered global brain signal in schizophrenia. Proc. Natl. Acad. Sci. U. S. A 111, 7438–43 (2014). [PubMed: 24799682]
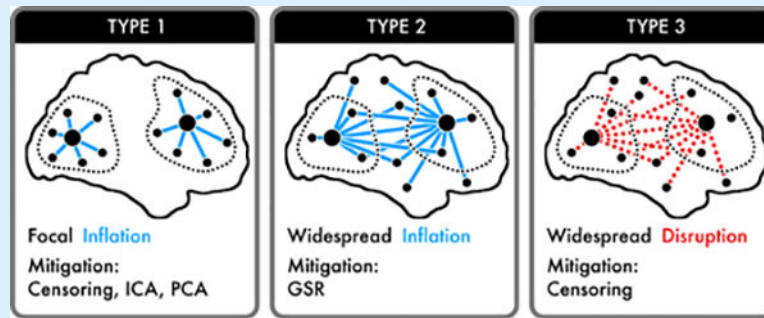
45. Hahamy A et al. Save the global: global signal connectivity as a tool for studying clinical populations with functional magnetic resonance imaging. Brain Connect. 4, 395–403 (2014). [PubMed: 24923194]

46. Carone D et al. Impact of automated ICA-based denoising of fMRI data in acute stroke patients. NeuroImage Clin. 16, 23–31 (2017). [PubMed: 28736698]

47. Tremblay M, Tam F & Graham SJ Retrospective coregistration of functional magnetic resonance imaging data using external monitoring. Magn. Reson. Med 53, 141–149 (2005). [PubMed: 15690513]

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

49. Avants BB et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033–2044 (2011). [PubMed: 20851191]

50. Smith SM et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, 208–219 (2004).

51. Lemieux L, Salek-Haddadi A, Lund TE, Laufs H & Carmichael D Modelling large motion events in fMRI studies of patients with epilepsy. Magn. Reson. Imaging 25, 894–901 (2007). [PubMed: 17490845]

52. Klein A et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46, 786–802 (2009). [PubMed: 19195496]

53. Tustison NJ et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 99, 166–179 (2014). [PubMed: 24879923]

54. Satterthwaite TD et al. Motion artifact in studies of functional connectivity: Characteristics and mitigation strategies. Hum. Brain Mapp 0, (2017).

55. Power JD A simple but useful way to assess fMRI scan qualities. Neuroimage 154, 150–158 (2017). [PubMed: 27510328]

56. Zuo X-N et al. An open science resource for establishing reliability and reproducibility in functional connectomics. Sci. data 1, 140049 (2014). [PubMed: 25977800]

57. Carp J Optimizing the order of operations for movement scrubbing: Comment on Power et alet al. Neuroimage 76, 436–438 (2013). [PubMed: 22227884]

58. Power JD, Barnes KA, Snyder AZ, Schlaggar BL & Petersen SE Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. Neuroimage 76, 439–441 (2013). [PubMed: 22440651]

59. Shehzad Z et al. The Preprocessed Connectomes Project Quality Assessment Protocol - a resource for measuring the quality of MRI data. Front. Neurosci (2015). doi:10.3389/conf.fnins.2015.91.00047

60. Rosen A et al. Data-driven Assessment of Structural Image Quality. bioRxiv 125161 (2017). doi:10.1101/125161

61. Gordon EM et al. Generation and evaluation of a cortical area parcellation from resting-state correlations. Cereb. Cortex 26, 288–303 (2016). [PubMed: 25316338]

62. Cammoun L et al. Mapping the human connectome at multiple scales with diffusion spectrum MRI. J. Neurosci. Methods 203, 386–397 (2012). [PubMed: 22001222]

63. Honnorat N et al. GraSP: Geodesic Graph-based Segmentation with Shape Priors for the functional parcellation of the cortex. Neuroimage 106, 207–221 (2015). [PubMed: 25462796]

64. Schaefer A et al. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb. Cortex 1–20 (2017). doi:10.1093/cercor/bhx179 [PubMed: 28365777]

65. Sikka S et al. Towards Automated Analysis of Connectomes: The Configurable Pipeline for the Analysis of Connectomes (C-PAC). Front. Neuroinform. (2012). doi:10.3389/conf.fninf.2014.08.00117

66. Whitfield-Gabrieli S & Nieto-Castanon A Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect. 2, 125–141 (2012). [PubMed: 22642651]

67. Zuo XN et al. Toward reliable characterization of functional homogeneity in the human brain: Preprocessing, scan duration, imaging resolution and computational space. Neuroimage 65, 374–386 (2013). [PubMed: 23085497]

68. Esteban O et al. FMRIPrep : a robust preprocessing pipeline for functional MRI. bioRxiv 1–20 (2018). doi:10.1101/306951

69. Gorgolewski KJ et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data 3, 1–9 (2016).

70. Jenkinson M, Bannister P, Brady M & Smith S Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841 (2002). [PubMed: 12377157]

71. Power JD, Plitt M, Kundu P, Bandettini PA & Martin A Temporal interpolation alters motion in fMRI scans: Magnitudes and consequences for artifact detection. PLoS One 12, 1–20 (2017).

72. Nichols TE Notes on Creating a Standardized Version of DVARS. 1–5 (2017).

73. Afyouni S & Nichols TE Insight and Inference for DVARS. bioRxiv 34 (2017). doi:10.1101/125021

74. Patel AX et al. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. Neuroimage 95, 287–304 (2014). [PubMed: 24657353]

75. Greve DN & Fischl B Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72 (2009). [PubMed: 19573611]

76. Jo HJ et al. Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state FMRI. J. Appl. Math 2013, (2013).

77. Cole MW, Bassett DS, Power JD, Braver TS & Petersen SE Intrinsic and Task-Evoked Network Architectures of the Human Brain. Neuron 83, 238–251 (2014). [PubMed: 24991964]

78. Lomb NR Least-squares frequency analysis of unequally spaced data. Astrophys. Space Sci 39, 447–448 (1976).

79. 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 82, 208–225 (2013). [PubMed: 23747457]

80. Smith SM & Brady JM SUSAN—A New Approach to Low Level Image Processing. Int. J. Comput. Vis 23, 45–78 (1997).

81. Blondel VD, Guillaume J-L, Lambiotte R & Lefebvre E Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp 10008, 6 (2008).

82. Girvan M & Newman MEJ Community structure in social and biological networks. Proc. Natl. Acad. Sci. U. S. A 99, 7821–7826 (2002). [PubMed: 12060727]

83. Zang Y, Jiang T, Lu Y, He Y & Tian L Regional homogeneity approach to fMRI data analysis. Neuroimage 22, 394–400 (2004). [PubMed: 15110032]

84. Yu-Feng Z et al. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. Brain Dev. 29, 83–91 (2007). [PubMed: 16919409]

85. Satterthwaite TD et al. Neuroimaging of the Philadelphia Neurodevelopmental Cohort. Neuroimage 86, 544–553 (2014). [PubMed: 23921101]

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

BOX 1 | Overview of confound models for motion correction.
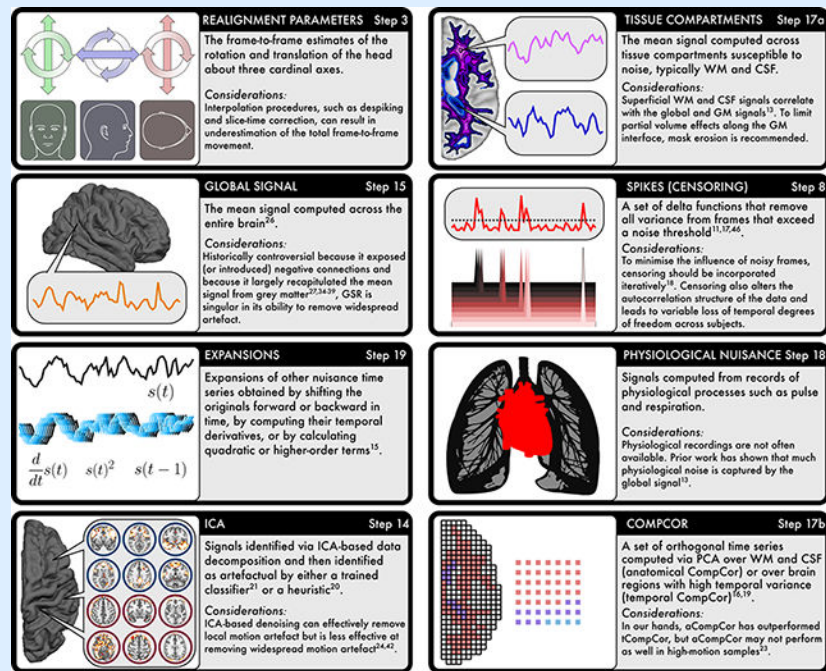
Numerous strategies have been used for removing motion-related variance from a functional MR time series. The schematic summarises several denoising options, along with the step of the protocol at which they can be added to the confound model, and discusses potential pitfalls and limitations of each approach. The design of a confound model suited to a particular data set and scientific objective is discussed under "Experimental Design".

BOX 2 | Taxonomy of motion artefacts in BOLD time series.

TYPE 1

Focal *Inflation*
Mitigation:
Censoring, ICA, PCA

TYPE 2

Widespread *Inflation*
Mitigation:
GSR

TYPE 3

Widespread *Disruption*
Mitigation:
Censoring

Motion artefact impacts functional connectivity data in three primary ways. This taxonomy, first introduced by Power and colleagues[28], is graphically summarised in the schematic shown. Type 1 effects occur when a movement drives the signal in proximal voxels homogeneously, resulting in spuriously inflated correlations among nearby regions. Type 2 effects, by contrast, occur when a movement globally drives the BOLD signal in a homogeneous manner, inducing widespread inflation of correlations. Finally, Type 3 effects occur when a movement induces heterogeneous signal fluctuations across the brain, disrupting correlations, particularly those between distal regions. In general, Type 2 artefacts are the most common and Type 3 are the least common. Notably, different confound regression strategies are more effective at targeting different features of motion artefact. For instance, global signal regression (GSR) models the widespread signal fluctuations that characterise Type 2 artefact. By contrast, strategies that model spatially localised artefactual variance, such as ICA and PCA, may be better suited for cleaning Type 1 artefact. Censoring approaches, which remove entire frames that have been contaminated by motion, remove the spatially varied Type 1 and Type 3 artefacts. (More aggressive censoring can also be effective against Type 2 artefacts[18].)

The contributions of different artefact types can also account for why motion artefact exhibits a *distance-dependent* spatial profile[10,11,17]: the connectivity among nearby regions of the brain is inflated by a combination of Type 1 and Type 2 effects, while the connectivity among more distant regions is inflated by Type 2 effects (and, to a lesser extent, disrupted by Type 3 effects). Consequently, motion on average inflates short-distance connections more than it does long-distance connections. Because GSR is highly effective at mitigating common Type 2 motion effects, its residual artefact profile is dominated by Type 1 effects; this is observed as elevated distance-dependence[22,28,54].

BOX 3 | The XCP cohort file

A pipeline cohort file defines the experimental sample — the set of subjects that the pipeline should process. In the XCP system, the cohort file is formatted as .csv and contains:

- A column for each identifier variable

- A column for each pipeline input

- A header naming each category of input

- A row corresponding to each subject

An example is provided below:

```
id0,id1,img,antsct
ACC,001,rawData/ACC_001_rest.nii.gz,processedData/ACC_001_antsct
DSQ,001,rawData/DSQ_001_rest.nii.gz,processedData/DSQ_001_antsct
DSQ,002,rawData/DSQ_001_rest.nii.gz,processedData/DSQ_002_antsct
CAT,001,rawData/CAT_001_rest.nii.gz,processedData/CAT_001_antsct
```

The header for identifier columns should begin with id and end with a non-negative integer. Identifiers should be ordered hierarchically – for instance, id0 could correspond to the subject identifier, id1 to the session identifier, and id2 to the scan identifier within the session. The primary image to be processed should be listed under the header img, and the data directory that directly contains all anatomical inputs (anatomical image, transforms, and segmentation) should be listed under the header antsct.

If the data are to be processed using the CBICA IPP, then the paths in the cohort file should not include any directories, since those directories will not be uploaded to IPP. For the archived and compressed anatomical processing directory, the extension (i.e., .tar.gz) should not be included in the path that is specified in the cohort.

> ## BOX 4 | Example subject list, model specification, and output files for qcfc.R
>
> The sample-level performance assessment script qcfc.R has a number of options for model, input, and output specification:
>
> ```
> ${XCPEDIR}/utils/qcfc.R -c <cohort> -o <output root> [-s <multiple
> comparisons correction> -t <significance threshold> -n <confound> -y
> <conformula>]
> ```
>
> Optional arguments are denoted in square brackets ([]). The primary input to qcfc.R (<cohort>) should be the subject list created in Step 34, an example of which is provided here:
>
> ```
> id0,id1,motion,connectivity
> ACC,001,0.0391,processedData/ACC_001_fc/connectome.txt
> ACC,002,0.0455,processedData/ACC_002_fc/connectome.txt
> ACC,003,0.0367,processedData/ACC_003_fc/connectome.txt
> DSQ,001,0.1532,processedData/DSQ_001_fc/connectome.txt
> CAT,001,0.0811,processedData/CAT_001_fc/connectome.txt
> ```
>
> The type of correction for multiple comparisons can be specified as fdr (default), bonferroni, or none. The maximal p-value threshold for significance can also be specified by the user; in the absence of user input, a default value of 0.05 will be used.
>
> The values of any model covariates (such as age and sex) should be included in another file containing the same subject identifiers as the sample matrix. The file containing model covariates should be provided as the <confound> argument, and the formula for the model should be provided as the <conformula> argument, with any categorical variables specified as factors (see example below). (If the user wishes to obtain only the direct correlation between motion and functional connectivity, then no formula or covariates file is necessary.)
>
> For example, to control for the participants' age and sex when computing motion effects, prepare a second file (<confound>) containing the same identifiers as the first, with additional columns for each of the covariates to be considered. In the example below, age is defined in months and sex is coded as a binary variable:
>
> ```
> id0,id1,age,sex
> ACC,001,217,0
> ACC,002,238,1
> ACC,003,238,1
> DSQ,001,154,0
> CAT,001,176,1
> ```

If this file is saved as sample-covariates.csv, then call qcfc.R as above, with the additional arguments <confound> set to sample-covariates.csv and <conformula> set to:

```
'age+factor(sex)'
```

Note that the categorical variable sex is specified as a factor. If the data set contains repeated measures (e.g., multiple scans from the same subject), then the subject identifier can be included in the model specification conformula as a random intercept:

```
'age+factor(sex)+(1|id0)'
```

When the output path is specified as –o <output root> as in the example call above, the outputs of qcfc.R include the following:

| Output path	Output description
<out>.txt	A matrix containing the QC-FC correlation for each edge in the input matrix.
<out>_thr.txt	The above matrix, thresholded to include only significant edges; can be used to plot glass brain visualisations of significant edges, for instance using BrainNetViewer.
<out>_absMedCor.txt	The absolute median QC-FC correlation over all edges.
<out>_nSigEdges.txt <out>_pctSigEdges.txt	The number and percentage of edges with significant QC-FC correlations.
<out>.svg	A visualisation of the QC-FC distribution (see Figure 3).
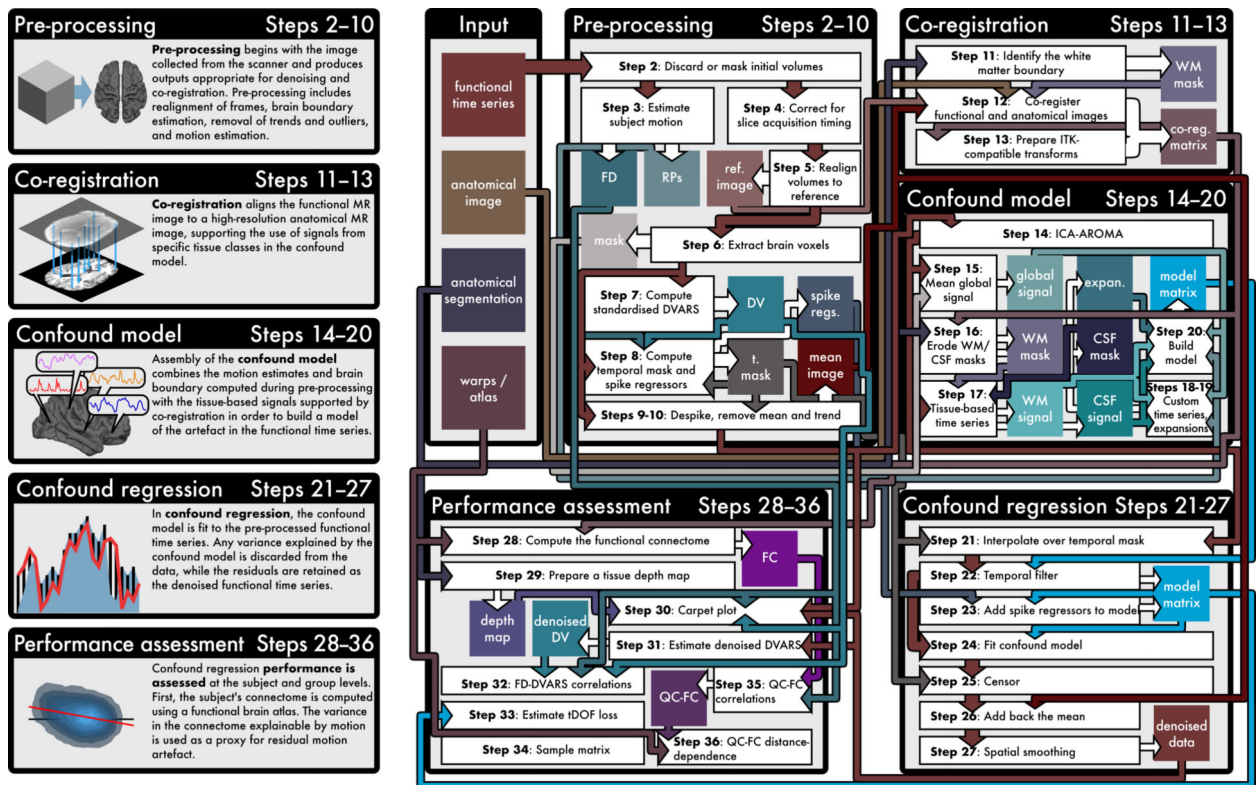
FIGURE 1 |. Workflow for motion correction of functional connectivity MRI data.
This processing stream estimates subject movement, minimally pre-processes the functional data, and aligns the data to a high-resolution anatomical reference (co-registration). Next, it builds a confound model using a combination of the global signal, signals from white matter and cerebrospinal fluid, motion estimates, time series expansions, and spike regressors. The confound model is then fit to the data in the confound regression step. A connectome is computed by warping a parcellation into the coordinate space of the time series and calculating functional connectivity between pairs of brain regions. Finally, diagnostic measures are produced to facilitate assessment of model performance and transparent reporting of denoising efficacy. A simplified schematic of key steps is shown at left, while a detailed flowchart of all protocol stages is shown at right. FD, framewise displacement; RPs, realignment parameters; DV, DVARS; WM, white matter; CSF, cerebrospinal fluid; FC, functional connectivity; QC-FC, quality control—functional connectivity correlations.
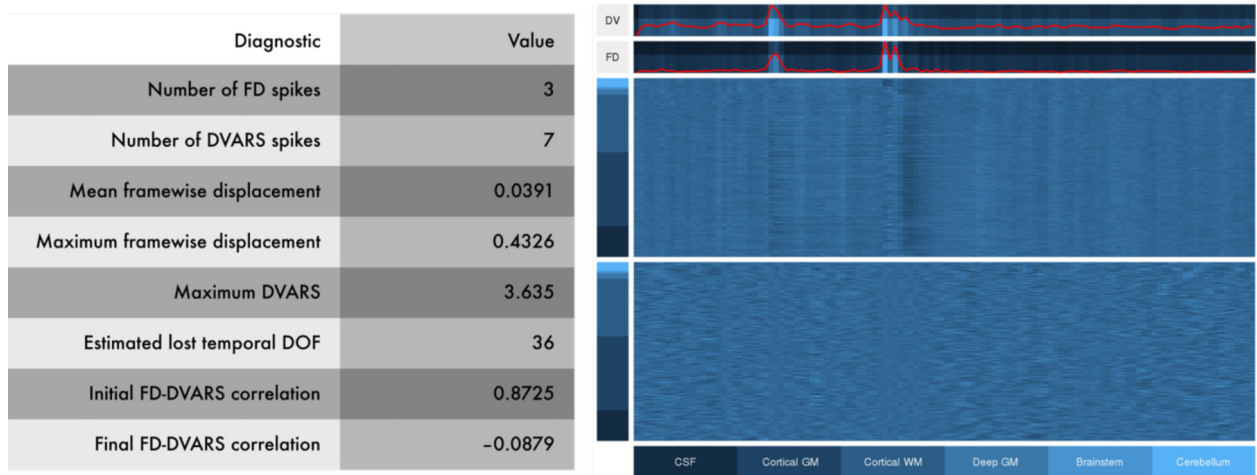
Diagnostic	Value
Number of FD spikes	3
Number of DVARS spikes	7
Mean framewise displacement	0.0391
Maximum framewise displacement	0.4326
Maximum DVARS	3.635
Estimated lost temporal DOF	36
Initial FD-DVARS correlation	0.8725
Final FD-DVARS correlation	−0.0879

FIGURE 2 |. Summary of subject-level performance diagnostics and anticipated results
(Steps 28—33). The illustrated results are from a single subject from the Philadelphia Neurodevelopmental Cohort (PNC)[85], processed using a 36-parameter stream that combines 6 realignment parameters, the mean WM and CSF time series, the mean global time series, and derivative and quadratic expansions. The table at left summarises quantitative diagnostics, while the panels at right display visual aids for performance assessment[55]. ***Left***, quantifications of subject movement indicate that this subject remained relatively still, apart from a few brief epochs of high movement. Notably, the GSR-based processing stream abolishes the strong FD-DVARS correlation[19] initially present in the data. ***Right***, a diagnostic visualisation[55] produced by voxts.R. ***Top right***, traces of the subject's framewise displacement and DVARS (DV). For this analysis, a frame was flagged as a spike if FD exceeded 0.25 or if standardised DVARS exceeded 2. The superthreshold region of the trace is demarcated by a darkened rectangle that covers the top fraction of each trace, allowing identification of flagged frames. In the traces shown, 3 frames exceed the FD threshold and 7 exceed the DVARS threshold. ***Middle right***, a voxelwise carpet plot of the subject's BOLD activity, computed over the minimally pre-processed image. Time is plotted on the abscissa and is synchronised to the quality traces at the top. Space is plotted on the ordinate, with voxels sorted according to their membership in tissue compartments (bottom right). Within each tissue compartment, time series are sorted from most superficial (bottom) to deepest (top). In this subject, movements are associated with global bands of signal loss (Type 2 artefact in Box 2), which is reflected in functional connectivity as global coupling. ***Bottom right***, after motion correction, the same subject's BOLD signal no longer exhibits global bands, although some loss of signal variance is visible in the most strongly contaminated frames.
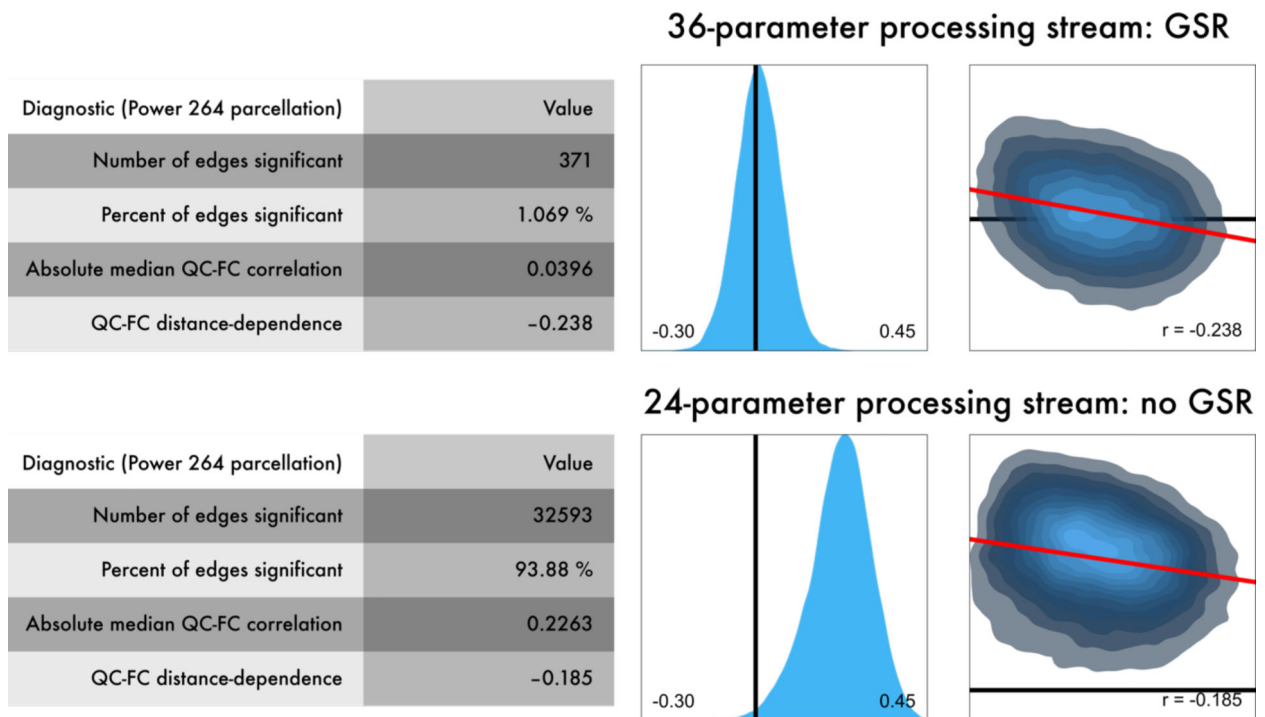
FIGURE 3 |. Summary of group-level performance diagnostics and anticipated results
(Steps 34—36). The illustrated results are from 500 low-motion subjects randomly sampled from the Philadelphia Neurodevelopmental Cohort (PNC)[85]. The top row shows performance diagnostics when the data are processed using a 36-parameter stream that combines the mean global time series, 6 realignment parameters, the mean WM and CSF time series, and derivative and quadratic expansions. The bottom row shows analogous results when the same data are processed using a 24-parameter stream that uses only the 6 realignment parameters with derivative and quadratic expansions. The table at left summarises quantitative diagnostics, while the panels at right display visual aids for performance assessment[22]. Results are shown for the 264-node Power parcellation[3]. ***Top left***, effective motion correction leaves only a small percentage of edges that are significantly related to motion, with a weak absolute median QC-FC correlation of approximately 0.04. ***Bottom left***, the 24-parameter model performs poorly by comparison, leaving a marked absolute median QC-FC correlation of 0.226, with nearly all edges exhibiting a significant relationship with motion. ***Centre right***, qcfc.R plots the distribution of QC-FC correlations across the two denoising schemes. In a high-performance processing stream such as the one presented at the top, the distribution is narrow and centred at approximately zero. An ineffective stream, in contrast, has a QC-FC distribution that is broader and positively centred. ***Far right***, a degree of QC-FC distance-dependence is unmasked by the GSR-based processing stream, as is evident in this visualisation produced by featureCorrelation.R. In comparison with the RP-only stream, motion artefact in the GSR-based stream is more strongly related to the Euclidean distance between network nodes. A processing stream that augments GSR with either censoring or signal decomposition techniques will typically exhibit less distance-dependence (see Box 2).

TABLE 1 |

Summary of subject-level quality indices.

Subject-level index	Description	Standalone implementations
Framewise displacement (FD)[11,12,70]	An estimate of the subjec s head movement from one frame of the time series to the next	•FSL: fsl motion outliers ($FD_{Jenkinson}$ and FD_{Power} •FSL: mcflirt ($FD_{jenkinson}$) •XCP: fd.R (FD_{power})
DVARS[11,72,73]	The temporal derivative of the root mean square intensity, an index of the frame-to-frame change in signal intensity across the brain	•FSL: fsl_motion_outliers XCP: dvars (standardised)
Outlier count	An index of the number of outlier values over all voxelwise time series within each frame	AFNI: 3dToutcount
"Quality index"	A measure of the dissimilarity of a frame with respect to the median value over all frames	AFNI: 3dTqual
FD-DVARS correlation[19]	The correlation between FD and DVARS; indexes the extent to which signal fluctuations relate to subject movement	XCP: featureCorrelation.R
Voxelwise displacement[17]	An estimate of each voxel's movement between consecutive frames of the time series	
Spike count (number of superthreshold frames)	The number or percentage of frames in a time series that exceed a motion threshold	
Loss of temporal degrees of freedom (tDOF)[25,86]	The number of temporal degrees of freedom lost during the de-noising process, typically estimated as the sum of the number of nuisance regressors and the number of censored frames	
Variance explained by nuisance time series and motion-BOLD contrasts[23,25,86]	Voxelwise maps or summary values indicating the fraction of variance in the data that can be explained by each term in the confound model, or the regression coefficients of a linear model fitting nuisance time series to BOLD time series	•AFNI: 3dTfitter • FSL: fsl_glm
Carpet plot / voxel plot[55]	A time-by-space matrix containing all values in the time series, often plotted alongside quality index time series, such as FD or DVARS	•plotdemo.m •XCP: voxts.R
Network identifiability[17,22,54,81,82]	The extent to which subnetwork structure can be resolved in the connectome; can be estimated as the modularity quality Q	
Test-retest reliability[23,25]	An estimate of the replicability of motion and functional connectivity estimates across repeated measures from the same subject	

TABLE 2 |

Summary of sample-level quality indices.

Sample-level index	Description	Standalone implementations
QC-FC correlations[10,11,22,23,28,54]	Correlations between mean framewise displacement and the functional connectivity between each pair of regions, computed across subjects	XCP: qcfc.R
Network-level QC-FC correlations[22,42]	Correlations between mean framewise displacement and network measures (such as modularity), computed across subjects	XCP: qcfc.R
QC-FC distance dependence[11,17,23,28,54]	The second-order correlation between edgewise QC-FC correlations and the Euclidean separation between region pairs	XCP: qcfcDistanceDependence
High- vs low-motion (HLM) contrasts[23,86]	Group-level comparisons between matched high- and low-motion subject bins	
Average tDOF loss, variability in tDOF loss[22,23,86]	The mean and variance in the loss of tDOF across subjects	
Discriminability[23,54]	Post-denoising sensitivity to between-group or individual differences	

TABLE 3 |

Troubleshooting table.

Step	Problem	Possible reason	Solution
8	Very few frames survive censoring	(i) Subject data are highly contaminated by noise; (ii) The censoring criterion is too stringent	(i) Exclude the subject from further analysis; (ii) Use a more lenient censoring threshold
16	Tissue time series are not produced	The tissue compartment mask is empty	Increase the retention criterion for mask erosion
22	Confound regression fails	The confound model is not well-formulated	Evaluate the confound model for missing values and collinearity; reduce or re-evaluate the model as necessary
26	Missing values (NA or NaN) in the connectome	(i) Poor registration quality; (ii) Parcels are of variable size	(i) Ensure that the sequence of transforms applied to the parcellation is correct, and then re-run registration steps as necessary, potentially using a different cost function; (ii) Select a different parcellation, or exclude smaller parcels from the connectome across the entire sample
33	The QC-FC distribution is very broad	Insufficient statistical power	Increase the sample size, or use subject-level diagnostics only.
33	No edges have a significant relationship with motion	(i) Insufficient statistical power (ii) Effective denoising model	(i) Increase the sample size, or use a less stringent correction for multiple comparisons. (ii) Ensure that statistical power is sufficient; otherwise, nothing to troubleshoot
33	Excessive residual QC-FC relationship	(i) Confound regression failed; (ii) High-motion subjects are driving the correlations; (iii) The confound model performs poorly in the data set under analysis	(i) See troubleshooting for Step 22; (ii) Apply subject-level exclusions based on motion; (iii) Use a different confound model, or benchmark performance of alternative models on a subsample of the data to select a model